

Côté Philo

Le journal de l'enseignement de la philosophie

Association pour la Création
d'Instituts de Recherche sur
l'Enseignement de la Philosophie

ACIREP

DOSSIER

L'Intelligence Artificielle en classe :
nouvel horizon critique pour la
philosophie ?

***Côté Philo* est une publication de l'ACIREPH**

*Association pour le Création d'Instituts de
Recherche sur l'Enseignement de la philosophie*

Éditeur : ACIREPH, 21 rue du Général Faidherbe, bâtiment A, 94130 NOGENT-SUR-MARNE

Directrice responsable : Fanny Bernard,
ACIREPH, 21 rue du Général Faidherbe, bâtiment A, 94130 NOGENT-SUR-MARNE

Rédacteur en chef : Serge Cospérec
ACIREPH, 21 rue du Général Faidherbe, bâtiment A, 94130 NOGENT-SUR-MARNE

Comité de rédaction : Delphine Blondel, Rémy David, Laurent Germain, Charlie Renard, Maxime Volta, Sylvain Theulle, Serge Cospérec

Imprimerie : Fadora, 55, rue Jean-Pierre Timbaud, 75011 PARIS

Les articles publiés par *Côté Philo* n'engagent que leurs auteurs.

Pour écrire dans *Côté Philo*

Adressez vos textes au comité de rédaction *email* : contact@acireph.org

Le Comité de rédaction informera l'auteur de sa décision : acceptation, acceptation sous réserve de modifications, ou non-publication.

Les textes envoyés ne sont pas retournés à leurs auteurs

**Retrouvez *Côté Philo* et les autres travaux de
l'ACIREPH sur notre site
www.acireph.org**

Côté Philo

Le journal de l'enseignement de la philosophie

ÉDITORIAL	Laurent Germain	3
DOSSIER	L'Intelligence Artificielle en classe : nouvel horizon critique pour la philosophie ?	
	Quelques applications utilisant les IA génératives pour l'enseignement de la philosophie Vivien Mirebeau	9
	Compte-rendu enthousiaste de l'atelier de Marion Schumm sur ChatGPT comme « outil critique pour les élèves » Jean-Pierre Dramisino	17
	De quelles valeurs nos techniques en général et l'IA en particulier sont-elles porteuses ? Adeline Barbin	23
	Le procès de la voiture autonome Théo Milana	31
	« "Intelligence artificielle" est un terme idéologique ». Entretien avec Anne Alombert, autrice de <i>Schizophrénie numérique</i> Anne Alombert (transcription Fanny Bernard)	35
	IA génératives : faut-il sauver l'école ou les élèves ? Sylvain Theulle	43
	Chronique de la critique de la dissertation : 2024, comment surmonter les copies ChatGPT ? Fanny Bernard	51
	Politique et IA : entre <i>1984</i> d'Orwell et le <i>Procès</i> de Kafka Serge Cospérec	61

Éditorial

Laurent Germain

Ce numéro est consacré aux ateliers et réflexions proposées et produites à l'occasion des Journées d'études (JE) de l'Acireph des 7 et 8 novembre 2024 au Lycée Jean Zay à Paris, sur *l'intelligence artificielle*. Il s'agissait de proposer, selon l'esprit même qui anime ces journées d'études et l'Acireph, des partages de pensées et de pratiques entre pairs, ainsi que des éclaircissements venus d'autres champs disciplinaires ; ceux-ci concernent l'impact et les enjeux qui suivent l'émergence et la banalisation de l'usage de l'IA dite générative, et en particulier des agents conversationnels désormais familiers du grand public, et, bien sûr, de nos élèves.

L'arrivée de ChatGPT a en effet rapidement affecté notre métier, et d'abord par l'apparition des devoirs rédigés par une IA, nous poussant à n'évaluer que des travaux faits en classe. Nous sommes contraints à ce pis-aller, faute de temps et de moyens pour réfléchir collectivement aux transformations pédagogiques impliquées par une IA qui semble là pour rester. Comment éviter de subir ? Comment faire des choix éclairés concernant l'usage possible de ces IA ? Faut-il travailler contre l'IA ? sans elle ? Ou avec elle, mais comment et pourquoi faire ?

L'IA modifie aussi certaines questions de philosophie. Quelle est la frontière entre l'humain et la machine ? Qu'est-ce au juste que l'« intelligence » ? Comment analyser les pratiques de pouvoir, les phénomènes de persuasion et de manipulation à l'époque de l'IA ? L'IA prépare-t-elle un monde plus juste ou plus inégalitaire ? Une culture technique, critique et en prise avec les questions contemporaines, est plus que jamais nécessaire aux élèves.

ChatGPT est-il intelligent ?

Les nouvelles IA sont capables d'apprendre et de se perfectionner de façon autonome. Elles sont savantes et patientes. On leur écrit, on leur parle ; elles nous écoutent ou nous lisent, elles nous comprennent et nous donnent des conseils. Elles sortent victorieuses du test de Turing : une machine peut tromper un humain dans une conversation. Leurs performances sont étonnantes.

Langage, perfectibilité, savoir : la frontière entre l'humain et la machine est-elle en train de tomber ? Les IA ont-elles égalé, voire dépassé, l'intelligence humaine ? Sont-elles capables de raisonner, d'inventer, de prendre des décisions, comme un être humain ? Auraient-elles même des états mentaux ?

Sur toutes ces questions, les débats sont vivants. Un cours de philosophie gagnerait à les aborder. Seulement, nous sommes peu familiers du domaine de l'IA. Comment fonctionne exactement ChatGPT ? Pourquoi est-il si performant ? Comment expliquer ses failles ?

Usages didactiques de l'IA : contrainte, duperie ou opportunité ?

L'IA outil. Le monde enseignant a d'abord découvert l'IA par la fraude et la triche. Mais l'hostilité à l'IA cède peu à peu la place à la curiosité, voire un intérêt. Et si ChatGPT pouvait nous aider dans notre travail ? Par exemple, à faire le plan d'un cours, sa description ou son résumé, à créer des QCM ou des questions sur un texte, à générer des exercices ou à corriger les copies.

L'IA, pour renouveler l'enseignement. Le fait que les IA génératrices de texte produisent des imitations tout à fait convaincantes d'une dissertation philosophique traditionnelle invite à questionner la pertinence de cet exercice. L'IA pourrait être l'occasion de repenser le format ou la nature des productions demandées aux élèves, de pratiquer l'écriture ou la lecture de textes philosophiques. Il s'agirait moins, alors, de diaboliser l'outil que d'apprendre à l'utiliser honnêtement et intelligemment, comme le font déjà certaines enseignantes ; à en percevoir aussi les limites, les dangers et les dérives.

Aucune de ces questions n'est taboue à l'ACIREPh. Elles ont été réfléchies dans certains ateliers sur des expériences faites par des collègues, partagées et soumises à la discussion. Ces journées ne se voulaient donc ni être une longue complainte technophobe stérile, ni un optimisme progressiste aveugle, mais un effort de compréhension collectif, et d'abord de cette technologie, — on retrouvera sur le site de l'ACIREPh la conférence « *Comment ça marche, l'IA ?* » de Bernard Victorri¹ —, ainsi que des présupposés qui la fondent, des enjeux philosophiques, cognitifs, sociaux, environnementaux qu'implique son utilisation, dont les effets n'ont fait que s'accroître depuis ces journées d'études. Il s'agissait enfin d'explorer les possibilités qu'elle ouvre ou les transformations qu'elle opère dans la pédagogie et l'enseignement de la philosophie, de questionner leur pertinence et leurs limites. Ainsi les articles présentés sont répartis en deux parties thématiques.

¹ Bernard Victorri, Directeur de recherche au CNRS, est un linguiste, spécialiste de la modélisation sémantique et des réseaux connexionnistes. La conférence « *Comment ça marche, l'IA ?* » est [disponible en ligne](#).

La première partie de ce numéro est consacrée à des recensions et reprises d'ateliers de pratique pédagogique présentés aux JE, par leur auteur ou par ceux qui y ont assisté. Vivien Mirebeau présente deux applications qu'il a lui-même programmées et l'usage et qu'on peut en faire avec les élèves : la première permet de dialoguer avec des philosophes du programme incarnés par des chatbots, la seconde, de s'entraîner à la méthodologie de la dissertation. Jean Pierre Dramisino présente l'atelier de Marion Schumm consacré à un « concours de triche » où les élèves s'exercent à produire une bonne introduction de dissertation en apprenant à formuler les instructions (le « prompt ») à donner à l'IA (ici ChatGPT). Adeline Barbin synthétise son travail effectué en classe sur la question de la neutralité axiologique de la technique, appliqué notamment à cette nouvelle technologie qu'est l'IA générative. L'article de Théo Milana résume sa proposition de séance de classe sur le procès d'une voiture autonome impliquée dans un accident et ses enjeux philosophiques, ainsi que la discussion des questionnements pédagogiques qui s'en est suivie. Enfin, Fanny Bernard transcrit les échanges avec Anne Alombert qui ont suivi l'atelier consacré à la lecture de son essai *Schizophrénie numérique*. Ce qui permet à l'autrice de revenir sur l'analyse des présupposés et des effets des « intelligences » artificielles, de l'idéologie qui accompagne leur promotion, en lien avec les projets des acteurs majeurs du domaine.

La seconde partie du numéro propose des réflexions et argumentations de trois membres de l'association sur l'impact de l'IA sur l'institution scolaire, sur l'enseignement et l'évaluation en philosophie, et plus généralement sur la société. L'article de Sylvain Theulle porte sur la triche - sans doute le premier prisme à travers lequel chacun de nous a été confronté à ChatGPT ou ses homologues -, phénomène que l'institution scolaire se refuse à affronter sérieusement, alors que la fraude ruine la justice des examens et des évaluations, et remet en cause ce qui donne sens à l'enseignement, à savoir l'apprentissage et le développement de l'autonomie intellectuelle. Prolongeant la réflexion sur les effets de l'IA sur l'évaluation en s'appuyant sur les ateliers des JE auxquels elle a assisté, Fanny Bernard questionne simultanément l'intérêt de maintenir les exercices écrits, à commencer par la dissertation, et la pertinence qu'il y a à encourager son utilisation ou s'en accommoder, au regard des enjeux dépassant de loin le champ scolaire. Elle propose de développer l'oral, parent pauvre de l'enseignement actuel, et des évaluations reposant sur des exercices oraux. Enfin Serge Cospérec propose une lecture critique des effets sociaux et politiques de l'IA à partir des métaphores orwellienne (1984) et kafaienne (*Le Procès*). Bonne lecture à toutes et tous !

Laurent Germain

Dossier

**L'IA en classe : nouvel horizon
critique pour la philosophie ?**

Quelques applications utilisant les IA génératives pour l'enseignement de la philosophie

Vivien Mirebeau

La publication de ChatGPT le 30 novembre 2022 (et sa diffusion fulgurante¹) a soulevé des craintes légitimes quand les enseignants se sont rendu compte que ce modèle était suffisamment « intelligent » pour égaler sans mal - et souvent dépasser - la qualité des productions d'un élève de Terminale, quelle que soit la matière. Le spectre d'une nouvelle forme de triche, plus efficace et plus redoutable parce qu'indétectable, commença à faire son chemin dans les esprits.

Il n'est pas question ici de contester cette utilisation possible de ChatGPT. Il est clair qu'il constitue à bien des égards le dispositif idéal pour un élève qui voudrait produire des devoirs solides sans fournir d'effort intellectuel. La question que je voudrais traiter est plutôt la suivante : même si les « grands modèles de langue » (LLM²) comme ChatGPT s'offrent naturellement à des usages vicieux, est-il possible d'employer ces dispositifs de façon *vertueuse*, c'est-à-dire de les mettre au service des apprentissages de l'élève ?

J'ai pour ma part été rapidement convaincu que de tels usages étaient possibles, à condition de résoudre un problème fondamental : comment faire usage des LLM dans un contexte pédagogique, alors que ces derniers reposent sur une technologie probabiliste particulièrement peu fiable, toujours sujette aux « hallucinations³ » ? Pour résoudre ce problème, une solution consiste à encadrer par

¹ Deux mois après son lancement, ChatGPT comptait 100 millions d'utilisateurs actifs mensuels, ce qui en fait l'application avec la croissance la plus rapide de l'histoire. Voir la note d'analyse de Krystal Hu, « ChatGPT sets record for fastest-growing user base ». [[Disponible en ligne](#)]

² Comme ChatGPT est souvent le seul grand modèle de langue connu du grand public, nous utiliserons ici abusivement les expressions « ChatGPT » et « LLM » comme des synonymes. Il y a bien entendu d'autres LLM disponibles : Llama, Claude, Deepseek, Gemini, Mistral, etc.

³ On appelle « hallucination » la façon dont un LLM peut produire des énoncés parfaitement erronés, en suivant exclusivement la logique probabiliste de la langue.

un « pré-prompt » les requêtes qui sont envoyées aux LLM. Il faut ici rappeler la différence entre un prompt et un « pré-prompt » :

- Un **prompt** est l’instruction que vous envoyez à votre modèle de langue. Par exemple, si vous ouvrez ChatGPT et que vous envoyez « Quelle est la plus grande ville du Brésil ? », cette phrase constitue votre prompt. L’application vous répondra « São Paulo », en vous offrant quelques détails supplémentaires.
- Un **pré-prompt** est un ensemble d’instructions que vous envoyez *avant votre prompt principal* : il peut en guider le sens, en informant votre LLM sur ce qu’il doit en faire. Par exemple, si vous envoyez à ChatGPT « Nous sommes au début du XXe siècle ; réponds en espagnol. », et que vous faites suivre ce pré-prompt du prompt précédent, ChatGPT vous répondra « Rio de Janeiro », dans la langue désirée.

L’idée est donc la suivante : on peut obtenir une réponse fiable de ChatGPT si on cadre le pré-prompt de façon suffisamment précise, en spécifiant les attentes pédagogiques. Si pour l’utilisateur qui dialogue en ligne avec ChatGPT cette distinction entre prompt et pré-prompt est relativement arbitraire et oiseuse, en revanche pour les programmeurs elle est tout à fait capitale. Il est en effet très facile de réaliser ce « cadrage » des prompts quand on programme une application : le prompt est fourni par l’utilisateur, et le pré-prompt est imposé par le programme pour préciser le cadrage désiré.

Par conséquent, il m’a semblé que la condition pour inventer des usages pédagogiquement vertueux de ChatGPT, c’était de développer des programmes qui encadrent suffisamment bien les instructions envoyées par l’élève. En suivant ce principe, j’ai développé plusieurs applications pédagogiques pour la philosophie, appuyées sur ChatGPT. Dans le présent article, je voudrais vous en présenter deux : MethodoPhilo, un assistant à la méthodologie de la dissertation, et PhiloGPT, un chatbot spécialisé dans la philosophie. Ces deux applications partagent le fait d’être gratuites et disponibles sans inscription ni installation.

I. MethodoPhilo : un assistant à la méthodologie de la dissertation

MethodoPhilo (<https://methodophilo.vmirebeau.fr>) est une application dont l’objectif premier est d’accompagner les élèves dans la rédaction de leur dissertation de philosophie. L’acquisition de la méthodologie en Terminale est toujours une étape difficile, particulièrement pour les élèves les plus fragiles. Il s’agit ici de découper la méthode de la dissertation en une série d’étapes simples, clairement expliquées.

Quelques applications utilisant les IA génératives pour la philosophie

Concrètement, l'élève qui se connecte à MethodoPhilo doit d'abord choisir son sujet. Il peut l'entrer manuellement, le trouver dans une base de données, ou charger l'un de modèles de dissertation complètement rédigés.

Une fois que le sujet est déterminé, l'interface principale apparaît. Dans la zone principale de progression en haut, il peut suivre l'enchaînement des grands moments : introduction, développement (décomposé en parties), conclusion. Quand on clique sur l'un de ces moments, on voit apparaître en-dessous un ensemble d'étapes, qui sont autant de tâches que l'élève doit remplir pour construire petit à petit sa dissertation. Par exemple, pour produire l'introduction, on demande à l'élève d'identifier les concepts principaux, de définir chaque concept, de construire les deux parties de la problématique, puis de rédiger l'annonce de chaque partie du plan.

Pour l'instant, les fonctions décrites reposent sur du codage tout à fait classique. Ce qui fait l'originalité de MethodoPhilo, c'est que chacune des tâches simples proposées à l'élève peut faire l'objet d'une évaluation par une IA. Par exemple, imaginons que l'élève traite du sujet « Le bonheur dépend-il de nous ? », et qu'il doive donc définir le concept de bonheur. Quand il a renseigné la zone de texte concernée, il peut cliquer sur le bouton situé juste en-dessous, qui représente un enseignant : une requête va être envoyée au LLM¹, qui va jouer le rôle d'un tuteur en jugeant si la production de l'élève est bien conforme aux attentes méthodologiques. Ce jugement est délivré sous deux formes : d'abord un commentaire en langue naturelle, qui explique ce qui convient et ne convient pas ; ensuite une évaluation synthétique, sous la forme d'un feu de signalisation rouge, orange ou vert. Dans tous les cas, des instructions claires sont données pour que l'IA ne remplisse pas la tâche à la place de l'élève : il s'agit de le guider dans son travail, jamais de le faire à sa place. De façon générale, ChatGPT est utilisé non pas comme un assistant conversationnel, mais comme un évaluateur, en s'appuyant sur tous les attendus méthodologiques nécessaires.

Que contient le pré-prompt ?

Pour que le LLM réponde d'une façon satisfaisante, MethodoPhilo lui fournit les instructions suivantes :

- Quelques informations sur son identité et son rôle
- Des conseils de méthodologie spécifiques pour la tâche concernée
- Des instructions de cadrage sur la façon dont il doit formuler la réponse

¹ Il s'agit du modèle gpt-4o-mini d'OpenAI dans la version actuelle, mais il devrait bientôt être remplacé par un modèle de langue plus ouvert.

- Des exemples de tâches réussies ou non, avec le commentaire et l'évaluation attendus

Le **prompt**, lui, est donné par l'utilisateur, qui essaye de remplir la tâche qui lui est assignée.

Si l'évaluation est insatisfaisante, l'élève est encouragé à retravailler sa production avant de demander à nouveau le jugement du tuteur. L'efficacité de MethodoPhilo repose sur ce principe de micro-évaluations successives. Quand une tâche de l'introduction est remplie, la tâche suivante se débloque. Quand toutes les tâches de l'introduction sont remplies, la première partie du développement (et la première tâche associée) se débloque. Tout au long de ce processus et pour chacune des tâches, des explications méthodologiques brèves sont fournies à l'élève qui en fait la demande. Dans l'idéal, l'élève s'approprie ainsi les bonnes pratiques de la dissertation sans se perdre dans l'angoisse de la page blanche.

Cette application est pensée pour favoriser l'autonomie de l'élève du point de vue de la méthode. Pour une mise en place concrète, on peut réserver une salle informatique et donner une heure aux élèves organisés en binômes pour produire une introduction complète sur un sujet déterminé, en essayant d'obtenir du tuteur le maximum de bonnes notes possible. Sur cette heure, l'enseignant peut circuler entre les postes pour clarifier les points sur lesquels l'IA pourrait se tromper ou être trop exigeante, et offrir son propre avis lorsqu'un élève conteste les remarques de l'IA. Cette organisation a pour mérite de donner un rythme naturel à la séance : à chaque étape, l'élève regarde la suggestion de l'IA, y réfléchit, et sollicite si besoin le professeur pour arbitrer. De son côté, le professeur n'est pas submergé par les questions répétitives sur la structure de la dissertation, puisqu'une partie de l'accompagnement est déléguée à l'IA.

MethodoPhilo n'est cependant pas sans défauts. Il se base sur une méthode particulière de la dissertation, qui ne fait pas consensus parmi les enseignants¹, et qui est implémentée de façon rigide. Par ailleurs, l'IA demeure un outil statistique qui peut proposer des évaluations inappropriées (souvent plus sévères que nécessaires, semble-t-il). Il faut donc bien cadrer l'utilisation de cette application auprès des élèves ; elle ne saurait se substituer ni à la réflexion personnelle, ni à un véritable accompagnement humain, qui peuvent seuls garantir une réelle progression.

MethodoPhilo contribue à la fois à la familiarisation des élèves avec un exercice complexe et au renforcement de leur autonomie. En structurant leur démarche, il leur évite de se perdre dans les méandres d'une dissertation désordonnée. En tant qu'enseignant, on peut y trouver un moyen de consacrer davantage de temps

¹ Certains enseignants, par exemple, conseillent de ne pas définir les termes du sujet dans une introduction.

à la discussion philosophique proprement dite, plutôt que de répéter sans cesse des consignes méthodologiques de base.

II. PhiloGPT : un chatbot spécialisé dans la philosophie

Contrairement à MethodoPhilo, qui se présente comme un tuteur méthodologique, PhiloGPT (<https://philogpt.v mirebeau.fr>) est un chatbot, un robot conversationnel comme l'est ChatGPT. A la différence de cette application cependant, PhiloGPT n'est pas un assistant généraliste qui serait utile dans toutes les situations. PhiloGPT donne plutôt accès à une galerie de personnages, entre lesquels on peut choisir pour discuter librement. Les personnages en question sont vingt-neuf grand-es philosophes du programme : Platon, Aristote, Descartes, Weil, etc. L'idée de l'application est de permettre aux élèves de philosophie de « tchatter » avec ces personnages exactement comme on parlerait à un proche. L'interface, directement inspirée de Whatsapp, renforce cette familiarité.

Sur le plan technique, PhiloGPT repose sur le même type de technologie que ChatGPT, mais chaque philosophe est construit à l'aide d'un prompt qui lui est propre.

Que contient le pré-prompt ?

- Des instructions de cadrage de l'application (« tu es une application de dialogue... »)
- Des informations biographiques sommaires sur l'auteur
- Ses thèses principales
- Quelques informations sur la façon de parler et d'argumenter propre à l'auteur
- La forme attendue pour les réponses
- L'instruction suivante : "récupère systématiquement de l'information avant de répondre"

Tout l'intérêt pédagogique de PhiloGPT tient dans cette dernière instruction. En fait, chacun des philosophes a été « nourri » avec ses œuvres principales, grâce à une technologie qu'on appelle le RAG (Retrieval-Augmented Generation, « Génération augmentée par de la récupération [de données] »). Le RAG implique de prendre un fichier contenant du texte, et de convertir ce texte en valeurs numériques qui en représentent le sens¹. À partir du moment où un texte est ainsi converti en vecteurs, il est très facile de procéder à des opérations

¹ Cette opération qui convertit du sens en nombres se nomme « embedding » (qu'on traduit souvent par « vectorisation »).

mathématiques pour identifier la proximité entre deux vecteurs distincts, et donc entre deux textes. Sans rentrer dans les détails techniques, cela permet de faire en sorte que quand on pose une question à un chatbot, celui-ci peut chercher dans l'ensemble des documents qui lui ont été fournis les passages dont le sens s'approche le plus de la question. Concrètement, quand vous posez une question à Platon dans PhiloGPT, tout se passe comme si votre interlocuteur relisait les œuvres complètes de Platon afin de trouver les cinq passages les plus pertinents pour construire sa réponse. Cela permet de réduire drastiquement le risque d'hallucinations, puisqu'il est demandé explicitement au LLM de se reposer sur les écrits authentiques du philosophe. Pour l'utilisateur, ces extraits sélectionnés sont accessibles en cliquant sur le bouton « Voir les sources », situé tout en bas de chaque message généré.

Quelques fonctionnalités supplémentaires sont disponibles : des questions écrites à l'avance sont suggérées à l'élève, on peut accéder à une courte fiche d'informations sur chaque philosophe, et on peut même organiser des groupes de discussion pour laisser les philosophes débattre entre eux autour de la question posée par l'utilisateur.

L'usage pédagogique le plus immédiat de PhiloGPT consiste à laisser l'élève dialoguer librement avec ces grandes figures de la philosophie. L'application est d'abord pensée pour les élèves qui se sentiraient étrangers à la matière, et tout est fait pour leur suggérer un rapport ludique et familier avec les auteurs.

Au-delà de cet usage spontané, on peut proposer des mises en place plus structurées :

Exercice 1 : Cartographier une notion (en préparation d'une dissertation)

Modalités : Travail individuel à la maison, puis rédaction de la dissertation en classe (entre 2 et 4 heures).

Déroulement : L'enseignant annonce aux élèves le thème du prochain devoir, qu'il ait déjà été abordé ou non en cours. Il fournit également une liste de philosophes qui ont traité ce sujet. Par exemple, si l'on aborde la notion de technique, on précise que Descartes, Hegel, Marx, Heidegger, Simone Weil et Hans Jonas ont des réflexions pertinentes sur le sujet. L'objectif pour l'élève est alors de composer une carte mentale de ces penseurs et de leurs liens : que dit Marx de la technique ? En quoi diffère-t-il de Hegel ? Comment Simone Weil critique-t-elle Marx à son tour ? On met l'accent sur l'aspect graphique de la carte. Pour nourrir sa réflexion, l'élève s'appuie sur les sources attachées à chaque réponse dans l'application, recueille les citations les plus marquantes, puis apporte sa carte en classe et l'utilise pour traiter le sujet imposé.

Avantages : Cette approche incite l'élève à participer activement à la construction de son propre savoir : c'est lui qui recherche et structure les contenus utiles à la dissertation. Les échanges avec les chatbots lui permettent en outre d'éclaircir sa

Quelques applications utilisant les IA génératives pour la philosophie

compréhension en soumettant directement ses questions et critiques aux philosophes concernés. On peut aussi demander aux élèves qui auront produit les meilleures copies de partager leur travail : comme chacun aborde des références et des analyses différentes, la lecture des travaux de ses camarades est particulièrement enrichissante.

Inconvénients : Le travail autonome comporte un risque d'erreur. Il est donc vivement recommandé de reprendre, au moins à l'oral pour toute la classe, les principaux éléments théoriques après la remise des copies.

Exercice 2 : Adopte un philosophe

Modalités : Travail à la maison, en individuel ou en petits groupes (une ou deux personnes).

Déroulement : On informe les élèves qu'ils devront choisir un philosophe dans la liste des penseurs disponibles via l'application, pour la semaine suivante. Ils disposent ainsi d'une semaine pour dialoguer avec différents philosophes virtuels et sélectionner celui ou celle qui les intéresse le plus. La semaine suivante, une feuille indiquant les noms des philosophes circule en classe : chacun s'inscrit soit seul, soit avec un camarade. Les élèves ont ensuite deux ou trois semaines pour réaliser une fiche sur le philosophe adopté : une courte biographie, ses problématiques clés, ses thèses et concepts, ainsi que les critiques qui lui ont été adressées¹. Les fiches réalisées sont ensuite relevées, évaluées, et après correction déposées sur un dossier partagé pour que chaque élève puisse y accéder. Tout au long de l'année, les auteurs de la fiche sont considérés comme les « spécialistes » du philosophe en question : on pourra se référer à eux en classe pour enrichir la discussion, et ils auront la tâche de défendre la validité de son point de vue.

Avantages : Cet exercice, très complet, mobilise de nombreuses compétences : autonomie, travail collaboratif, réflexion conceptuelle, rédaction et expression orale. Son intérêt principal réside dans l'acquisition d'un savoir différencié : chacun s'approprie un contenu philosophique distinct, ce qui alimente les échanges entre élèves en classe et au-delà.

Inconvénients : Ce dispositif prend du temps. Il peut donc être préférable de favoriser le travail en binôme plutôt que strictement individuel.

PhiloGPT ne remplace ni l'échange professoral ni la rigueur du texte original, mais il offre à l'élève l'occasion de se confronter à des idées, de tester ses hypothèses et de mieux comprendre le jeu des arguments. En cela, il offre une alternative à la simple recherche documentaire. L'élève a affaire non pas à des articles descriptifs, mais à un interlocuteur virtuel capable de réagir, de clarifier et parfois de se tromper, ce qui le force à exercer sa vigilance. Ce faisant, il se

¹ Il est possible de contrôler les connaissances obtenues sur le philosophe en soumettant à chaque élève un QCM très simple sur l'auteur choisi.

familiarise peu à peu avec ce qu'est véritablement la philosophie : un travail de questionnement critique et de construction d'un savoir étayé.

MethodoPhilo n'écrit pas la dissertation, PhiloGPT ne remplace pas les textes originaux : chacune de ces applications apporte des aides ciblées, que l'enseignant peut ensuite intégrer à sa propre pédagogie. Cette intégration suppose toutefois une vigilance constante pour ne pas laisser se diffuser des contenus erronés et pour éviter que l'IA ne devienne un palliatif systématique à la réflexion personnelle. Pourtant, si ces dérives sont maîtrisées, ces outils numériques constituent une proposition intéressante pour adapter l'enseignement de la philosophie aux pratiques d'une génération souvent déjà familiarisée avec l'usage d'Internet et des chatbots, en leur montrant qu'on peut faire de la philosophie autrement : en expérimentant, en discutant et en construisant peu à peu un savoir fondé sur des exigences rigoureuses.

Pour aller plus loin, on pourrait imaginer que les enseignants de philosophie ne se contentent pas de faire usage d'outils construits autour de l'IA, mais puissent eux-mêmes paramétrer et configurer leurs chatbots, quand bien même ils ne disposent pas de compétences de codage. C'est pour cette raison que j'ai développé EduBot (<https://edubot.fr>). N'importe quel enseignant disposant d'une adresse mail académique peut accéder gratuitement aux fonctions de création de bots, et générer des ressources interactives (QCM, audios, cartes mentales...) à partir de son cours.

Il reste encore beaucoup d'usages pédagogiques à inventer autour des IA génératives. Cela implique cependant des questions plus difficiles : *quels modèles de langue* devons-nous intégrer dans nos applications et dans nos cours ? Des modèles fermés et efficaces, ou des modèles plus ouverts mais parfois relativement instables ? Coûteux sur le plan de l'énergie, ou frugaux ? Avec quels filtres de modération ? Derrière la question des IA génératives se nouent aussi des tensions économiques, sociales, écologiques et politiques qui ne doivent pas être ignorées, et qui nécessiteraient un cadrage clair de la part des institutions.

Vivien Mirebeau

Compte-rendu enthousiaste de l'atelier de Marion Schumm sur ChatGPT comme « outil critique pour les élèves »

Jean-Pierre Dramisino

Introduction

Je me propose de faire la présentation à la fois enthousiaste et subjective d'un atelier de Marion Schumm auquel j'ai participé lors des Journées de l'ACIREPH en novembre 2024. Il était intitulé « ChatGPT : un outil critique pour les élèves ? » – J'y suis allé en tant qu'utilisateur de ChatGPT¹, parce que j'avais essayé, sans grande réussite, de faire travailler les élèves à partir de textes produits par cette intelligence artificielle générative (IAgen). Je le dis tout de suite : j'aurais aimé trouver le dispositif que Marion Schumm nous a proposé ! Elle se dit non spécialiste de l'intelligence artificielle, je veux bien la croire ; c'est en tout cas un dispositif intelligent, qui participe à l'apprentissage de la dissertation en philosophie, et plus largement, qui permet de développer l'esprit critique de nos élèves. Il peut, de plus, nous amener à réfléchir sur notre enseignement tout en adoptant une posture critique à l'égard de cette machine – sans pour autant en bannir l'usage.

Le contexte est le suivant : Marion Schumm cherche à répondre à la difficulté que l'on peut parfois rencontrer face aux élèves qui pourraient ne plus considérer comme intéressant ou important de faire l'effort de rédiger des propos développés, construits et argumentés. Je reprends son propre texte de présentation de l'atelier : « À quoi bon tous ces efforts, si la simple probabilité sémantique, guidée par des prompts pertinents, semble pouvoir aboutir, et ce en quelques instants, à un résultat comparable, voire meilleur ? » L'idée est donc de prendre le contrepied de ce découragement, autant que des représentations biaisées que peuvent avoir nos élèves de ChatGPT. Son idée est d'obliger les élèves à s'en servir de façon critique. Par son atelier, elle vise également à nous montrer, en tant

¹ On en reste ici à ChatGPT, car il est le plus connu ; mais on peut pratiquer d'autres IAgen (Intelligence Artificielle générative) comme Gemini (Google), Perplexity, Mistral AI, DeepSeek, etc. Je parlerai, plus généralement dans cet article, d'IAgen (Intelligence Artificielle générative).

qu'enseignant-es de philosophie, en quoi un usage pertinent et critique de cette IAgen – pertinent parce que critique – est possible...

Son atelier nous a justement fait vivre ce qu'elle met en place avec ses élèves.

Faire une dissertation à l'ère des IAgen

Tout cela s'inscrit dans un travail sur la méthode de la dissertation. Marion Schumm nous explique ne pas trop développer en cours les consignes préparatoires à cet exercice ; et ce, afin d'amener ses élèves à une réflexion critique sur leurs productions pour les faire progresser. L'idée est de promouvoir une approche *normative* dans la dissertation et pas seulement d'en rester à l'application mécanique de règles formelles – ce qui de toute façon achoppe à produire une réflexion de type philosophique...

C'est pourquoi elle se refuse de faire un corrigé. Par flemme, dit-elle humblement. Mais au-delà de cette boutade, admettons avec elle que les corrigés que nous pouvons proposer sont à peine lus et ne sont pas véritablement des outils de travail pour nos élèves pour faire les dissertations suivantes. Au mieux, les élèves vont répéter des formules toutes faites, parfois à contre-emploi, mais le plus souvent iels ne vont pas s'en servir et/ou ne pas savoir qu'en faire. Donc derrière cette flemme, il y a la prise en compte d'une difficulté de notre enseignement... et des élèves.

Paradoxalement, c'est par le truchement de ChatGPT qu'elle se propose de faire travailler le corrigé d'une dissertation.

« Du vent... »

Notre collègue a donc donné à ses élèves une dissertation dont le sujet est : « Comment découvrir la vérité ? ». Comme je le disais plus haut, elle ne donne pas trop de consignes. Elle se limite par exemple, nous a-t-elle dit, à indiquer la finalité de l'introduction comme « règle de méthode » pour en rédiger une. Les élèves vont ainsi produire leur première dissertation, elle les corrige et vient la séance de rendu des copies et surtout, de la correction... sans corrigé. Pour faire réfléchir les élèves à leur introduction et plus largement à l'introduction d'une dissertation de philosophie, Marion Schumm propose alors un « concours de triche » à l'aide de ChatGPT. Manière ludique de faire réfléchir les élèves sur l'écriture philosophique, mais aussi sur l'usage de cette machine.

L'idée est d'abord de « demander » à ChatGPT de faire l'introduction de cette dissertation et de la critiquer en classe. Le *prompt* est réduit à sa plus simple expression : « Fais une introduction à la dissertation de philosophie "comment découvrir la vérité ?" ». Indépendamment des progrès actuels dans les réponses de

L'atelier de Marion Schumm sur ChatGPT comme « outil critique pour les élèves »

cette IAgen¹, celle-ci produit souvent des généralités creuses, « du vent » comme le dit Marion Schumm. Exemple de la première phrase d'une introduction produite par ChatGPT : « La vérité est une notion centrale en philosophie, souvent considérée comme le but ultime de la connaissance et de la réflexion ». L'introduction « venteuse » est projetée ou distribuée aux élèves et notre collègue commence à catégoriser la réponse proposée phrase par phrase. Alors que les élèves trouvent souvent ce type d'introduction comme satisfaisante, la catégorisation a ainsi pour but de leur faire prendre conscience que l'on pourrait utiliser ces phrases creuses dans n'importe quelle introduction : il suffit de changer la notion convoquée. D'autres phrases peuvent être intéressantes, mais offrir d'autres imperfections, etc.

L'intérêt de cette première étape dans la démarche est ainsi de critiquer ce que nos élèves produisent souvent – des généralités creuses – sans que pour autant ils ne se sentent visés. ChatGPT est le bouc émissaire parfait : il donne un exemple qu'on peut critiquer à loisir sans craindre que nos élèves ne s'en offusquent. Si une part assez minime de l'introduction de l'IAgen est satisfaisante, comment en produire une pleinement opératoire... avec la machine ?

Le concours de triche

Il s'agit maintenant de rédiger, et surtout de faire « rédiger » une introduction digne de ce nom. C'est l'objet du **concours de triche**.

La proposition est la suivante : chaque groupe constitué doit faire une introduction à l'aide de chatGPT². Charge à chacun de produire la meilleure introduction. C'est un concours, de triche certes, mais un concours tout de même. Marion Schumm insiste sur le ton de l'humour : il faut respecter des règles strictes. Celles-ci consistent à produire un texte introductif en n'utilisant que cette IAgen, sans copier-coller, sans recherches annexes, sans améliorer les phrases, juste à partir des « demandes ». C'est aux groupes de trouver les *prompts*³ pertinents pour éviter le « vent ». Pour le dire autrement, il faut faire en sorte que ChatGPT problématise la question. Mais cela, c'est aux élèves de le découvrir par leur « dialogue » avec la machine. Et tout l'intérêt va être de comparer les productions et les processus par lesquels ils y sont parvenu·es. Le paradoxe, on le comprend aisément, étant de faire comprendre comment produire un texte de philosophie

¹ À ce propos, mes derniers essais avec ChatGPT montrent à quel point les progrès sont patents. Il peut être alors stratégique de passer par une IAgen moins performante.

² Que l'on peut utiliser de manière limitée sans compte et donc, peut-on supposer, sans siphonage de données. RGDP oblige.

³ Une remarque ici : c'est par une *succession* de « demandes » qu'on parvient à ce que la machine réponde de manière plus pertinente. Ce qui est déjà induire un usage « expert » de cette IAgen.

— pour peu qu'on admette que la dissertation en soit un — à partir d'une « discussion » avec une machine.

L'atelier

Nous avons donc nous-mêmes essayé de faire une introduction à cette dissertation avec ChatGPT, par groupe, dans le cadre d'un concours de triche¹. Ce qui nous a permis de prendre conscience de l'intérêt de la démarche proposée par Marion Schumm.

Je vais faire rapidement état des discussions au sein de mon groupe, car, tel que je l'ai perçu, il s'est passé quelque chose d'intéressant qui peut faire comprendre les difficultés de nos élèves — et peut dans le même temps éclairer nos propres pratiques. En effet, nous avons passé beaucoup de temps sur la formulation d'un *prompt* qui devait éviter une accroche « vendeuse ». Indépendamment du fait que l'on peut, à mon avis, se passer d'accroche, on en est resté à une approche formelle qui ne permettait pas de produire ce qu'on voulait, et surtout, on a mis du temps à en venir à ce qui fait le cœur de l'introduction — je passe sur le fait que nous n'étions pas d'accord sur les étapes de l'introduction, ou sur l'importance des définitions... Nous sommes arrivé·es à produire une introduction plus ou moins satisfaisante à force de questions. Mais elle était encore bien imparfaite.

La bonne introduction que j'ai pu lire, résultat d'une « discussion » pertinente avec la machine, n'est pas le produit d'un ou de plusieurs *prompts* qui indiqueraient point par point, de manière précise, ce qu'il faut faire, mais plutôt celui d'une approche plus « philosophique » si on peut dire. La succession de *prompts* a amené la machine à préciser ce qui fait difficulté, à définir les termes, mais en demandant à ce que les définitions soient prises en compte dans la problématisation, à justifier les affirmations, etc. On retrouve ainsi la perspective normative dont je parlais plus haut. Ce qui va permettre de faire émerger des règles méthodologiques relativement claires et simples. Je discute de cela en conclusion.

Je vous invite à essayer. Seul·e ou en classe. Ce peut être une séance intéressante, limitée dans le temps, qui peut être un support ou une étape dans l'apprentissage de la dissertation de philosophie.

¹ Au vu du temps contraint, Marion Schumm a proposé que les groupes qui pensent avoir une bonne introduction « challengent » les autres. — Ce qui est efficace également dans le cadre de la classe comme j'ai pu l'expérimenter. Le gain de temps est patent. Sur un plan pratico-pratique, j'ai également projeté les introductions des groupes qui pensaient avoir réussi à faire (faire) une bonne introduction. J'ai aussi demandé à tous les groupes de m'envoyer leur introduction par l'ENT (Espace Numérique de Travail) pour pouvoir comparer les productions et revenir le cas échéant sur ce qui était intéressant ou non sur le plan méthodologique. Pour compléter, je pense qu'il faudrait également demander un compte-rendu des *prompts* qui ont permis aux élèves de réussir ou non l'exercice.

Remarques conclusives

Pour finir, j'aimerais revenir sur deux points.

Le premier point concerne l'intérêt de ChatGPT comme *bouc émissaire*. Comme je l'ai rappelé plus haut, la critique de l'introduction produite par la machine est facilitée par le fait que les élèves s'autorisent à être très acerbes, alors qu'ils auraient tendance à être très mesurés dans la reprise de celle d'un·e camarade de classe. Il faut aller plus loin. Dans le concours de triche, Marion Schumm nous dit observer que les élèves considèrent que même si ils n'arrivent pas à une introduction satisfaisante, c'est ChatGPT qui n'y arrive pas : l'erreur n'est pas de leur fait, mais du fait de la machine. Alors que cela devient leur introduction quand ils y arrivent ! C'est donc un moyen très intéressant de mettre à distance l'introduction, de la regarder comme un objet qu'on peut travailler et non un prolongement d'eux/elles-mêmes.

Pour le deuxième point, j'aimerais souligner l'idée que, contrairement à ce que j'ai moi-même pu dire lors de l'atelier, le concours de triche proposé ici *n'est pas un concours de prompts*. Ce serait sortir de la finalité normative de l'exercice sur laquelle j'ai insisté plus haut. Même si, paradoxalement, ce sont bien des *prompts* auxquels répond la machine, penser l'exercice sous l'angle des seuls *prompts* ne permet pas de faire (faire) des introductions pertinentes. L'idée est d'amener les élèves à comprendre à quoi répond une bonne introduction, et non pas à produire le bon *prompt*. – Ce qui me permet de revenir sur notre propre concours de triche. La difficulté à laquelle mon groupe s'est heurté renvoie au fait que nous avons abordé l'exercice avec ChatGPT comme les élèves le font avec la dissertation elle-même. C'est-à-dire par une approche formelle, à partir d'une « recette ». C'est donc un des intérêts de l'atelier de Marion Schumm que de nous faire vivre cette approche critique d'une introduction par le concours de triche : en prenant ainsi conscience du travail que doit produire un élève en classe de philosophie pour faire faire une bonne introduction à ChatGPT, cela permet une réflexion de niveau 2, dit-elle, par laquelle nous réfléchissons aux consignes que nous donnons aux élèves. Les meilleures introductions s'intéressent plus à ce qui fait le cœur d'une introduction qu'à la production de *prompts*. Elles indiquent peut-être comment faire travailler nos élèves sur la dissertation de philosophie.

Jean-Pierre Dramisino

De quelles valeurs nos techniques en général et l'IA en particulier sont-elles porteuses ?

Adeline Barbin

Dans le cadre d'un cours sur les relations entre technique et liberté intitulé « La technique nous rend-elle libres ? », je propose aux élèves de réfléchir aux relations entre techniques et valeurs. Les techniques ne sont en effet pas neutres comme le soutient la thèse instrumentaliste qui veut que tout dépende de l'usage qu'on en fait. La deuxième partie de l'article a pour objectif d'étendre cette réflexion au cas particulier du numérique et des intelligences artificielles.

I. Techniques et valeurs

1. L'idée que nos choix techniques dépendent du critère de l'efficacité

Je m'appuie sur deux extraits d'Ellul pour présenter les idées reçues dont il faut se défaire pour penser pertinemment la technique. Le premier extrait permet de questionner la place de l'efficacité dans les techniques.

Nous voyons donc que cette double intervention [de la raison et de la conscience] dans le monde technique qui produit le phénomène technique peut se résumer comme « la recherche du meilleur moyen dans tous les domaines ». C'est ce « one best way » qui est à proprement parler le moyen technique et c'est l'accumulation de ces moyens qui donne une civilisation technique.

Le phénomène technique est donc la préoccupation de l'immense majorité des hommes de notre temps, de rechercher en toutes choses la méthode absolument la plus efficace. Car on est actuellement passé à la limite dans les deux sens. Ce n'est plus aujourd'hui le moyen *relativement* le meilleur qui compte, c'est-à-dire comparé à d'autres moyens également en action. Le choix est moins en moins affaire personnelle entre plusieurs moyens appliqués. Il s'agit en réalité de trouver le moyen supérieur dans l'absolu, c'est-à-dire en se fondant sur le calcul dans la plupart des cas.

Jacques Ellul, *La technique ou l'enjeu du siècle*, 1954, Economica, pages 18-19

1. Relevez dans le texte les expressions qui permettent d'expliquer de quelle façon nous choisissons nos techniques.
2. Que pensez-vous de cette analyse ? Aidez-vous d'exemples que vous connaissez pour y réfléchir.

Nous considérons généralement que si nous utilisons une technique plutôt qu'une autre, c'est parce que c'est la plus efficace et que serait donc là le seul critère qui expliquerait son usage. Après avoir mis en évidence la thèse d'Ellul

avec la première question, il s'agit de s'appuyer sur les exemples des élèves pour montrer que ce critère ne fonctionne pas. Si les élèves ne proposent pas d'exemples, on pourra utiliser celui du clavier Qwerty qui s'est imposé alors qu'il ne permet pas la plus grande vitesse de frappe¹.

Le second extrait doit permettre de comprendre que la place donnée au critère d'efficacité par Ellul repose sur l'affirmation que le changement technique ne dépendrait, à notre époque, que de causes techniques.

Or qu'est-ce qui détermine actuellement cette progression [de la technique] ? Ce ne sont plus ni les conditions économiques ou sociales, ni la formation intellectuelle ; ce n'est plus le facteur humain qui est déterminant, mais essentiellement la situation technique antérieure. Lorsque telle découverte a lieu, il s'ensuit presque par nécessité telles autres découvertes. L'intervention humaine dans cette succession apparaît comme occasionnelle et ce n'est plus un homme déterminé qui seul pouvait faire ce progrès, mais n'importe qui suffisamment au courant des techniques peut faire une découverte valable qui succède raisonnablement aux précédentes et qui annonce raisonnablement la suivante. [...]

Pour faciliter la vie de la ménagère, par exemple, on met en application un nouvel appareil qui broie les ordures et permet de les éliminer par l'évier. Ce procédé provoque une énorme pollution des rivières. Il faut alors chercher un moyen de purification de cette eau qui doit, en définitive, être consommée : une plus grande quantité d'oxygène sera nécessaire pour permettre aux bactéries de détruire les matières organiques. Comment oxygéner les rivières ?... Voilà comment la technique s'engendre elle-même.

Jacques Ellul, *La technique ou l'enjeu du siècle*, 1954, Economica, p. 81-85.

1. D'après Ellul, quelle est la cause du changement technique ?
2. Quelles autres causes possibles ce texte exclut-il ?
3. Que pensez-vous de cette analyse ? Aidez-vous d'exemples que vous connaissez pour y réfléchir.

2. La contingence des techniques

Cette étape du raisonnement est l'occasion d'aborder ou de réviser avec les élèves la distinction entre nécessaire et contingent. D'après Ellul, le développement technique suit un cheminement nécessaire. Divers exemples doivent permettre aux élèves de comprendre que d'autres facteurs interviennent. On leur demande pour cela de chercher ce qui explique qu'une technique ou une forme (*design*) précise de cette technique soit utilisée. Voici quelques exemples possibles.

¹ David Paul A., « Comprendre les aspects économiques de QWERTY : la contrainte de l'histoire », trad. Zeitlin E., *Réseaux*, 1998, vol. 16, n° 87, p. 9-21.

1. Les systèmes de production d'électricité

Dans les années 1920, on trouvait à Berlin une demi-douzaine d'usines de production de taille importante ; Londres en comptait plus de cinquante, de petite taille. Le pouvoir administratif en charge des systèmes électriques était centralisé à Berlin, alors qu'à Londres, c'étaient les autorités municipales de chaque quartier qui régulaient le système de production et de distribution.

Cause : administrative.

2. La stérilité des graines en agriculture

Certaines semences génétiquement modifiées vendues aux agriculteurs sont stériles : les plantes qui en sont issues ne produisent pas à leur tour de graines utilisables. Cette stérilité permet à la fois de contrôler la diffusion de ces plantes et d'assurer le respect de la propriété privée des séquences ADN, protégées par des brevets. Il faut en effet se tourner vers les industries propriétaires des semences pour en obtenir d'autres.

Cause : sécurité et inscription dans l'artefact d'un droit de propriété (qui repose sur un système juridique qui autorise à breveter des séquences ADN).

3. La machine à laver

La machine à laver a mis plus longtemps en France que dans d'autres pays à devenir un bien de consommation accessible au grand public. Les modèles étaient chers, notamment parce qu'ils étaient très lourds et qu'il fallait donc les fixer pour qu'ils ne se déplacent pas pendant l'essorage. Il s'agissait par ailleurs de modèles semi-automatiques : l'essorage ne se lançait pas automatiquement après le lavage et quelqu'un (idéalement une domestique) devait donc le lancer. En 1960, ni l'Allemagne, ni la Suisse, ni les Etats-Unis n'utilisaient plus ces modèles. La machine à laver a ainsi longtemps été en France un objet de valorisation sociale.

Cause : statut social de l'artefact.

4. À votre tour : avez-vous des idées d'exemples ?

Le bilan de l'exercice permet de revenir sur la place de l'efficacité dans le changement technique et de souligner que l'émergence d'une technique et sa forme sont conditionnées par des facteurs culturels, sociaux et politiques et non pas seulement techniques.

La technique se définit brièvement et généralement comme un agir efficace, plus précisément comme un moyen matériel ou intellectuel permettant d'obtenir volontairement et efficacement un résultat déterminé. Le problème avec cette définition est la place imprécise donnée à l'efficacité. L'efficacité n'est en effet pas une valeur absolue. La voiture, par exemple, est-elle un moyen de transport efficace ? Oui, sans doute, si on pense à la facilité de déplacement qu'elle offre ; non, si on compare l'énergie dépensée par rapport au nombre de personnes transportées.

Il nous faut donc définir plus précisément la technique : elle est un agir efficace qui prend une forme (au sens d'incarnation) culturellement organisée (au sens où elle dépend des valeurs de cette culture). Toute technique porte et exprime ainsi une conception du monde et une relation possible à celui-ci. En sociologie des techniques, on dira que toute technique est socialement façonnée (*socially-shaped*). On peut à nouveau inviter ici les élèves à proposer d'autres exemples.

3. Le déterminisme technique¹

À quel point le fait qu'une technique existe et la forme qu'elle prend ont-ils des effets sur l'organisation d'une société, en particulier sur ses valeurs ? En plus d'être socialement façonnées, elles seraient socialement « façonnantes » (*socially-shaping*) : les techniques structureraient la société. C'est la thèse du déterminisme technique : la technique serait la cause principale de l'organisation de la société.

L'exemple le plus connu de déterminisme technique strict est sans doute celui de l'étrier tel que le présente Lynn White Junior : parce qu'il permet de tenir en selle quand il y a des chocs lors des combats, l'étrier aurait permis la création de la chevalerie ; mais puisque la chevalerie nécessite un groupe humain entraîné et équipé (armures), elle nécessite une structure politique et sociale, celle du féodalisme. Donc le féodalisme serait le fruit de l'étrier.

La thèse du déterminisme technique au sens strict a été rarement défendue et son ridicule a contribué à masquer un fait : il semble bien qu'il existe un déterminisme technique au sens où les techniques ont des effets sur l'organisation de la société et des relations humaines. Le déterminisme technique dit « souple » défend l'idée que la technique est un facteur parmi d'autres. Pour le mettre en évidence, on invite les élèves à lire des exemples pour répondre à la question suivante : les techniques sont-elles neutres ou ont-elles des effets sociaux et politiques ?

1. Le groom automatique

Un groom automatique est un dispositif mécanique qui assure la fermeture automatique d'une porte. Dans nos sociétés, la politesse veut que l'on referme la porte derrière soi si elle fermée quand on arrive. Comme les humains ne sont toujours très disciplinés, on a donc délégué à un mécanisme la mise en œuvre du respect de la norme sociale en question².

¹ On distingue deux sens à l'expression : les techniques se développent de façon nécessaire indépendamment de toute intentionnalité humaine (c'est la thèse de l'autonomie de la technique qu'on trouve chez Ellul) ; les techniques sont la cause dont la structure de la société est l'effet. Nous nous référons au second sens.

² Bruno Latour, « Where are the missing masses? », in Wiebe E. Bijker and John Law (dir.), *Shaping*

2. Le travail à la chaîne

En 1872, Engels¹ juge qu'une organisation hiérarchique est nécessaire dès lors que des machines coordonnent les activités des ouvriers, comme dans le cadre du travail à la chaîne : « Prenons comme exemple une filature de coton. Pour que le coton se transforme en fil, il doit subir au moins six opérations successives et différentes qui, la plupart du temps, s'effectuent dans des locaux différents. En outre, il faut un ingénieur pour tenir les machines en marche et les surveiller, des mécaniciens, chargés des réparations courantes, et un grand nombre d'ouvriers pour le transport des produits d'un atelier à l'autre, etc. Tous ces travailleurs hommes, femmes et enfants sont obligés de commencer et de finir leur travail à des heures déterminées par l'autorité de la vapeur qui n'a cure de l'autonomie des individus. »

3. Le protocole TCP/IP

Le protocole *Transmission Control Protocol/Internet Protocol* organise la façon dont les ordinateurs transfèrent des données en ligne. Ce protocole inscrit dans l'architecture d'Internet une certaine protection de la liberté d'expression puisqu'il s'agit d'un système décentralisé et neutre : les informations sont transmises sans que le contenu en soit examiné, ni que le réseau sache qui sont l'expéditeur et le destinataire réels. Cette architecture du système en est une parmi d'autres possibles, dont certaines permettent plus de contrôle.

4. À votre tour : avez-vous des idées d'exemples ?

L'exercice doit permettre de mettre en avant le fait que la thèse de la neutralité des techniques, aussi appelée instrumentalisme, ne fonctionne pas : on ne peut pas se contenter de dire que le fait qu'une technique soit bonne ou mauvaise dépend de l'usage qu'on en fait. L'instrumentalisme considère que l'évaluation axiologique ne peut porter que sur la fin et non sur le moyen. Si on considère que les techniques sont porteuses de conceptions du monde et encadrent les relations possibles à celui-ci, alors les techniques sont porteuses de valeurs et on peut les évaluer axiologiquement. On peut donc notamment évaluer leur relation à la liberté.

Il me semble qu'il est alors nécessaire d'introduire une distinction conceptuelle si on veut réfléchir pertinemment aux relations entre technique et liberté : celle de la capacité et de la liberté, la première n'étant au mieux qu'une forme particulière de la seconde. Que les techniques augmentent nos capacités comme espèce biologique (aller plus vite, vivre plus longtemps, etc.) ne veut pas dire qu'elles augmentent nos libertés, comme groupe ou comme individu. On pourra

technology/Building Society. Studies in sociotechnical change, MIT Press, Cambridge (Mass.), 1992, p. 225-259.

¹ Engels Friedrich, « De l'autorité », *Almanacco repubblicano*, décembre 1873.

revenir sur l'exemple du travail à la chaîne : elle augmente nos capacités de production, mais pas la liberté des ouvriers et ouvrières.

4. Le code technique dominant

Je reprends l'expression de « code technique » à Andrew Feenberg qui le définit comme « la réalisation d'un intérêt ou d'une idéologie sous la forme d'une solution techniquement cohérente à un problème¹ ». Le code technique est la rencontre de la dimension proprement technique, qui exige que nos moyens permettent d'atteindre des fins et des valeurs d'une société. Dans le code technique de nos sociétés, par exemple, on trouve une forte exigence de sécurité. On y trouve également, avec l'industrialisation, la déqualification du travail qui autorise ce que Feenberg nomme « l'autonomie opérationnelle ». Elle permet « la liberté du propriétaire ou de son représentant de prendre des décisions indépendantes sur la façon de mener les affaires de l'entreprise », indépendante en particulier des ouvriers. L'autonomie opérationnelle apparaît ainsi comme une organisation juridique qui se redouble dans le développement technique.

II. Le numérique, l'IA et leurs valeurs

1. Le numérique nous rend-il libres ?

Le numérique n'est pas différent du monde physique : il encode, lui aussi, des valeurs. Nous en avons déjà croisé un exemple, celui du protocole TCP/IP. La suite de mon propos vise à proposer des pistes de réflexion pour aborder le sujet avec les élèves. On remarquera qu'il est possible à travers cette question d'aborder de nombreuses questions au programme : la justice, le travail, le libre arbitre, les libertés et la démocratie, ou encore les sciences.

Le numérique et l'IA² ne favorisent-ils pas l'émancipation des individus en mettant à leur disposition un grand nombre de savoirs et d'outils ? C'est évidemment un aspect tout à fait notable. On peut cependant se demander pourquoi ces innovations échapperaient au code technique dominant de l'autonomie opérationnelle dans la mesure où ces technologies se développent au sein du même cadre social et économique que les précédentes.

Concernant le numérique en général, l'idéologie dominante de la Silicon Valley a été largement étudiée et commentée : elle s'articule autour du libéralisme, du

¹ Andrew Feenberg, *Between reason and experience*, 2010, MIT Press, p. 68.

² Je distingue l'IA en général des *large language models*.

transhumanisme, du cornucopianisme¹ et du technosolutionnisme. L'individu y est considéré comme devant être entrepreneur d'un soi qui est un capital à faire fructifier. Pour mettre ce point en évidence, on peut s'appuyer sur des exemples d'applications : les applications de régime impliquent que la responsabilité de la santé incombe à l'individu au détriment de toute approche collective et de questionnements sur les causes sociales, par exemple, de l'obésité.

On a pu récemment voir apparaître les concepts de techno-féodalisme ou néo-féodalisme. Ils désignent non seulement la place occupée par quelques entreprises gigantesques qui concentrent des richesses immenses, mais également les importantes inégalités de pouvoir liées à la collecte des données personnelles et au contrôle de l'information qui se mettent en place, indépendamment de tout consentement individuel ou de tout contrôle démocratique.

2. Les valeurs de l'intelligence artificielle

On pourra aborder avec les élèves le phénomène aujourd'hui bien documenté des algorithmes de recommandations sur les réseaux sociaux qui n'ont pas grand-chose de neutres et favorisent la polarisation des idées. Autour du thème de la justice, on peut étudier les biais de l'algorithme Compass utilisé aux États-Unis pour prédire le risque de récidive et dont le racisme a été mis en évidence. Dans les deux cas, ces biais proviennent des humains, soit à travers le choix du fonctionnement des algorithmes, soit à travers les données d'entraînement.

On peut également se demander à quel point le développement des intelligences artificielles rejoue historiquement ce qui s'est produit avec l'invention du moteur thermique et la révolution industrielle. On retrouve, tout d'abord, la même importance du capital de départ. Ici, il ne prend pas la forme d'argent, mais de *data* (pas d'IA sans des quantités massives de données). Ces données ont pour une large part été obtenues ou utilisées en toute illégalité². Les IA paraissent également propres à entraîner la déqualification d'un certain nombre de tâches qu'elles sont capables de réaliser. Ensuite, on peut s'inquiéter de retrouver un phénomène semblable à celui qui s'est produit lors du passage de la production artisanale à la production industrielle, celui de l'homogénéisation des contenus, au détriment de l'originalité et des minorités, et au profit des contenus dominants trouvés dans les données d'entraînement. On retrouve de surcroît, dans la sphère technologique et médiatique, le même type de discours que celui croisé au XIX^e siècle, empreint de fatalisme (il ne servirait à rien de vouloir s'opposer aux transformations en cours, ni même d'espérer les orienter ou les encadrer) et mettant

¹ Le cornucopianisme (ou mythe de la corne d'abondance) désigne l'idée que le génie humain permettra toujours d'accéder à des ressources illimitées.

² Shoshana Zuboff, *L'âge du capitalisme de surveillance*, Zulma, 2022. On pourra penser également au non-respect du droit d'auteur dans les données d'entraînement.

en scène la diminution de la valeur d'un être humain désigné comme inférieur à ses créations. Enfin, il apparaît nécessaire de construire un nouvel appareil de droit de la même manière que le droit social et le droit du travail sont venus répondre à l'industrialisation. Tout cela n'empêche pas bien sûr de souligner les gains attendus (de connaissance, de productivité, etc.) et de les mettre également en parallèle avec ce qu'a apporté l'industrialisation.

Ces différents éléments posent la question de la menace pour la liberté au sens de libre arbitre et pour nos libertés individuelles. La captation de nos données personnelles s'inscrit dans un enjeu qui est celui de la connaissance et de la modification de nos comportements. Le scandale *Cambridge Analytica* par exemple a soulevé la question de la possibilité d'influencer les votes des électeurs. Le développement de l'Internet des objets, aussi appelé internet ubiquitaire, en multipliant les sources d'approvisionnement en données, prolonge ces préoccupations. De grands noms de la tech sont en faveur de la disparition de la vie privée, pourtant aujourd'hui considérée comme un support fondamental de la démocratie.

Il est enfin possible d'aborder la question des IA en s'interrogeant sur ce que sont les sciences. La quantité de données disponibles (*big data*) et la possibilité d'utiliser des systèmes d'IA pour les analyser et en dégager des corrélations a conduit certains à affirmer que les sciences allaient (et devaient) abandonner le modèle de la causalité et la construction de modèles théoriques.

Puisque le cours s'est ouvert sur la contingence des techniques, il paraît pertinent de soulever cette interrogation au sujet des IA : d'autres IA sont-elles possibles ? Comment envisager des IA qui permettent de surmonter ces difficultés ? D'un côté, certains dangers semblent bien intrinsèques au phénomène de numérisation lui-même : il facilite la surveillance et le contrôle. De l'autre, certains aspects relèvent bien de la forme, du *design*, donné à la technique. Lorsque Google explique que ses serveurs conservent les données, l'entreprise affirme quelque chose de très inexact : les serveurs de Google sont conçus pour conserver les données, ce qui est très différent. Il n'y a là rien d'une inéluctabilité technique.

« Le procès de la voiture autonome »

Théo Milana

L'atelier que j'ai présenté n'avait pas initialement été proposé par moi-même, mais le collègue qui l'avait proposé ne pouvait plus le présenter. J'ai proposé de prendre la relève, car le sujet suscitait un certain intérêt pour moi. Je n'avais jamais pratiqué de procès en classe ; j'avais donc conscience que ma proposition ne serait pas exempte de défaut, et sans doute bancal sur certains points.

J'ai donc commencé par me renseigner sur la forme du procès, notamment à partir de la présentation donnée dans le guide pratique "Enseigner la philosophie" de l'ACIREPH, avec l'article de Nicole Grataloup, complété par d'autres témoignages de cette pratique sur le groupe Enseigner la philosophie, ainsi que quelques articles. Le grand mérite de l'exercice, par rapport à d'autres exercices comme un colloque des philosophes, me semblait être dans le fait qu'il *force* les élèves à prendre une décision.

J'ai ensuite effectué des recherches sur les voitures autonomes : à la fois sur leur fonctionnement, mais aussi sur des cas de procès réels, car il y a déjà eu des accidents impliquant des voitures autonomes. Bien sûr, cette recherche a pu se montrer parfois un peu fastidieuse, car je sortais de mes compétences habituelles. Mes recherches m'ont permis notamment de réfléchir au type de voiture autonome dont il serait question - car il existe différents niveaux d'automations - ainsi que de découvrir la jurisprudence qui semble en train de se construire, et qui semble pour l'instant souvent favorable aux constructeurs.

Une première difficulté, dont on verra qu'elle n'a pas été surmontée, a été de circonscrire le sujet : de quoi devait parler le procès ? Sur quel enjeu devait-il porter ? En effet, il pouvait porter ou bien sur une attribution de responsabilité (si une voiture autonome cause un accident, qui doit être responsable : les personnes présentes dans la voiture ? le constructeur ? l'État ?), avec des questions d'ordre morales et politiques ; ou bien sur la légitimité même des voitures autonomes (est-ce une bonne chose ? devrions-nous les interdire ?) - avec des réflexions qui porteraient sans doute davantage sur le rapport à la technique.

Il me semblait nécessaire de répondre à cette question, afin de construire mon procès ; en effet, il faut bien décrire la situation qui sera à juger. Or, selon la façon dont le cas est construit, les réflexions ne seront pas les mêmes, ni les enjeux.

Je ne parvenais pas à trancher entre ces deux aspects, et j'ai fini par construire un procès qui les prenait tous les deux en compte, d'une façon sans doute un peu bancal. La situation est devenue la suivante : les voitures autonomes ont été rendues obligatoires ; Gigi Rousseau, qui déteste les voitures autonomes, doit malgré tout l'utiliser pour ses déplacements, car l'Etat l'a rendue obligatoire (pour des raisons de fluidité du trafic et de baisse de la mortalité principalement) ; elle se retrouve prise dans un accident, causé par la voiture dans laquelle elle se trouvait. La voiture avait un niveau d'automation de 4, c'est-à-dire : elle était censée être autonome, mais le conducteur pouvait reprendre la main à tout moment. Gigi, distraite dans la voiture, n'a pas pu empêcher l'accident de se produire. La société Autovroum a donc tenté de remettre la faute sur Gigi. Cette dernière a choisi de profiter du procès qui lui est fait pour non seulement se défendre, mais remettre en cause l'existence même des voitures autonomes.

Ainsi, le procès avait un double enjeu : 1. Déterminer qui était responsable 2. Déterminer si les voitures autonomes étaient une bonne chose pour l'homme, ou si on devait les bannir. Je m'étais inspiré des *Animaux dénaturés* de Vercors, dans l'idée d'un procès au double enjeu (à la fois question d'attribution de responsabilité, qui se mêle à une question plus large et essentielle).

La deuxième difficulté a été ensuite de construire un dispositif approprié pour que les élèves puissent travailler. Après avoir construit la situation qui devait être jugée, j'ai réfléchi aux différents groupes selon lesquels répartir les élèves. J'ai opté pour la solution suivante :

- Un groupe de juges, qui devrait trancher le procès.
- Un groupe d'avocats de Gigi, qui devrait à la fois défendre Gigi, et obtenir l'interdiction des voitures autonomes
- Un groupe de représentants d'Autovroum, qui devrait faire condamner Gigi, et défendre l'usage des voitures autonomes.

Ces trois groupes étaient donc centraux pour le déroulé du procès, et possédaient un dossier commun. Ce dossier rappelait les faits ; contenait une définition de la voiture autonome ; des définitions juridiques de la responsabilité ; un texte d'Aristote et de Hume sur la notion de responsabilité.

« Le procès de la voiture autonome »

Mais je craignais que trop d'élèves ne soient alors pas impliqués. C'est pour cela que j'ai choisi de construire en plus trois groupes experts. Ces groupes auraient accès à des dossiers différents, et devraient se construire une expertise sur leur sujet ; les juges, avocats et représentants devaient aller librement les consulter durant l'atelier :

- Un groupe était expert en intelligence artificielle.
- Un groupe est expert sur le fonctionnement des voitures autonomes.
- Un groupe est un comité d'éthique, expert sur les rapports en morale et intelligence artificielle.

Il me semblait en effet, pour comparer cela à la façon dont le procès d'Antigone est construit, que ce dernier faisait intervenir une question finalement assez intuitive pour les élèves, ou du moins pour laquelle leur intuition était suffisante pour fournir déjà des bases de réflexions intéressantes. Or, il me semblait que sur un sujet tel qu'un procès de voiture autonome, il était nécessaire d'avoir des connaissances concernant le fonctionnement de ces dernières.

Le but était donc de prévoir un temps de préparation du procès : les juges, avocats et représentants travaillant les éléments du dossier, avant de se renseigner auprès des experts. Puis le procès s'effectuerait, avec les avocats et représentants échangeant des arguments, et pouvant demander aux experts de témoigner.

Déroulement effectif du procès durant les ateliers :

J'ai eu l'occasion de présenter deux fois l'atelier aux journées d'étude de l'Acireph, à deux groupes distincts, dans des conditions un peu différentes, puisque le deuxième groupe était moins nombreux. N'ayant pas pensé à demander une salle adaptée aux déplacements requis pour l'atelier, cela a un peu rendu difficile d'aller consulter les experts.

Les collègues m'ont demandé plusieurs précisions sur le procès, notamment sur la jurisprudence actuelle. Il semble que la forme procès peut entraîner une réponse qui se contenterait d'être purement dans les bornes du droit, sans forcément faire un effort réflexif sur le bien-fondé desdites lois - ce qui était un effet que je craignais quelque peu. J'ai donc pensé préférable de "tricher" quelque peu, en modifiant la consigne : le procès devait faire jurisprudence ; il devait partir d'un flou juridique, et ne pouvait donc pas entièrement se fonder sur une législation préexistante.

Si les procès ont globalement pu se dérouler, les collègues ont fait de nombreuses remarques, qui me semblent devoir pousser à revoir l'exercice :

1. Le double enjeu du procès risquerait sans doute de rendre l'exercice confus. En effet, le procès interroge à la fois sur un problème de responsabilité, et sur un problème lié à la technique. Or, les deux enjeux ne paraissent, en tout cas tel que l'exercice a été présenté, pas très organiquement liés. Plusieurs collègues pensaient qu'il aurait été plus opportun de ne garder qu'un des deux enjeux, en amendant dès lors le cas proposé aux élèves. Il ne me semble pas qu'un consensus s'est dégagé sur quel aspect était à privilégier.

2. Les dossiers fournis aux élèves étaient trop fournis. En effet, cela rendait difficile d'imaginer qu'ils pourraient faire leur préparation dans un temps raisonnable. Une solution s'est présentée d'elle-même me semble-t-il : en effet, les textes philosophiques étaient assez souvent difficilement mobilisables d'une manière directe par les élèves, car ils abordent le problème de biais. Il conviendrait alors de supprimer ces textes : on gagnerait sur le temps de préparation.

3. Certains rôles semblaient plus difficiles à intégrer. Par exemple, les experts en Intelligence Artificiel étaient moins consultés.

Ainsi, il me semble que si les collègues n'ont pas passé un mauvais moment, et s'il y a pu y avoir des moments d'échanges intéressants, cela doit peut-être plus au public auquel s'est, de fait, adressé l'exercice, qu'à la construction de ce dernier. En l'état, il est assez bancal. Pour le rendre praticable en classe, il conviendrait sans doute de trancher sur ce qui doit être traité (problème portant sur la responsabilité, ou problème portant sur l'automatisation et la technique), et sans doute faire un peu plus confiance aux élèves, en leur donnant moins d'éléments. Si des éléments techniques me semblent toujours nécessaires, on pourrait sans dommage supprimer les textes philosophiques.

Une autre piste qui a été soumise était de changer la forme de l'exercice. Ce qui me semblait intéressant dans la forme du procès, c'était de forcer les élèves à trancher et prendre une décision. Or, on peut arriver à cet objectif en procédant différemment. Il a été suggéré d'imaginer le dispositif suivant : l'État devrait choisir de rendre obligatoire ou non la voiture autonome. On pourrait à partir de cette situation imaginer un grand nombre d'interlocuteurs différents (comité d'éthique, constructeur de voiture, ingénieurs, représentants de victimes, etc...). Cette piste me semble intéressante, pour conserver la prise de décision, tout en centrant le sujet d'une façon peut être plus simple que selon un enjeu de responsabilité.

« L' "intelligence artificielle" est un terme idéologique »

Entretien avec Anne Alombert

« L' "intelligence artificielle" est un terme idéologique »

Cet article retrace les échanges que nous avons eus avec Anne Alombert lors d'un atelier aux Journées d'Études de novembre 2024 sur l'intelligence artificielle. Il a été rédigé par Fanny Bernard qui a animé l'atelier, en collaboration avec Maxime Volta et Ana Ilić qui y ont participé. Anne Alombert est maîtresse de conférence à Paris VIII et spécialiste des technologies numériques. Elle a relu cette retranscription de nos échanges.

*Fanny Bernard avait proposé un nouveau dispositif d'atelier : le premier jour, le groupe a « arpenté » l'essai d'Anne Alombert *Schizophrénie numérique*¹ et en a dégagé des questions. Puis le deuxième jour, le groupe a échangé avec la philosophe à partir de sa lecture.*

Les professeur·es ayant participé à l'atelier : *Comment en es-tu venue à ce sujet : l'intelligence artificielle, la technologie numérique ? Pourquoi ce livre ?*

Anne Alombert : J'ai rédigé une thèse² qui confrontait les pensées de Simondon et Derrida au sujet de la technique et de l'écriture. Simondon et Derrida m'ont permis de penser en dehors des schémas dualistes qui opposent nature et technique, le premier à travers sa critique du « substantialisme » et sa philosophie relationnelle et le second par sa critique du « logocentrisme ». Pendant ma thèse, j'ai aussi travaillé avec l'association de Bernard Stiegler « Ars industrialis » sur les technologies numériques : Stiegler montre que les évolutions technologiques ne sont pas des évolutions industrielles seulement mais aussi des évolutions de nos esprits, de notre conscience et de nos expériences, du rapport à l'espace et au temps. Cette association développe des projets pratiques : ces évolutions technologiques ont des effets prolétarisants, c'est-à-dire qu'elles dépossèdent les individus de leurs capacités et de leurs savoirs. L'association cherchait au contraire à développer des technologies contributives et déprolétarisantes à travers des projets de recherche-action. Puis, j'ai fait partie du Conseil National du Numérique, et dans ce cadre, j'ai participé à la rédaction d'un rapport sur l'économie de l'attention, qui étudiait les fonctionnements et les modèles d'affaires des réseaux sociaux commerciaux et qui proposait des alternatives face à la captation de

¹ *Schizophrénie numérique*, Anne Alombert, Allia, 2023.

² Le livre tiré de thèse est [disponible en ligne](#), :Anne Alombert, *Penser l'humain et la technique*, ENS Éditions, 2023.

l'attention et la collecte des données, qui ont des effets nocifs pour nos capacités psychiques et sociales.

Avec *Schizophrénie Numérique*, je cherchais à témoigner et à poursuivre le travail mené par l'association Ars Industrialis et le CNNum, en ouvrant des pistes pour transformer les technologies numériques contemporaines, dans une perspective plus concrète que ma thèse qui reste très théorique. L'essai a été retenu dans le cadre du prix du livre philosophique de l'APPEP pour être lu par des élèves de lycée.

Les professeur·es : On se demandait d'ailleurs comment le livre a été reçu par des élèves de terminale, car il ne nous a pas paru si facile à lire...il y a beaucoup de concepts abstraits pour nos élèves qui découvrent la philosophie. Nous nous sommes interrogés sur ce que tu appelles « l'industrialisation des esprits » et « l'automatisation des esprits ».

A.A. : En effet, les élèves m'ont parfois dit que j'utilisais des mots trop compliqués ! Mais la thématique leur parlait beaucoup. Ils me demandaient souvent s'il ne fallait plus utiliser de smartphone par exemple. J'essayais de les amener à ne pas rester sur des pistes individuelles, mais plutôt collectives, en analysant les modèles économiques et les idéologies politiques sous-jacents aux dispositifs technologiques dominant, plutôt qu'en essayant de solutionner un dilemme moral. Je voulais les pousser à réfléchir au fait que le problème est plus politique qu'éthique.

Les professeur·es : Dans notre arpentage de ton essai et pendant ces JE, on s'est beaucoup questionnés sur l'usage que l'on peut faire de l'IA en cours. Quel traitement est possible et souhaitable en cours de philosophie de l'IA selon toi ? Tu parles d'histoire et d'anthropologie des techniques et de développement de l'esprit critique concernant les technologies. Que fais-tu, toi en tant que professeure, à l'université ? Y a-t-il des usages pédagogiques et éthiques de l'IA en cours de philosophie ? Peut-elle développer notre intelligence ? Si ce n'est pas le cas, quelles sont les autres pistes ? Et as-tu fait des recommandations pour des politiques publiques, peux-tu nous les expliquer ?

A.A. : Selon moi, il faut introduire des cours sur l'histoire des techniques et des cours d'anthropologie des techniques. Simondon parle de « culture technique » : aujourd'hui, une culture technique numérique serait indispensable pour permettre aux futurs citoyens de comprendre leurs environnements technologiques quotidiens. Il ne suffit pas d'utiliser un outil technologique pour le comprendre. Les jeunes savent très bien se servir de tous les outils numériques, notamment du smartphone, mais ils ne comprennent pas forcément leurs enjeux politiques, sociaux, culturels. Il faut développer une conscience historique de la façon dont la

« L' "intelligence artificielle" est un terme idéologique »

technologie a évolué. Le fait d'adopter une perspective de sciences humaines pour penser les techniques permettrait aussi d'être moins soumis et plus critique face à ces dispositifs.

Les professeur·es : On connaît mal nous-même l'histoire des techniques, et notamment de l'IA. Peux-tu nous expliquer la généalogie de ce dispositif technologique et ses enjeux philosophiques ?

A.A. : Simondon écrit dans les années 1960 après la conférence sur l'Intelligence Artificielle à Darmouth en 1956. Il déconstruit le paradigme behavioriste et comportementaliste qui a nourri l'IA¹, les sciences cognitives et la cybernétique. Il faut faire l'histoire du rapport entre le néolibéralisme économique et une certaine conception de la psychologie humaine. Herbert Simon est à la fois lauréat du prix Turing et le prix Nobel d'économie en 1975 et développe la théorie de la rationalité limitée, qui est à l'origine de l'économie comportementale. Il est considéré comme l'un des contributeurs principaux au développement de l'Intelligence Artificielle.

La conception de l'être humain véhiculée par ces disciplines est en effet problématique et les spécialistes de l'IA intériorisent cette façon de penser l'être humain sans la réinterroger, en comparant l'organisme à une machine ou le cerveau à un ordinateur, par exemple, comme dans l'idéologie transhumaniste².

Je cherche pour ma part à réinscrire les technologies numériques et l'intelligence artificielle dans une généalogie des supports de symboles. Ainsi, Jack Goody dans *La raison graphique*³ montre que les techniques conditionnent la pensée. Il faut donc s'interroger sur la manière dont les supports numériques affectent nos esprits.

Outre cette mise en perspective historique, on peut créer avec les élèves des contenus intéressants avec des technologies numériques alternatives, comme en leur proposant de pratiquer des plateformes collaboratives ou d'enregistrer des podcasts par exemple. Je n'ai moi-même pas eu le temps de développer beaucoup de projets de ce type, mais je sais que beaucoup de profs le font déjà.

¹ Voir l'enregistrement de la conférence de Bernard Victorri sur [le site de l'ACIREPh](#).

² Raymond Kurzweil par exemple est un des fondateurs du mouvement transhumaniste et il a été directeur de l'ingénierie chez Google. Il parle de « machines spirituelles » (Ndlr).

³ Je me permets de recopier la quatrième de couverture de cet essai peu connu des professeurs de philosophie : « Les recherches de Jack Goody sur les cultures orales et écrites ont profondément contribué à faire prendre conscience que la langue est un mode de production du social et du cognitif. Tout changement dans les moyens de communication introduit ainsi des modifications majeures dans la nature et la distribution des pouvoirs et des savoirs (papiers d'identité, billets de banque, billets doux...). » (Ndlr)

Les professeur·es : *Dans ton essai, tu mets en valeur les technologies contributives contre les algorithmes de recommandation des GAFAM, peux-tu nous dire pourquoi ?*

A.A. : Il y a deux paradigmes des technologies numériques.

- La « simulation », c'est-à-dire les technologies qui cherchent à simuler et à remplacer les activités, le modèle en est l'automate et cela suppose une délégation de nos facultés à la machine. Par exemple, avec « ChatGPT », nous déléguons nos activités d'écriture ou d'expression.
- La « collaboration », c'est-à-dire les technologies qui mettent en avant la mise en commun et la contribution collective. Par exemple, l'encyclopédie collaborative Wikipédia, qui permet la construction collective de savoirs ou des plateformes qui permettent l'annotation collective d'une vidéo, etc.

Il existe aussi des algorithmes de recommandation collaborative : ces algorithmes se fondent sur des jugements de groupe de pairs, les individus votent en fonction de critères et les algorithmes de recommandation se fondent sur ces votes. Par exemple, les critères peuvent être : clarté du contenu, utilité publique de la vidéo, fiabilité des sources, etc. Cela valorise la vidéo en fonction des avis des utilisateurs, et non des intérêts de la plateforme. C'est ce que propose le projet Tournesol¹. Les algorithmes qui reposent sur des statistiques, comme sur X ou Facebook, créent au contraire une homogénéisation et un nivellement des contenus proposés.

Les professeur·es : *Tu défends l'idée qu'il ne faut finalement pas appeler les grands modèles linguistiques (Large Language Model, LLM) des « intelligences » artificielles. Pourquoi ces technologies ne relèvent-elles pas de l'intelligence selon toi ? Y a-t-il un véritable enjeu dans cette appellation ?*

A.A. : L'« intelligence artificielle » est un terme promotionnel inventé par John McCarthy à la conférence de Dartmouth en 1956 pour récupérer des financements mais ce n'est pas un terme scientifique. L'idée est qu'on peut décrire objectivement l'esprit de l'humain et le simuler avec une machine, ce qui conduit à une automatisation des facultés cognitives. Avec la notion d'intelligence artificielle,

¹ Le projet est ainsi présenté sur leur site internet : « *Tournesol est un projet de recherche participatif mené par l'Association Tournesol dont le but est de faire avancer la recherche dans l'éthique des algorithmes et des systèmes de recommandation. Pour atteindre cet objectif, nous développons la plateforme Tournesol. Tous.les les contributeur.trices y sont invité.es à construire ensemble une grande base de données ouverte de jugements sur la qualité des vidéos. L'effet immédiat de cette base de données est d'améliorer les recommandations de milliers de personnes utilisant la plateforme. L'effet à long terme est de faciliter et d'améliorer les futurs projets de recherche en mettant à disposition cette base de données.* » (Ndlr)

« L' "intelligence artificielle" est un terme idéologique »

on projette une faculté psychique ou cognitive (l'intelligente) dans les machines, alors qu'il s'agit d'une automatisation des facultés de pensée.

Il y a eu deux façons¹ de concevoir l'intelligence artificielle :

- Le paradigme symbolique : un programme déterminé à l'avance formalise et simule logiquement et informatiquement les processus du raisonnement logique (modélisation de la pensée humaine)
- Le paradigme connexionniste : ce sont des calculs statistiques sur des quantités massives de données qui s'inspirent du fonctionnement des neurones biologiques (modélisation du cerveau humain)

Avec l'internet des plateformes, on peut récupérer beaucoup de données et l'IA connexionniste est devenue très performante. En même temps, il y a eu des progrès dans le champ de la micro-électronique qui ont permis de produire des processeurs spécifiques qui effectuent des calculs complexes.

Les professeur·es : Mais quels sont les enjeux de l'IA connexionniste ?

A.A. : Il y a d'abord des enjeux écologiques dans le développement de ces systèmes : la nécessité de construire des *data centers* crée déjà de nombreux conflits. En Uruguay, il y a des conflits d'usages sur l'eau potable et l'électricité avec la construction par Google d'un *data center*, et en Espagne, c'est le plus grand data center de Meta qui crée des conflits d'usage des ressources en eau dans une région souvent touchée par la sécheresse. Ce sont des dispositifs qui ne sont pas vertueux sur le long terme !

Il y a aussi des enjeux psycho-sociaux. Ce sont des nouveaux savoirs et capacités qui sont extériorisés à travers ces dispositifs technologiques. Ce sont des capacités psychiques qui sont extériorisées, comme la mémoire, l'attention, la réflexion, l'expression. Avec *ChatGPT*, on assiste à l'extériorisation de la capacité à s'exprimer. Le calcul prédictif a permis la recommandation algorithmique et la génération de textes et d'images. Frédéric Kaplan parle de « capitalisme linguistique »². En effet, Google vend aux enchères chaque mot qui est recherché dans le moteur de recherche. La langue est commerciale. Cela nous invite à nous conformer à du plus probable dans notre façon de nous exprimer. De même, *ChatGPT* génère des séquences de mots probables et uniformise nos capacités expressives. Il y a un effet prescriptif des calculs prédictifs sur notre esprit. On assiste à une standardisation des pensées : c'est cela que j'appelle « l'industrialisation des esprits ».

¹ Pour mieux comprendre ce point, voir l'enregistrement de la conférence de Bernard Victorri aux JE sur l'IA, sur [le site de l'acireph](#).

² Frederic Kaplan, « Quand les mots valent de l'or - vers le capitalisme linguistique », Novembre 2011, *Le Monde Diplomatique*

Les professeur·es : Mais la pensée n'a-t-elle pas toujours été standardisée ? Ça ne date pas de l'IA ! N'y a-t-il pas quelque chose de véritablement nouveau dans l'IA ?

A.A. : Le problème est d'utiliser ce type de technologie dans l'apprentissage au moment où se forment nos capacités. L'IA est utilisée tout le temps pour tout. C'est une habitude quotidienne permanente dans tous les champs de l'existence dès qu'il faut écrire : une dissertation, un mail, etc. Le risque est de ne pas développer nos propres capacités si nous les déléguons trop tôt à ces systèmes.

Les professeur·es : Mais est-ce qu'il y a une réappropriation possible de ces outils par les élèves ? Dans d'autres ateliers lors de ces JE, des collègues ont proposé des utilisations en classe de l'IA avec les élèves. Y a-t-il des usages subversifs possibles de ce système ou est-ce absolument impossible ?

A.A. : Il y a des usages subversifs, par exemple, créer un modèle de langage avec les élèves pour leur faire comprendre les enjeux de l'IA, mais je propose une troisième alternative : ni utiliser l'IA générative, ni l'utiliser de façon subversive, mais utiliser d'autres dispositifs numériques. Actuellement, il y a un problème de technodiversité qui fait qu'on est centrés sur des dispositifs technologiques qui relèvent seulement de l'IA générative. Or, il y a des technologies numériques qui ne provoquent pas une prolétarianisation et une industrialisation des esprits. Par exemple : *Wikipédia*, *Mastodon*, *Tournesol*, *Pol.is*... *Wikipédia* est un système d'écriture collaborative : c'est une organisation à but non lucratif financée principalement par des dons et animée par des bénévoles, qui permet aux individus de construire des articles ensemble. *Mastodon* est un réseau social auto-hébergé, libre, distribué et décentralisé dans lequel il n'y a pas de publicité ciblée et de recommandation automatique. *Pol.is* est une plateforme de délibération citoyenne qui valorise le consensus et non la concurrence ou la polarisation comme sur les réseaux sociaux commerciaux tels *X* ou *Facebook*.

Toutes ces technologies pourraient donner lieu à des projets pédagogiques passionnants. Il est intéressant de diversifier nos imaginaires technologiques, surtout avec les jeunes. On pourrait aussi demander aux élèves d'imaginer et de concevoir le réseau social ou le dispositif numérique de leurs rêves.

Le design de l'outil conditionne son utilisation, par conséquent la technologie n'est pas neutre. Par exemple, le fonctionnement des réseaux sociaux qui récompensent les producteurs de contenu en fonction du nombre de vues est complètement différent du fonctionnement des forums internet qui proposent un regroupement autour de projets ou d'intérêts communs. Dans le champ de l'éducation, il faut pratiquer les technologies qui favorisent le commun et la capacitation, et non les technologies commerciales prolétarianisantes.

« L' "intelligence artificielle" est un terme idéologique »

Les professeur·es : *On critique souvent l'IA pour ces coûts et enjeux environnementaux, quelles sont les limites matérielles au développement de l'IA ?*

AA : Avec Gaël Giraud, dans *Le capital que je ne suis pas*, nous avons voulu décrire les conditions matérielles nécessaires à l'intelligence artificielle. En tant qu'économiste, Gaël Giraud montre qu'à court terme, le développement de l'intelligence artificielle n'est pas soutenable. Il s'interroge aussi sur la possibilité de fixer des limites aux entreprises de la Tech. D'autre part, un article de la revue *Nature*¹ évoque aussi l'effondrement de la pertinence des modèles de l'IA générative qui s'entraînent aujourd'hui massivement sur des données générées automatiquement. L'une des limites au développement de l'IA générative pourrait résider dans le fait qu'à force de s'entraîner sur des contenus déjà automatiquement générés, les modèles produiraient des contenus de moins bonne qualité.

Les professeur·es : *Tu as travaillé pour le Conseil National du Numérique sur plusieurs rapports² . Quels sont les recommandations de politiques publiques que tu as pu faire ? Quel rôle peut jouer le droit face à l'IA ?*

A.A. : L'Union Européenne a produit récemment un *IA Act* qui introduit un cadre réglementaire et juridique commun pour l'Union Européenne. La Quadrature du Net a produit un article³ sur l'*IA Act* dans lequel l'association montre que ce texte est présenté au départ comme un instrument de protection des droits et libertés face aux risques systémiques de l'IA, mais qu'à l'arrivée, le texte est taillé sur mesure pour les géants du numérique. De mon côté, je pense que le texte ne va pas assez loin dans la mesure où il permet de lutter contre les « usages malveillants » de l'IA générative, mais je pense aussi que même dans le cadre des usages quotidiens, les systèmes d'IA génératives produisent des effets psychiques et sociaux et qu'il est nécessaire de protéger les citoyens. La technologie n'est jamais neutre, il y a donc une responsabilité des designers.

Par exemple, Facebook est responsable de ce qui est vu ou pas vu par les utilisateur·ices, à travers leur algorithme de recommandation. Le règlement européen sur les services numériques exige la transparence des algorithmes de recommandation des réseaux sociaux commerciaux. Il serait possible d'exiger aussi la

¹ « AI models collapse when trained on recursively generated data », *Nature*, 24/07/2024. [Disponible en ligne]

² On peut les retrouver en ligne : « Assurer nos libertés à l'ère de l'intelligence artificielle ». [Disponible en ligne] ; « Cultiver la richesse des réseaux : contribution aux États Généraux de l'Information ». [Disponible en ligne] ; « Votre attention s'il vous plaît ! Quels leviers face à l'économie de l'attention ? », [Disponible en ligne] ; « Pour un numérique au service des savoirs : de l'informatisation à la capacitation » [Disponible en ligne] (Ndlr)

³ À lire en ligne : <https://www.laquadrature.net/2024/05/22/le-reglement-ia-adopte-la-fuite-en-avant-techno-solutionniste-peut-se-poursuivre/> (Ndlr)

transparence des algorithmes des IA génératives, pour comprendre les biais idéologiques. Mais je crois qu'il faut aller plus loin que la transparence.

Dans le cadre du Conseil National du Numérique, j'ai défendu l'idée d'un « pluralisme algorithmique » pour produire de la diversité dans les contenus valorisés sur les réseaux sociaux, en mobilisant plusieurs algorithmes de recommandation différents et transparents. Je pense que les systèmes d'IA génératives devraient être diversifiés aussi, et que nous n'avons pas besoin de grands modèles de langage très coûteux écologiquement et qui exploitent des quantités massives de données. Au contraire, nous pourrions valoriser des petits modèles locaux et frugaux à usages spécifiques, qui pourraient être beaucoup plus fiables et efficaces.

Pour prolonger la réflexion, on pourra lire le nouvel ouvrage d'Anne Alombert paru juste après la rédaction de cet article : De la bêtise artificielle, Allia, 2025.

IA génératives : faut-il sauver l'école ou les élèves ?

Sylvain Theulle

L'immense majorité des professeurs de philosophie a pris conscience depuis assez longtemps que les IA génératives (chat-GPT, Claude, Deepseek, etc.) sont très efficaces pour tricher lors des devoirs à la maison. Mais nous découvrons désormais qu'elles se révèlent tout aussi redoutables pour la fraude pendant les devoirs surveillés en classe.

Que les élèves se fassent aider, voire trichent en copiant des manuels ou des corrigés tout faits lors des devoirs maison, n'a certes rien de nouveau. Mais les IA génératives compliquent considérablement la détection de la fraude : chaque production étant inédite, il n'existe plus d'original auquel confronter un texte suspect. Les élèves peuvent ainsi exploiter cette nouveauté technique pour contester toute accusation de plagiat.

Quant aux devoirs en classe, la triche est également devenue plus facile. Certes, il était déjà possible de dissimuler un smartphone dans sa trousse ou dans ses poches, mais cette stratégie supposait de consulter des sites internet susceptibles de fournir une réponse pertinente au sujet ; tâche fastidieuse et risquée, en raison du temps nécessaire pour effectuer la recherche, lire les résultats, etc. Les IA génératives changent la donne : une seule application peut répondre à toutes les questions. Et comme pour les devoirs maison, le caractère original de chaque production rend la preuve de la fraude difficile, sauf à surprendre l'élève en flagrant délit.

L'arrivée des IA génératives ne constitue pas une différence de nature par rapport aux pratiques antérieures, mais d'abord une différence de degré : tricher est devenu très facile, les copies produites avec ces outils, sans être excellentes au point d'éveiller les doutes, sont d'une qualité suffisante ; et les élèves savent que les probabilités de contester avec succès une accusation de fraude sont assez élevées. Tout concourt donc à favoriser la généralisation de la triche, devenue massive et permanente.

Certes, il est difficile de mesurer précisément un tel phénomène. Mes propres observations, et les échanges que j'ai pu avoir avec les enseignants d'autres disciplines, m'amènent à penser que, dans chaque devoir surveillé en classe d'un lycée, au moins un élève tente de frauder en utilisant ces outils. Toutes les

disciplines littéraires sont concernées, car ces IA génératives sont capables de produire des textes de bonne qualité sur n'importe quel sujet, et sont très pratiques sur les sujets reposant sur une question courte et ouverte, comme les dissertations.

Dans cet article, je voudrais m'opposer à deux réactions fréquentes au problème de la triche avec les IA. La première revient à nier la nouveauté et la gravité du problème et à se contenter de prescrire des consignes de bon sens aux surveillants. Cette réaction, conservatrice, est souvent celle des directions d'établissements, et elle peut s'expliquer par le fait que celles-ci n'ont connaissance que des cas de fraude qui ont été détectés et qui leur ont été communiqués. La seconde revient à adopter une posture de retrait consistant à abandonner les exercices traditionnels pour leur substituer de nouveaux exercices, pour lesquels les IA seraient moins pertinentes. Cette réaction, pessimiste sans être conservatrice, est souvent adoptée par les professeurs, et bien qu'elle puisse aussi être légitimement motivée par l'intérêt intrinsèque de ces nouveaux exercices, elle l'est souvent par le constat que les exercices traditionnels sont dorénavant irréalisables en classe. Ces deux réactions me semblent insatisfaisantes, la première parce qu'elle néglige l'ampleur de la fraude, la seconde parce qu'elle conduit à renoncer à des exercices qu'elle estime intellectuellement formateurs. Plus positivement, je voudrais défendre dans cet article qu'il y a de bonnes raisons de lutter beaucoup plus sévèrement contre la triche avec l'IA, et que la sanction, loin d'être liberticide, est ici la condition de la justice scolaire et de l'autonomie intellectuelle des élèves.

Il y a de bonnes raisons de s'inquiéter de la fraude aux examens devenue massive. La première tient au fait que l'école est une institution qui assure l'évaluation, le classement, et l'orientation des élèves. Cette fonction suppose que les notes soient un indice fiable du niveau des élèves. Si tel n'est plus le cas, c'est l'ensemble du système de classement qui s'effondre : les notes des élèves tricheurs n'étant plus représentatives, celles des élèves honnêtes ne le sont pas davantage, dans la mesure où les notes sont essentiellement comparatives. La fraude pose donc un problème classique d'action collective : le tricheur est un passager clandestin qui profite de la fiabilité du système des notes sans en respecter les contraintes. La généralisation des passagers clandestins mène à l'effondrement de l'institution qu'ils parasitent.

Toute institution scolaire ainsi comprise doit donc lutter contre eux. Or, les IA génératives sont si simples d'usage que le risque de voir se multiplier les passagers clandestins devient très crédible. Lorsqu'un délit devient trop fréquent car trop difficilement détectable et punissable, le seul levier possible reste d'alourdir la

IA génératives : faut-il sauver l'école ou les élèves ?

sanction⁷. Les agents font des calculs de coûts et de bénéfices, et s'ils savent qu'ils ont peu de chances d'être pris sur le fait, que les gains de la fraude sont élevés, et que les coûts encourus en cas de sanction sont faibles, alors ils ont tout intérêt à tricher. Et de fait, ils le font : les élèves trichent énormément, à tous les niveaux de scolarité.

Parmi les institutions disciplinaires au sens de Foucault, l'école semble de loin la plus ouverte et tolérante. Les punitions y sont assez rares et généralement peu sévères ; les professeurs ne se voient pas comme des gardiens de prison mais comme des éducateurs soucieux d'émanciper les élèves ; sanctionner est pour eux l'ultime recours quand les autres moyens ont échoué ; ils pensent qu'il faut éduquer justement pour éviter d'avoir à réprimer. Les exclusions pour fraude sont donc très rares, de même que les mentions explicites sur les bulletins et livrets scolaires. D'ordinaire, la sanction se limite à un avertissement du chef d'établissement et à l'attribution d'un zéro au devoir concerné, alors que ce zéro, pris dans une moyenne constituée de nombreuses notes, ne fait baisser cette moyenne que de façon marginale. D'autant que, dans un contexte où la fiabilité des notes est déjà largement questionnée, de faibles variations ne sont guère interprétées comme significatives.

D'autre part, signaler à l'enseignement supérieur qu'un élève a de mauvaises notes, en suggérant que son niveau scolaire est faible, au lieu d'indiquer explicitement qu'il a triché, pose un problème éthique évident. Le secondaire participe à couvrir la triche de ses élèves, à leur assurer des débouchés corrects, et à reporter le coût de la triche sur le supérieur, au lieu de faire elle-même ce travail de lutte contre la fraude. En résumé, l'école lutte mollement contre les tricheurs, ce qui n'est pas très dissuasif ; et maintenant que les IA génératives existent, les bénéfices de la fraude augmentent, alors que les coûts sont restés les mêmes : il n'est donc pas surprenant que la fraude augmente.

Il est regrettable de constater que ce combat pour la fiabilité des notes n'est pas nouveau, qu'il aurait dû être mené depuis longtemps, et qu'il ne l'a jamais été sérieusement. Tout le monde sait que la fraude est très répandue, ce qui affaiblit à la fois la valeur des diplômes et l'équité du système. L'efficacité et la justice du système sont remises en cause. L'efficacité du système est compromise puisqu'on ne forme plus les élèves les plus qualifiés pour la formation qu'ils ont choisie. Et la justice l'est tout autant puisque les élèves évincés peuvent légitimement se plaindre d'avoir perdu une place qui leur revenait de droit.

D'un point de vue humain, c'est un scandale considérable. Les élèves mettent beaucoup d'espoir dans leur formation scolaire, leurs parents également. On ne compte plus les élèves si stressés qu'ils contestent les notes, demandent des devoirs bonus, ou même incitent leurs parents à intervenir auprès des professeurs. L'injustice est massive, manifeste, facile à identifier et à combattre, et pourtant

nous ne faisons rien. Les textes réglementaires relatifs au baccalauréat prévoient des sanctions théoriquement très sévères (jusqu'à cinq ans d'interdiction de se présenter à l'examen), mais, lorsqu'il s'agit des devoirs en classe, dont beaucoup comptent pour le contrôle continu, le laxisme est de mise. Je comprends et partage l'envie de sanctionner le moins possible, d'accorder une seconde chance, mais ce souci de bienveillance peut aussi conduire à l'injustice.

Il ne m'appartient pas, pas plus qu'à l'ACIREPh, de déterminer quels sont les meilleurs moyens techniques de lutter contre la fraude. En revanche, il me semble que nous devrions militer pour la création de comités d'éthique au sein des lycées et des académies, chargés d'observer les pratiques et de formuler des propositions d'amélioration. Les élèves sont extrêmement sensibles à ces enjeux de justice, ils vivent très mal de voir leurs camarades obtenir de meilleurs résultats qu'eux en trichant. Cette demande de justice doit être prise au sérieux. Car un dilemme social nous guette : nous hésitons à sanctionner nos élèves sévèrement si les autres établissements ne le font pas aussi. Ces comités d'éthique pourraient précisément avoir pour mission de réfléchir à cette difficulté et de proposer des solutions collectives. Il n'est pas nécessaire de créer de nouvelles structures *ex nihilo* : les conseils déjà existants (conseil de la vie lycéenne, conseil de discipline, etc.) pourraient se voir confier des missions élargies.

À titre de simple suggestion, pourquoi ne pas investir dans des solutions technologiques plus performantes ? Pourquoi se contenter de surveillants humains passant dans les rangs, alors qu'il existe de nombreuses solutions plus efficaces ? Les caméras sont désormais dotées d'algorithmes d'intelligence artificielle qui peuvent être entraînées pour détecter des comportements suspects chez les élèves. Une caméra filmant l'ensemble d'une pièce serait assurément plus efficace qu'un surveillant humain. Elle présenterait en outre un autre avantage : en cas de contestation, il serait aisé de revoir les enregistrements. Je me doute que cette suggestion peut sembler orwellienne. Mais si on considère que la triche constitue une grave injustice commise à l'encontre de tous les autres élèves respectant les règles, il me semble que ce recours aux technologies de surveillance n'a rien d'excessif. Le refus des technologies repose souvent sur des craintes mal déterminées ou des arguments de type « pente glissante », alors que le danger à combattre est lui bien réel. Certes, le coût financier d'un tel dispositif serait sans doute élevé, mais a-t-on jamais évalué le coût de la fraude scolaire ? La collectivité finance les études d'élèves qui n'ont pas légitimement gagné leur place, ce qui est à la fois injuste pour ceux qui ont été évincés dans la compétition scolaire, et coûteux pour la collectivité, puisqu'un élève suit alors une formation pour laquelle il n'a pas réellement les qualifications. Nous disposons donc d'arguments éthiques et prudents pour justifier une telle lutte contre la fraude, même si les moyens sont coûteux.

IA génératives : faut-il sauver l'école ou les élèves ?

Mais la lutte contre la triche n'a pas seulement pour but de sauvegarder l'institution scolaire, elle vise aussi à préserver l'autonomie intellectuelle des élèves. Le second motif de s'opposer à la fraude réside dans le fait qu'elle prive l'élève des efforts intellectuels nécessaires à son développement intellectuel, et le maintien dans la servitude et la dépendance vis-à-vis des technologies qu'il n'a ni conçues ni comprises.

Cet argument mérite toutefois d'être discuté. Les IA génératives peuvent en effet être envisagées de deux manières :

1. Elles privent les utilisateurs de l'occasion de penser par eux-mêmes, en leur fournissant des réponses rapides et automatiques sur n'importe quel sujet. En cela, elles génèrent de la dépendance et freinent les progrès personnels.
2. Elles offrent aux utilisateurs des réponses bien supérieures à celles qu'ils auraient pu produire seuls, et leur permettent ainsi d'aller beaucoup plus loin dans leur réflexion que s'ils n'y avaient pas eu accès. Un élève qui s'interroge sur un problème philosophique pourra obtenir une réponse d'une qualité tout à fait convenable. Je comprends les critiques que les philosophes de profession peuvent émettre contre les réponses parfois stéréotypées des IA, mais ces réponses restent bien meilleures que celles d'un adulte ordinaire sans formation spécifique.

Il est difficile de trancher entre ces deux perspectives. La réponse suppose de s'interroger sur la valeur que nous accordons à l'autonomie intellectuelle. A-t-elle une valeur procédurale, importante en elle-même, parce qu'il serait toujours souhaitable de comprendre et de justifier les discours que nous tenons ? A-t-elle une valeur instrumentale, importante seulement parce qu'elle nous conduit plus sûrement à la vérité que le fait de s'en remettre à autrui ?

D'un point de vue instrumental, et en toute généralité, suivre aveuglément l'opinion d'autrui n'est pas une méthode fiable pour atteindre la vérité. Mais les IA génératives, sans être infaillibles, sont relativement fiables et plus performantes qu'une personne non experte d'un sujet quelconque. Certes, elles « hallucinent » occasionnellement, c'est-à-dire qu'elles produisent des réponses grossièrement fausses et que n'importe quel humain reconnaîtrait aisément comme telles. Mais ces hallucinations restent rares, et le niveau moyen des réponses est très au-dessus de ce que les humains produisent. Si l'autonomie n'a qu'une valeur instrumentale, alors il semblerait plus rationnel de s'appuyer sur une IA que de tenter de réfléchir seul.

On peut formuler un second argument sur la valeur instrumentale de l'IA, cette fois du point de vue d'un étudiant qui y aurait recours. La grande majorité des

élèves ne poursuivra pas d'études supérieures en philosophie, et une partie importante d'entre eux ne lira plus de philosophie, ou de façon très marginale. Dès lors, l'apparition des IA génératives ne peut pas être comparé à l'apparition des calculatrices en mathématiques. En mathématiques, savoir calculer constitue une compétence de base indispensable, sur laquelle reposent des savoirs plus complexes. Renoncer à l'entraînement au calcul mental sous prétexte que l'on dispose d'une calculatrice fragiliserait les fondements de l'apprentissage des mathématiques plus avancées, qui font sans cesse appel aux opérations élémentaires (tables de multiplication, factorisations, etc.). En philosophie, le problème est différent : le premier cours de philosophie ne constitue pas une compétence de base sur laquelle viendraient se greffer des connaissances plus techniques ; et même si chaque professeur peut établir une certaine progressivité des apprentissages, il n'y a pas vraiment de niveau en philosophie ; chacun s'efforce de penser le mieux possible en fonction de ses connaissances et références. Certes, un professeur va toujours plus loin que ses élèves, mais ils accomplissent en un sens les mêmes opérations intellectuelles. Le fait de ne pas avoir fait beaucoup de philosophie ne pénalisera pas l'élève dans le supérieur, ou dans sa vie quotidienne. Pour cette raison, disposer d'un outil qui lui permet de produire des phrases philosophiques qu'il ne serait pas capable de formuler seul ne paraît pas spécialement handicapant dans sa vie ou dans ses études. L'IA, ici, est un « bonus » qui ne prive l'élève de rien.

La question de savoir si l'autonomie intellectuelle a une valeur intrinsèque est aussi importante. Il est difficile de répondre à une telle question sans se heurter à un socle de valeurs fondamentales qui ne se justifient pas aisément davantage. À titre personnel, j'attache une grande importance à l'individualité, au fait que chacun puisse faire des choix qui soient vraiment les siens, et former des idées qui soient vraiment les siennes. Tout comme John Stuart Mill, je préfère encore un individu qui se trompe après avoir examiné sincèrement une question, plutôt qu'un individu qui répète mécaniquement des idées vraies qu'il ne comprend même pas. Les individus ne sont pas seulement des mécanismes de diffusion d'idées qu'ils sont incapables de comprendre. Ils sont responsables de ce qu'ils disent et pensent, et doivent être à la hauteur de cette responsabilité. L'autonomie intellectuelle n'implique pas de se priver de l'usage des IA génératives, et rester volontairement dans l'ignorance n'est jamais souhaitable. Mais en raison de leur mode de fonctionnement fondé sur les probabilités, elles ont tendance à reproduire l'opinion majoritaire sur chaque sujet, et se contenter de celle-ci, en raison même du fait qu'elle est majoritaire, est une forme d'abdication de sa responsabilité intellectuelle. Ce qui était souhaitable avant l'apparition des IA le reste après leur apparition : chacun a le devoir de mettre ses idées à l'épreuve des faits et de la raison.

IA génératives : faut-il sauver l'école ou les élèves ?

D'autre part, on ne doit pas négliger les effets délétères d'un usage massif et dérégulé des IA sur les IA elles-mêmes. En effet, les IA génératives sont elles aussi exposées à des problèmes d'action collective. Ces technologies reposent sur des bases de textes qui leur servent à s'entraîner et à générer de nouveaux contenus. Or, la qualité de leurs productions dépend étroitement de la qualité des textes sur lesquels elles ont été formées. Si les individus renoncent à se former convenablement sous prétexte que les IA existent, ils ne produiront plus eux-mêmes de textes de qualité, ce qui appauvrira progressivement les bases de données des IA et dégradera leurs productions. Autrement dit, les IA génératives produisent, elles aussi, des passagers clandestins, qui utiliseront leurs productions sans contribuer à maintenir un niveau d'exigence suffisant dans les contenus produits par les humains. Elles ont besoin d'humains bien formés, donc des humains à qui on a déconseillé de les utiliser à tort et à travers.

La lutte contre la triche poursuit deux objectifs solidaires : établir la justice scolaire et la fiabilité des notes d'une part, et garantir l'autonomie intellectuelle des élèves d'autre part. Ces deux objectifs sont solidaires car on ne peut pas, sans hypocrisie, faire l'éloge de l'autonomie intellectuelle dans une école où les tricheurs sont impunis et obtiennent les meilleures places dans le supérieur. L'autonomie intellectuelle ne devrait jamais se retourner contre les intérêts matériels des élèves, et c'est au système scolaire de faire en sorte que les intérêts matériels et intellectuels convergent effectivement. En d'autres termes, celui qui a réellement fait l'effort de penser par lui-même doit aussi, toute chose égale par ailleurs, avoir les meilleurs résultats et les meilleures chances d'accéder aux formations sélectives. À défaut, les élèves adopteront une attitude cynique mais rationnelle : délaissé le travail intellectuel pour privilégier les stratégies opportunistes visant à obtenir des points sans effort (triche, devoirs bonus, etc.).

Si l'école est incapable de garantir la fiabilité des notes, alors son discours sur l'autonomie intellectuelle n'est rien de plus qu'une idéologie mensongère dont seraient victimes les élèves qui n'ont pas compris les véritables règles du jeu. On a raison de dire que « prompter n'est pas penser » (pour reprendre le titre d'un communiqué de l'APPEP sur le sujet). Mais encore faut-il que prompter ne soit pas la solution la plus rationnelle pour réussir sa scolarité.

Sylvain Theulle

Chronique de la critique de la dissertation : 2024, comment surmonter les copies ChatGPT ?

*Synthèse des JE : comment une mauvaise humeur peut
aboutir sur une bonne*

Fanny Bernard

1. Une inquiétude devenue réflexion

L'IA générative s'est imposée à nous comme un changement technologique majeur. À l'avènement de ChatGPT, l'IA générative gratuite lancée par OpenAI, je n'y ai vu qu'un nouveau gadget numérique. Mais après en avoir observé des traces dans des copies d'élèves, j'ai voulu le tester moi-même. J'ai compris que l'IA pouvait générer une très bonne dissertation.

La vidéo de Monsieur Phi¹ n'a fait qu'apporter une preuve de plus. La conférence de Bernard Victorri aux JE de novembre 2024 m'a également confortée dans cette idée : il suffit à un élève de maîtriser l'IA pour écrire une bonne dissertation.

Dès lors, que faire en tant que professeure de philosophie ? Faut-il continuer à demander aux élèves d'écrire des dissertations ? Peut-on encore, et si oui comment, continuer à enseigner la dissertation, et plus largement, à donner sens à l'écriture philosophique à l'heure des textes générés par IA ?

Lors de nos journées d'études, les conférences et les différents ateliers auxquels j'ai assisté ainsi que les discussions que j'ai eues avec les collègues ont été très riches. Elles m'ont permis de comprendre que même si plusieurs options s'offrent à nous, il semble à nouveau nécessaire et urgent de réinterroger la place de la dissertation dans nos pratiques². En outre, l'IA ne bouleverse pas seulement les exercices écrits : elle questionne la façon même d'enseigner.

2. Enseigner avec ou sans IA ?

Plusieurs collègues ont récemment proposé des pistes pour intégrer l'intelligence artificielle à l'enseignement de la philosophie, chacune selon un rapport différent au métier.

1 Voir la vidéo en ligne du site « Monsieur Phi », « De l'art superflu d'écrire des dissertations à l'heure de ChatGPT ». [[Disponible en ligne](#)]

2 Voir Jean-Jacques Rosat, « Penser et dissenter », *Côté Philo*, n°9, octobre 2006. [[Disponible sur le site de l'ACIREPh](#)]

Cédric Eyssette l'utilise avant tout comme une aide pour l'enseignant : un outil de conception, de reformulation ou d'entraînement, pensé pour soutenir le travail pédagogique sans modifier la relation directe entre élèves et savoirs.

Vivien Mirebeau¹, quant à lui, explore l'IA comme partenaire de dialogue : un interlocuteur virtuel capable de stimuler la réflexion et de déclencher des questionnements philosophiques.

Enfin, Marion Schumm s'interroge sur la possibilité d'intégrer l'IA dans la pratique de la dissertation elle-même, en invitant les élèves à écrire ou à réécrire avec son aide. Son ingéniosité l'a poussée à proposer un atelier de méthodologie de la dissertation, ou plutôt, la déconstruction de ChatGPT par son utilisation. Elle fait faire aux élèves un concours de triche : qui produira la meilleure introduction sur un sujet de dissertation en utilisant uniquement ChatGPT ? Aucun copier-coller ni écriture libre n'est toléré ! L'introduction finale doit être entièrement générée par ChatGPT, sans retouche de la part des élèves. C'est un excellent exercice de méthode : les élèves doivent donner les bonnes consignes à l'IA pour qu'elle produise une bonne intro. Cela nécessite plusieurs essais : il faut alors lire, évaluer et corriger le texte généré par l'IA pour qu'il soit amélioré. Marion affirme que les élèves se rendent alors compte que tricher demande plus de travail que de ne pas tricher. Le tour est joué ! La séance de méthodologie a permis aux élèves de comprendre les attendus d'une dissertation et la nécessité du travail rédactionnel pour y parvenir. Ce qui se pratiquait avant au GFEN sur des copies du bac, rappelle Cécile Victorri, peut se faire plus efficacement avec l'IA, puisque l'élève peut instantanément corriger la copie : en retouchant le prompt, c'est-à-dire en donnant des conseils méthodologiques plus précis, l'élève voit tout de suite ce qui rend l'introduction meilleure.

Mais lorsque les élèves ont compris comment écrire un bon prompt, qu'est-ce qui les empêche de l'utiliser à chaque fois que l'on demande de rédiger une dissertation ? Et surtout, dans cette perspective proposée par Marion Schumm, nous apprenons aux élèves à écrire en utilisant l'IA. Faut-il enseigner à écrire avec l'IA ou sans IA ? Marion Schumm part du constat que la majorité des élèves utilisent déjà l'IA, même lorsque les enseignants ne le font pas : les condamnations de droit semblent avoir perdu leur pertinence. Il faudrait alors réfléchir à une façon d'enseigner la dissertation avec une IA.

Ces trois approches dessinent un continuum d'usages : de l'appui à l'enseignant jusqu'à l'accompagnement direct de la rédaction des élèves.

Toutes, à leur manière, interrogent ce que signifie enseigner la philosophie à l'ère de l'IA : non seulement ce que l'on enseigne, mais aussi ce qu'est pour l'élève apprendre à philosopher.

¹ Voir son article dans ce même numéro.

2024, Comment surmonter les copies ChatGPT ?

Mais tout de même, en tant que professeur·es, est-ce notre rôle d'enseigner à utiliser l'IA correctement ? ChatGPT peut-il être un outil pour apprendre à philosopher ? Doit-il devenir la calculatrice de la philosophie, au sens où cet outil pourrait nous faire gagner du temps sur la rédaction comme la calculatrice le fait dans le calcul mental ? Ne faut-il pas plutôt protéger les élèves de l'IA ? N'est-ce pas céder au marketing d'OpenAI qui prétend faire de ChatGPT la calculatrice¹ des mots du XXIème siècle ?

D'abord, par l'utilisation en classe de l'IA, nous banalisons et légitimons son usage : faut-il participer à ce mouvement d'ancrage irréversible de l'IA dans nos vies, alors qu'il s'agit d'un outil technologique qui ne répond à aucun de nos besoins préalables ?

De plus, on ne peut pas négliger les impacts écologiques qui ont été rapidement évoqués par Camille Girard-Chanudet dans sa conférence à nos JE : la consommation énergétique (électricité, eau) pour l'entraînement des LLM et pour le maintien des serveurs est insoutenable, sans parler de la construction de ces mêmes serveurs sur des terres qui devraient servir à nous nourrir ou tout simplement maintenir une biodiversité, ni des émissions de CO². L'IA nécessite la multiplication de la puissance des puces graphiques et des centres de données, ainsi qu'une intensification de l'extraction de matières premières comme alerte la tribune publiée par « Hiatus », un collectif composé de plusieurs organisations, dont la Quadrature du Net, Changer de Cap, la Ligue des Droits de l'Homme³. Ils dénoncent également ce que Camille Girard-Chanudet soulignait dans sa conférence : « l'exploitation des travailleurs et travailleuses qui participent au développement et à la maintenance de ses infrastructures, notamment dans les pays du Sud global où elle prolonge des dynamiques néo-coloniales ». Cédric Eyssette évoque dans son atelier les LLM Open source et locaux, mais la construction de ces infrastructures numériques implique une course dans l'exploitation des ressources rares dont l'extraction provoque de graves conflits à l'étranger, mais aussi en France.

Enfin, on retrouve avec l'IA les problèmes sociaux et politiques classiques de l'utilisation du numérique.

On a peu évoqué le temps passé devant des écrans par les adolescents⁴, mais les spécialistes alertent sur un véritable problème de santé publique, notamment en termes d'addiction.

1 Voir Florian Bayard, « ChatGPT : OpenAI annonce une révolution des « devoirs à la maison », *01net*, 13 juin 2023. [Disponible en ligne]

2 Voir Nastasia Hadjadji « L'insoutenable coût écologique du boom de l'IA », *Reporterre*, 4 juillet 2024, [Disponible en ligne]

3 Collectif Hiatus, « L'intelligence artificielle accélère le désastre écologique, renforce les injustices et aggrave la concentration des pouvoirs », *Le Monde*, 6 février 2025. [En ligne]

4 Rapport « Enfants et écrans. À la recherche du temps perdu » par une commission d'experts pour la présidence de la République. [[Disponible en ligne](#)]

Ces outils posent aussi des questions éthiques relatives à la protection des données et au consentement¹. Cédric Eyssette nous rappelle dans son atelier que l'utilisation de ChatGPT en classe est d'ailleurs problématique à cet égard. Comme l'écrit Anne Alombert, on peut toujours se demander si c'est nous qui utilisons l'IA ou si ce n'est pas plus l'IA qui nous utilise² puisque le principe même des IA est de se servir de nos requêtes pour « s'améliorer ». Un tel outil de génération de textes pose en dernière instance le problème de l'automatisation du travail : que deviendront nos facultés de rédaction si elles s'exercent uniquement par la relecture de textes produits par l'IA ?

Enfin, on peut se questionner sur les nouvelles dépendances que peut créer l'IA : si l'on n'apprend plus à écrire sans IA, la rédaction demeure tout de même essentielle dans nos vies quotidiennes (mails, SMS, sites internet, lettres de motivation, etc.). De fait l'IA deviendra un outil numérique dont on ne pourra plus se passer puisqu'on n'aura pas appris à faire sans.

Ce rapide résumé des enjeux écologiques, politiques, sociaux et anthropologiques de l'IA me conduit à refuser son usage en classe. Mais il y a aussi des raisons pédagogiques.

L'objectif de notre enseignement est d'apprendre à problématiser, conceptualiser et argumenter, si l'on reprend les termes de Michel Tozzi³. L'utilisation correcte d'une IA générative pour écrire une dissertation me semble développer des compétences d'interprétation et de commande technique. Mais l'évaluation d'un texte écrit par l'IA ne permet pas de mesurer si la personne qui l'a prompté sait problématiser, conceptualiser et argumenter. Prompter implique une certaine forme de pensée, mais qui n'est pas celle enseignée en philosophie.

Apprendre aux élèves à générer des textes par l'IA non seulement reconnaît ce type d'écriture comme équivalent à la rédaction humaine, mais cela laisse croire que les textes s'écrivent ainsi.

Le travail de rédaction d'une dissertation permettait de faire comprendre aux élèves les allers-retours, les capacités cognitives, le temps nécessaires à la production d'un texte philosophique. Développons-nous véritablement leur intelligence philosophique⁴ en leur faisant produire un texte par une intelligence artificielle ?

Comme l'a montré Christine Barré-De Miniac (2015), écrire ne consiste pas seulement à produire un texte conforme à des attentes scolaires : c'est un travail de

1 Voir l'article de Martin Untersinger « CNIL : les propositions-chocs de la Commission de l'IA pour faciliter l'accès aux données personnelles », *Le Monde*, le 13 mars 2024. [Disponible en ligne]

2 Anne Alombert, « ChatGPT : derrière le mythe de l'intelligence artificielle, les dangers des automates computationnels », *Le Vent Se Lève*, 27 mars 2023. [Disponible en ligne]

3 Voir Michel Tozzi, *Penser par soi-même*, Chronique sociale, 2005.

4 Voir la critique Anne Alombert du terme « intelligence artificielle » qui lui préfère celui « d'automate computationnel », dans ce même numéro.

pensée et de subjectivation. L'écriture engage le sujet dans un rapport à lui-même et au savoir, par les détours, les reprises, les hésitations qu'elle impose. C'est pourquoi déléguer l'acte d'écrire à une intelligence artificielle, c'est risquer de priver l'élève de cette expérience formatrice : celle d'un rapport réflexif à sa propre pensée, où se construisent à la fois le sens et la posture philosophique. Autrement dit, ce n'est pas seulement la qualité du texte produit qui est en jeu, mais le processus d'élaboration de soi par l'écriture, ce que l'IA, par nature, ne peut ni éprouver ni transmettre.

3. L'IA comme révélateur d'une crise de sens

Je pensais avant les JE qu'il suffisait de ne plus donner de dissertation en devoir maison et d'exercer une surveillance de geôlier lors des devoirs en classe (car les montres connectées permettent aux élèves d'accéder à ChatGPT). Et qu'on pouvait alors corriger des copies écrites avec certitude par les élèves. Mais outre le fait qu'on retrouve de la triche avec ChatGPT même lors des devoirs surveillés (on peut se cacher aux toilettes pour consulter internet, cacher une montre connectée dans une trousse, etc.), l'enjeu me semble ailleurs. L'IA ne fait que révéler une fragilité déjà ancienne : celle du sens donné aux exercices scolaires, et notamment à la dissertation. D'ailleurs la question de la triche n'est pas nouvelle : l'élève pouvait déjà avoir appris par cœur un texte qu'il rédige le jour de l'évaluation. Cela m'était arrivé avant ChatGPT : un élève avait appris par cœur un corrigé du commerce sur le sujet que j'avais choisi pour le bac blanc.

Comme le souligne Philippe Meirieu (2023), le véritable enjeu n'est pas tant celui de la fraude que celui du rapport au savoir qu'elle révèle. L'IA met en évidence la confusion croissante entre le désir de savoir (immédiat, sans épreuve ni distance) et le désir d'apprendre, qui suppose un effort, un cheminement, une lente transformation de soi.

Or, c'est précisément ce passage du "savoir" à "l'apprendre" que la dissertation visait, à l'origine, à accompagner. Si cet effort perd son sens, alors le recours à l'IA n'est que le symptôme d'un désinvestissement plus profond : celui du goût d'apprendre et du plaisir de penser par soi-même.

Le problème n'est pas simplement de savoir, lorsqu'on corrige sur Santorin les copies du bac, si elles ont été écrites par une IA ou non. Il faut surtout réussir à expliquer aux élèves pourquoi on leur apprend toute l'année à accomplir une tâche difficile (la dissertation) qu'un site internet peut faire en quelques secondes. En somme, que ce que cette tâche n'est pas seulement une performance mais un outil de formation. Mais le sens de l'effort et le goût de la gymnastique intellectuelle risquent de ne pas suffire aux futures générations pour avoir le désir d'apprendre la dissertation.

Il faut savoir aussi ce que nous, enseignant.es, voulons faire apprendre par la dissertation et plus généralement les finalités de notre enseignement ? Qu'apprendre à des élèves du XXI^{ème} siècle qui auront à travailler dans un monde où de nombreux textes seront générés par des machines ?

Si c'est juste maintenir la dissertation coûte que coûte, nous risquons de confondre la forme scolaire de la philosophie et la philosophie elle-même. L'abandon de la dissertation n'est pas un deuil de la philosophie. Comme l'écrivait Jean-Jacques Rosat : « La dissertation a été inventée pour l'école, pas pour la philosophie »¹. Rappelons que ni Platon, ni Descartes ni Marx n'ont écrit de dissertation. Il serait donc possible de la remplacer par un autre exercice philosophique tout en enseignant toujours notre discipline.

Allons même plus loin : faudrait-il abandonner tout exercice écrit en cours de philosophie puisqu'on ne peut plus détecter par qui il a été écrit ? Renoncer à toute évaluation ? Est-ce alors la mort de notre enseignement ?

Non !

5. Quel mode d'évaluation valable en 2025 ?

Et si on évaluait à l'oral ?

La solution peut sembler farfelue au premier abord. Et pourtant faire du baccalauréat une évaluation orale a déjà été pensé à une autre époque ou dans d'autres lieux. En France, où jusqu'en 1830, les épreuves étaient uniquement orales, puis mixtes jusque dans les années XXXX. Au Portugal², 10 % de l'évaluation se fait à l'oral selon la présentation qu'en ont faite Serge Cospérec et J.N. Vicente dans le Côté Philo consacré à l'enseignement de la philosophie au Portugal. En Italie, l'examen comparable au baccalauréat appelé « la maturità » était en majeure partie oral. Enfin, Derrida et Bouveresse affirmaient déjà dans leur rapport sur l'enseignement de la philosophie : « au baccalauréat d'enseignement technique, nous proposons que l'épreuve devienne orale et consiste en une interrogation à partir d'un dossier constitué par le candidat durant l'année »³.

De même, l'agrégation de philosophie comporte elle-même plus d'épreuves orales (quatre) que d'épreuves écrites (trois). L'idée d'un exercice philosophique oral mérite d'être repensée, d'autant que les épreuves orales actuelles (comme la leçon d'agrégation) peuvent elles aussi être générées par ChatGPT.

1 Voir Jean-Jacques Rosat, « Penser et dissenter », *Côté Philo*, n°9, octobre 2006. [[Disponible sur le site de l'ACIREPh](#)]

2 Voir Serge Cospérec, « L'enseignement de la philosophie au Portugal », *Côté Philo*, n° hors-série, avril 2023. [[Disponible sur le site de l'ACIREPh](#)]

3 Voir le Rapport de la Commission de philosophie et d'épistémologie co-présidée par Jacques BOUVERESSE et Jacques DERRIDA en 1989. [[Disponible sur le site de l'ACIREPh](#)]

2024, Comment surmonter les copies ChatGPT ?

Il ne s'agit donc pas d'opposer oral et écrit (l'oral à l'école étant d'ailleurs scriptural), mais d'inventer des formes d'évaluation où la pensée se déploie dans l'échange plutôt que dans la simple restitution d'un discours figé.

Il ne s'agit pas de ne plus enseigner l'écrit puisque pour produire un oral, il est souvent nécessaire de passer par l'écrit : prendre des notes, structurer sa pensée, élaborer des formulations. De même que pour lire un texte philosophique, il est utile de prendre des notes à côté et sur le texte. L'écrit reste un outil essentiel du travail de pensée. Mais il ne peut plus, à lui seul, incarner toute l'évaluation du philosophe. Et il peut être fécond de penser l'écrit et l'oral comme deux modalités complémentaires du philosophe.

Il ne s'agit pas non plus de survaloriser l'oral comme mode d'expression contemporain. L'oral philosophique demande un apprentissage exigeant et un cadre rigoureux. Il ne s'agit pas d'une spontanéité du type conversationnel : la réponse faite dans le cadre d'un échange oral philosophie est nécessairement nourrie d'une réflexion préalable, qu'il s'agisse d'un cours de philosophie, d'une lecture, d'une pensée personnelle... La spontanéité est en réalité un leurre dans le cadre d'une interaction orale philosophique scolaire : sont formulées des pensées qui ont eu le temps de se construire auparavant. En tant qu'enseignante, je vois bien que les élèves qui ont des interventions pertinentes et solides lors des colloques des philosophes par exemple, sont celles et ceux qui ont travaillé le sujet et leur position philosophique. L'élaboration d'une réflexion écrite pourrait avoir lieu avant ou après, sans être le cœur de l'évaluation comme il l'est aujourd'hui.

Il faut ici distinguer deux choses trop souvent confondues : apprendre à philosopher et évaluer la philosophie. L'écrit, par exemple, demeure un lieu privilégié de formation : il permet à l'élève de revenir sur sa pensée, de la stabiliser, de prendre conscience de ses biais et de ses intuitions. Mais en tant qu'outil d'évaluation sommative, il atteint aujourd'hui ses limites : le risque de délégation à l'IA y rend l'authenticité de la production difficile à garantir, et l'épreuve sur table évalue davantage la performance que le processus d'apprentissage.

6. Quels exercices oraux pour évaluer sans trahir le philosophe ?

Quel type d'exercice oral permettrait alors d'enseigner la philosophie sans être un texte générable par une IA appris par cœur ?

Il s'agirait d'un exercice qui implique un dialogue avec autrui, qu'il s'agisse d'un·e professeur·e ou d'un·e autre élève, et non le monologue d'une dissertation.

Nous sommes un certain nombre de professeurs à évaluer les activités orales de nos élèves¹ : colloque des philosophes, procès, communauté de recherche philosophique, débat mouvant, discussion à visée démocratique et philosophique, etc.² Nous réfléchissons déjà à une grille de notation en tentant de prendre en compte des critères classiques de l'évaluation philosophique : argumentation, problématisation, conceptualisation, références philosophiques, analyse d'exemples ; en y ajoutant des critères propres à l'interaction orale : respect des règles de parole, niveau de langue correct, réaction à des propos entendus, etc.

Il existe des grilles d'évaluation des communautés de recherche philosophique dans des pays étrangers francophones. On peut les retrouver par exemple dans *La pratique du dialogue philosophique au secondaire* de Mathieu Gagnon et Sébastien Yergeau. La notation peut rester individuelle même si l'exercice se fait en groupe. On l'observe de façon empirique dans nos classes : ces interactions orales de groupe conduisent les élèves à des réflexions beaucoup plus profondes philosophiquement que les copies que nous corrigeons le jour du bac. À ce titre, il faut véritablement prendre de la distance par rapport aux exercices scolaires que nous enseignons quotidiennement pour nous rappeler ce qui fait philosophie, comme l'écrit *PhiloCité*³ : « tout travail véritable philosophique collectif naît de la coïncidence de ces divers éléments : au sein du flux de sa propre existence, chacun doit d'abord entamer un travail philosophique sur lui-même pour ensuite l'élargir à la collectivité qu'il rencontre et avec laquelle il dialogue. Car il n'y a pas de travail philosophique collectif sans deux aspects : la rencontre entre individus et le dialogue. »

On peut cependant aussi envisager comme évaluation du baccalauréat une interaction entre l'élève et le jury afin que l'élève ait plus le temps de développer sa propre parole qu'au cours d'une interaction avec d'autres élèves. Il ne s'agirait pas d'exposer que l'élève aurait appris par cœur à l'avance, car dans ce cas, il pourrait être construit par une IA, comme ça l'est déjà en grande partie pour le Grand Oral. L'élève pourrait produire un discours qui donnerait lieu ensuite à des questions du jury. Si la notation surévalue les réponses aux questions par rapport

1 Lors des JE sur le féminisme j'avais proposé un colloque sur sexe et genre, en se demandant comment faire en sorte que tous les élèves puissent participer sans que soient reconduites dans la prise de parole des dominations (sexistes notamment). Aux JE sur les séries technologiques, j'avais proposé un atelier sur les inégalités dans la prise de parole entre filles et garçons, dont on pouvait tirer des modalités d'organisation prenant en compte les rapports de force pour tenter de les éviter. Voir le compte-rendu de mon atelier fait Mathieu Bompont, dans *Côté Philo*, n°26, décembre 2024. [[Disponible](#) sur le site de l'ACIREPh]

2 Ces quatre types d'exercices d'interaction orale sont présentés dans *PhiloCité, Philosopher par le dialogue*, Vrin, 2020.

3 *PhiloCité, Philosopher par le dialogue*, Vrin, 2020, page 22.

à l'exposé¹, le risque de psittacisme d'une IA est amoindri. Les questions amènent forcément à de la nouveauté.

Enfin, d'un point de vue purement social, la préparation d'une interaction orale permettrait de travailler également le lien humain entre élèves. Les 16-19 ans passent en moyenne 5h10 sur un écran par jour selon une étude du centre national du livre sur les jeunes Français et la lecture. Il est de notre responsabilité de leur proposer autre chose en classe que d'utiliser un écran pour faire de la philosophie. Je ne bannis pas tout usage du cinéma ou du vidéoprojecteur, mais les élèves le disent : ils aiment ce qu'ils appellent « débattre », échanger oralement avec leurs camarades en classe. Outre cette dimension sociale importante de l'exercice oral en philosophie, *PhiloCité* et Michel Tozzi notent dans leurs travaux que les discussions philosophiques ont également une dimension démocratique : sans en faire la visée principale comme eux, elle peut être un effet bénéfique non négligeable.

Toutefois, l'évaluation orale ne saurait devenir l'unique horizon. Elle comporte ses propres limites : la spontanéité apparente peut masquer des inégalités d'aisance (gestion du stress dû à la forte exposition de soi), des rapports différenciés au langage ou des difficultés à formuler des idées complexes dans l'instant. De plus, l'oral privilégie la dynamique du processus, tandis que l'écrit permet une élaboration lente, critique et réflexive : celle qui permet de revenir sur sa pensée, de la réviser, de la formuler avec rigueur. Il est vrai que lors d'interactions orales, la spontanéité est nécessaire. Est-elle un critère d'évaluation pertinent en philosophie ? Enfin, les professeurs n'ont pas été formés pour évaluer l'oral individuel ou collectif des élèves.

7. Pour conclure :

En résumé, le déploiement massif et généralisé de l'Intelligence Artificielle et l'engagement financier de l'État français dans son développement nous obligent à repenser notre enseignement et ce que nous entendons par « philosophie ». On ne peut pas se contenter d'exiger du ministère des conditions pour enseigner et évaluer la même chose que nous avons toujours faite, comme le demande l'APPEP dans son communiqué. L'IA est un tournant dans notre rapport à l'écrit, même si je me refuse à l'utiliser pour des raisons écologiques, politiques et sociales. L'Inspection Générale devra prendre acte qu'il va être difficile de savoir aux prochaines sessions du baccalauréat si une copie a été écrite par un·e élève ou par l'IA que l'élève aura consulté en cachette. Seule une évaluation orale des élèves me semble possible et pertinente en philosophie afin de ne pas passer notre temps à « corriger ChatGPT » (que l'on soit élève ou prof) et afin de retrouver le plaisir de philosopher ensemble. Si ce qui est important à l'école est le travail personnel

¹ Je remercie pour cette suggestion Vanina Mozziconacci qui pratique déjà cette forme d'évaluation à l'université.

de l'élève tout au long de l'année, et non le produit fini, comme me l'avait objecté Cécile Victorri lors d'une discussion après les JE, pourquoi ne pas valoriser dans l'évaluation ce chantier, plutôt que l'œuvre achevée en temps limité sur table le jour de l'examen ? Loin de signifier la mort de l'enseignement de la philosophie, l'utilisation par les élèves de ChatGPT nous invite à repenser d'urgence ce que nous souhaitons enseigner quand nous enseignons la philosophie.

Ce texte n'a pas été écrit par une IA !

Fanny Bernard

Politique et IA : entre 1984 d'Orwell et le Procès de Kafka

Serge Cospérec

I. IA et Big data. Vers un monde orwellien ?

Être surveillé tout le temps et partout

Les nouvelles formes de surveillance et de contrôle que permettent l'IA et les *Big Data* rendent notre monde de plus en plus semblable au monde dystopique imaginé par Orwell dans *1984*, dans lequel un gouvernement totalitaire contrôle ses citoyens par la surveillance généralisée, la manipulation de l'information et la répression.

IA et Big data

Beaucoup d'outils de surveillance utilisés aujourd'hui restent classiques, l'innovation technique les a seulement rendus plus efficaces et plus discrets : micros et systèmes d'écoute, logiciels espions et trackers, activation à l'insu des utilisateurs des micro et caméra de leur smartphone, vidéosurveillance sur les lieux publics et privés. Si l'IA ne change pas tout, la nouveauté c'est le *Big data* (la collecte de *données massives*) et le *data mining*, c'est-à-dire l'extraction d'informations utiles à partir de ces ensembles de données grâce aux puissants algorithmes statistiques et analytiques de l'IA. La nouveauté, c'est aussi que les individus fabriquent eux-mêmes leur « dossier numérique ». D'abord, en consentant à l'enregistrement de leurs données à travers toutes sorte d'objets connectés : smartphones, PC, tablettes, enceintes et assistant vocaux, TV intelligente, montres, vélos, appareils ménagers, etc., dont beaucoup « écoutent » et enregistrent en permanence. Peu de personnes ont conscience de la quantité de données qu'elles acceptent de divulguer, *via* les paramètres de ces objets (les smartphones en particulier) et l'installation d'applications (qui lit les mentions légales avant d'en installer une ?)¹. Ensuite, en *exposant* leur vie (leur

¹ On pourra se reporter au grand classique de David Lyon, *The Electronic Eye : The Rise of Surveillance Society* (University of Minnesota Press, 1994). Pour un ouvrage récent et bien fait, voir Gabriel Ganascia, *Servitudes virtuelles*, Seuil, 2022. Les premières pages de l'introduction à *La Société d'exposition, Désir et désobéissance à l'ère numérique* (Seuil, 2020) de Bernard E. Harcourt, sont également très stimulantes. En revanche dire que *1984* représenterait un paradigme dépassé, parce que dans les « société

« *story* »), en partageant *volontairement* toutes sortes de photos et d'informations personnelles sur les réseaux sociaux, sur leur blog, leurs pages personnelles, sans qu'aucun *Big Brother* ne les y contraigne ou n'ait d'efforts à faire pour les acquérir. Les GAFAM – Google (Alphabet), Apple, Facebook (Meta), Amazon et Microsoft – et leurs satellites (WhatsApp, Instagram, X, TikTok, Snapchat) accomplissent le fantasme des régimes autoritaires et des entreprises capitalistes : que les citoyens-consommateurs livrent d'eux-mêmes une masse d'informations personnelles qu'ils se chargeront d'agréger et d'exploiter « pour leur bien ». Les GAFAM sont aussi prêts à livrer ces données à toute dictature qui menacerait de les interdire sur son territoire. Tout cela nous rend dociles (on s'habitue à consentir), vulnérables, et complices de l'utilisation à grande échelle des instruments de contrôle et d'asservissement, puisque non seulement nous les acceptons, mais nous en usons en permanence à travers les avis, les étoiles, servant à « évaluer » les restaurants, les livreurs, les chauffeurs de taxis, etc., que nous déposons volontiers ou consultons.

Intérêt de la métaphore orwellienne

Daniel Solove dans *The Digital Person : Technology and Privacy in the Information Age*, souligne à juste titre que les *Big Brothers* d'aujourd'hui n'ont pas intérêt à ce que nous les percevions comme une menace. Ils veulent au contraire être aimés. La métaphore orwellienne serait obsolète. Cette conclusion, trop rapide, est erronée. D'abord, parce que la bienveillance des GAFAM ne dure que *tant que dure notre consentement* à leur pouvoir. En témoigne, la violence de leurs attaques contre les tentatives de régulation étatique ou supra-étatique (l'Union Européenne) de leurs activités. Ensuite, parce que les régimes totalitaires (Russie, Chine, Iran, Corée du Nord) ou autoritaires (Turquie, Arabie Saoudite, Singapour, Biélorussie, Birmanie, etc.) se servent de l'IA pour *manipuler, surveiller* et *réprimer* leurs opposants d'une façon classiquement orwellienne : désinformation, contrôle des médias, surveillance des individus, enlèvements, rééducation, disparitions, tortures, procès politiques). Enfin, parce que dans 1984, le dictateur d'Océania, *Big Brother*, travaille à supprimer l'idée même de révolte de l'esprit de ses sujets par la manipulation de l'information, la réécriture de l'histoire (comme aujourd'hui en Inde, en Chine, en Russie, et désormais aux États-Unis), et une révision de la langue pour en faire disparaître tous les termes politiquement incorrects ou subversifs, dans l'espoir que l'idée même d'une critique, d'une opposition ou d'une révolte ne puisse plus se développer dans l'esprit (j'y reviens ensuite).

d'exposition » la collecte des données est volontaire et le contrôle consenti, est plutôt naïf. Les régimes autoritaires progressent partout dans le monde, y compris en Europe. Ils sont l'exact contraire des « sociétés d'exposition » ; leurs citoyens évitent de communiquer et de s'exposer sur les réseaux qui sont constamment surveillés grâce à l'IA.

Orwell, Bentham, Foucault

Orwell expose dès le début de son roman le principe des sociétés de surveillance. Pour soumettre des individus, la force n'est pas nécessairement requise. La crainte d'être vu suffit à les tenir. Et la surveillance n'a même pas besoin d'être permanente car l'essentiel est que les individus se croient surveillés.

Le télécran recevait et transmettait simultanément. Il captait tous les sons émis par Winston au-dessus d'un chuchotement très bas (...). [Il] pouvait être vu aussi bien qu'entendu. Naturellement, **il n'y avait pas moyen de savoir si, à un moment donné, on était surveillé.** Combien de fois, et suivant quel plan, la *Police de la Pensée* se branchait-elle sur une ligne individuelle quelconque, personne ne pouvait le savoir. **On pouvait même imaginer qu'elle surveillait tout le monde, constamment.** Mais de toute façon, elle pouvait mettre une prise sur votre ligne chaque fois qu'elle le désirait. On devait vivre, on vivait, car l'habitude devient instinct, en admettant que tout son émis était entendu et que, sauf dans l'obscurité, tout mouvement était perçu.

Orwell, 1984.

1984 applique à la société du principe d'architecture carcérale imaginé par Jeremy Bentham dans son *Panopticon*¹ (1791). Où qu'ils se trouvent les prisonniers peuvent toujours être vus par des surveillants qu'ils ne peuvent pas voir. Foucault analyse le dispositif dans *Surveiller et Punir* [Gallimard, 1975, SP désormais] :

L'effet majeur du Panoptique : induire chez le détenu un état conscient et permanent de visibilité qui assure le fonctionnement automatique du pouvoir. Faire que **la surveillance soit permanente dans ses effets, même si elle est discontinuée dans son action.** [SP, 202]

Dispositif important, car il **automatise et désindividualise le pouvoir.** Celui-ci a son principe moins dans une personne que [...] dans un appareillage dont les mécanismes internes produisent le rapport dans lequel les individus sont pris. [...] Il y a une machinerie qui assure la dissymétrie, le déséquilibre, la différence. Un assujettissement réel naît mécaniquement d'une relation fictive. **De sorte qu'il n'est pas nécessaire d'avoir recours à des moyens de force.** [SP, 203]

¹ Le titre exact est *PANOPTICON or THE INSPECTION-HOUSE*, et le sous-titre : *containing the idea of a new principle of construction applicable to any sort of establishment, in which persons of any description are to be kept under inspection ; and in particular to penitentiary-houses, prisons, houses of industry, work-houses, poor-houses, lazarettos, manufactories, hospitals, mad-houses, and schools, with a plan of management.* L'objectif de Bentham est humaniste : trouver un moyen de maintenir l'ordre sans constamment punir, battre ou supplicier les prisonniers. Foucault ne s'y trompe pas.

Orwell met aussi en évidence un effet plus subtil et envahissant. Ce type de surveillance conduit à intérioriser les normes sociales qui, bientôt, ne sont même plus perçues comme répressives « *On devait vivre, on vivait, car l'habitude devient instinct* ». Foucault analyse cet effet :

Celui qui est soumis à un champ de visibilité, et qui le sait, reprend à son compte les contraintes du pouvoir ; il les fait jouer spontanément sur lui-même ; il **inscrit en soi le rapport de pouvoir** dans lequel il joue simultanément les deux rôles ; il devient le principe de son **propre assujettissement**. [...] Le pouvoir externe [...] tend à l'incorporel ; et plus il se rapproche de cette limite, plus ces effets sont constants, profonds, acquis une fois pour toutes, incessamment reconduits : perpétuelle victoire qui évite tout affrontement physique et qui est toujours jouée d'avance. [SP, 204]

Au-delà de la prison, Foucault montre que le principe panoptique caractérise les *sociétés disciplinaires* : « L'agencement panoptique programme, au niveau d'un mécanisme élémentaire et facilement transférable, le fonctionnement de base d'une société toute traversée et pénétrée de mécanismes disciplinaires ». [SP, 210]

Ce n'est pas un hasard si, le slogan le plus célèbre du régime khmer rouge était « *L'Angkar a les yeux de l'ananas* »¹. L'aphorisme vaut avertissement : l'Angkar voit tout, l'Angkar sait tout. Pour Henri Locard, 1984 décrit « nombre des mécanismes de la société polpotienne 30 ans avant qu'elle n'existe » ; l'aphorisme de l'ananas est « la version khmère rouge du célèbre « *Big Brother is watching you* » orwellien. [...] Pol Pot se faisait appeler « Frère Numéro Un » ou même carrément "*Big Brother*", seulement entre Khmers Rouges s'entend, puisque la population ordinaire ne le connaissait pas »², il restait invisible, sans visage.

Le système du « crédit social » chinois, accomplissement du schéma orwellien

La collecte massive de données rend aujourd'hui possible un gouvernement omniscient, qui surveille ses citoyens, les espionne et les contrôle par la peur, comme on le voit en Russie et plus encore en Chine, avec le système du crédit social [SCS

¹ L'ananas « qui possède une centaine d'« yeux » tournés dans toutes les directions, renvoie à la vision panoptique qui permet de tout voir sans être vu, source de tout pouvoir et de toute autorité. Les yeux de l'Angkar sont des espions (« *chhlop* »), recrutés parmi de jeunes adolescents qui devaient dénoncer les traîtres, y compris leur propre famille. » Commentaire des *Yeux d'ananas*, peinture de l'artiste cambodgien Bor Hak, dans la revue */art absolument/*, n° spécial d'avril 2013, « *Création et postmémoire* », page 37 [[Disponible en ligne](#)]. Un analogue scolaire est la phrase efficace avec des petits : « Attention ! même quand j'écris au tableau, je vous vois ; j'ai des yeux dans le dos ».

² Henri Locard, *Le "Petit Livre Rouge" de Pol Pot ou les paroles de l'Angkar*, L'Harmattan, 2019, page 88.

désormais]¹. Yongxi Chen et Anne S.Y Cheung expliquent que le SCS a « pour objectif d'évaluer la fiabilité des citoyens chinois dans la tenue de leurs promesses et le respect des règles juridiques, des normes morales et des normes professionnelles et éthiques ». Mais ils soulignent, dès 2017, leur crainte « que la totalité de la vie des individus soit enregistrée, que les citoyens soient surveillés et que l'État **orwellien** devienne une réalité ». Et cela d'autant plus, que les individus ne savent pas « ce qui contribue à leur score de crédit social, ignorent la façon dont il est composé [...] et la façon dont leurs données sont interprétées et utilisées.² » Le SCS repose sur la collecte d'une myriade de données sur le comportement des citoyens chinois, dans les entreprises, dans la rue, dans les transports en commun, y compris leur respect du code de la route et leur « moralité » sur les réseaux sociaux. Les mégadonnées sont analysées par une IA qui calcule la « qualité » du citoyen et lui attribue un score, dont dépend le périmètre de ses droits. Un score trop bas peut les priver « du droit de postuler à certains emplois, d'inscrire leurs enfants dans certaines écoles ou encore de prendre les transports en commun », « l'organe de presse du gouvernement chinois, *Global Times*, déclarait ainsi récemment que « 11 millions de personnes avaient été empêchées de prendre l'avion, 4 millions et demi de prendre le train »³. Les convictions religieuses sont également surveillées. Appartenir aux cultes « mauvais » ou « hérétiques » (bouddhisme, islam, une dizaine d'Églises chrétiennes, Falun Gong, etc.) vaut inscription sur les listes noires du « crédit social »⁴.

En principe, un État démocratique ne surveille qu'à de bonnes fins : identifier des activités criminelles, cybercriminelles, pédopornographiques, etc. Mais il collecte lui aussi d'énormes quantités d'informations et dispose grâce à l'IA d'une puissance inédite de surveillance des individus, et rien ne garantit qu'une démocratie soit protégée. En 2025, on a vu comment l'administration de Donald Trump s'est servie du fichier recensant 500.000 migrants latino-américains — qui s'étaient vu accorder un statut légal sous Joe Biden — pour les identifier, les traquer, les arrêter et les expulser, après avoir purement et simplement révoqué ce statut.

¹ Sur le SCS, on peut se reporter à la première étude publiée en France, celle du chercheur Emmanuel Dubois de Prisque « Le système de crédit social : comment la Chine évalue, récompense et punit sa population », *Institut Thomas More*, « Note 36 », juillet 2019. [[Disponible en ligne](#)]. Je signale le bon documentaire de Sébastien Le Belzic, *Ma femme a du crédit*, diffusé sur LCP en 2022.

² Chen, Yongxi et Cheung, Anne SY, « The Transparent Self Under Big Data Profiling: Privacy and Chinese Legislation on the Social Credit System », *The Journal of Comparative Law*, 2017, vol. 12, n° 2, pages 356-378. [[Disponible en ligne](#)]

³ « Surveiller pour punir : la notation des citoyens chinois », *France Culture*, 9 janvier 2019. [[Disponible en ligne](#)]

⁴ Emmanuel Dubois de Prisque, *La Chine et ses démons : Aux sources du sino-totalitarisme*, Odile Jacob, 2022, Chapitre VI, « La Chine, le christianisme et la liberté ».

Surveiller et faire punir ses semblables sans passer par la Justice

Épier ses semblables. À la surveillance verticale et hiérarchique des États, s'ajoute la surveillance horizontale et « égalitaire » de nos pairs sur les réseaux sociaux. Les « justiciers » et délateurs du Net traquent les maladroites des utilisateurs, leurs fautes et abus, supposés ou réels, pour s'en moquer, les dénoncer ou exiger leur punition, même longtemps après qu'ils se sont produits. La puissance des algorithmes de recherche facilite la recherche d'informations, y compris privées. Chacun se sentant désormais surveillé et enregistré, à des fins potentiellement malveillantes, se censure dans ce qui, initialement, était un espace de liberté.

Détruire sa vie en 63 signes. Justine Sacco, avant de prendre un avion pour Le Cap, tweete aux 500 abonnés de son compte personnel : « *Going to Africa. Hope I don't get AIDS. Just kidding. I'm white !* ». En français « *En route pour l'Afrique. J'espère que je ne vais pas choper le sida. Je blague, je suis blanche* ». À l'atterrissage au Cap, dix heures plus tard, elle rouvre son smartphone et découvre son licenciement. Le tweet est devenu viral. Le premier hashtag créé (#HasJustineLandedYet) agrège des milliers de tweets indignés. Le taux de partage record attire l'attention des médias en ligne qui le diffusent l'info, provoquant un lynchage médiatique.¹

Licencié à cause d'une photo prise à son insu. Adama Cissé, agent de propreté à Paris est licencié après la diffusion sur les réseaux sociaux d'une photographie le montrant allongé sur rebord de vitrine, en tenue de travail, chaussures et chaussettes à côté de lui, avec ce commentaire « *Voilà à quoi servent les impôts locaux des Parisiens à payer les agents de propreté à roupiller, on comprend pourquoi Paris est si dégueulasse* ». Le buzz se développe, la Mairie de Paris interroge son employeur, qui décide de le licencier².

Licenciée pour des messages « déplacés » après l'assassinat de l'influenceur trumpiste Charlie Kirk. Laura Sosh-Lightsy employée d'une université publique du Tennessee écrit sur sa page Facebook : « *La haine engendre la haine. ZERO sympathie* ». Une sénatrice républicaine, Marsha Blackburn, dénonce aussitôt la publication : « *Cette personne devrait avoir honte (...). Elle doit être virée de son poste* ». Dans la foulée, l'université annonce son licenciement.³

Ces trois affaires sont bien différentes. Elles ont pourtant deux points en commun : 1° le rôle des algorithmes d'IA dans la création, l'ampleur et la propagation

¹ Pierre Haski, « Elle tweete une horrible blague raciste, sa vie devient un cauchemar », *Le nouvel Observateur*, 2013. [[Disponible en ligne](#)]

² Sur les circonstances : « Peut-on être licencié à cause d'une photo prise à son insu ? », *Blog Cadre averti*, 2020. [[Disponible en ligne](#)]. Sur l'affaire et le jugement final, 3 ans et demi plus tard (!), « Rouen. Licencié après avoir été dénoncé par « un passant », l'éboueur gagne aux prud'hommes », *Actu 76*, [[Disponible en ligne](#)]

³ « États-Unis. Licenciements massifs pour ceux qui se sont amusés de la mort de Charlie Kirk », *Courrier International*, 18 septembre 2025, [[Disponible en ligne](#)]

du buzz médiatique ; 2° l'atteinte au droit fondamental en démocratie d'être jugé équitablement, publiquement et dans un délai raisonnable, par un juge indépendant et impartial.

Noter ses semblables. La surveillance s'exerce aussi par l'évaluation des employés des entreprises et des services publics. Nous évaluons le chauffeur UBER, le livreur Amazon, le serveur du restaurant, le médecin du quartier, le personnel de l'hôpital, et désormais, les établissements scolaires et le professeur d'université¹. Les algorithmes d'IA moulinent les avis et attribuent un score pour notre plaisir et celui des employeurs. En collectant et diffusant les avis et les score, le système *fait savoir* et *donne du pouvoir*. Dans certains cas, il permet de punir ou de récompenser automatiquement, comme chez Uber.

L'épisode *Nosedive (Chute libre)* de la série *Black Mirror* illustre plaisamment les effets de ce type de contrôle social. Dans un monde dystopique, chacun note les autres de 0 à 5 en fonction de la qualité de son interaction avec eux. Lacie, l'héroïne, cherche à augmenter son statut social pour accéder à la maison de ses rêves. À la suite d'une bévue, rien ne se passe comme prévu. Comme sa note baisse, son statut social baisse aussi, ce qui lui fait perdre des relations « intéressantes » et des privilèges (en matière de logement ou transport). Chaque baisse de note entraîne ainsi la suivante. Elle devient une paria et bientôt une délinquante. Personne n'a voulu la punir. Chacun a juste noté *honnêtement* sa qualité relationnelle.

Les *Big Brothers* contemporains ont compris la leçon d'Orwell. Plus on surveille de façon douce, invisible, moins on a besoin de recourir à la force. Ils disposent avec nous d'une armée de petites mains évaluatrices, toujours prêtes à espionner, noter et signaler, toujours à l'affût comme les « yeux de l'ananas » polpotien. Nous sommes bien proche d'adopter le système chinois du crédit social, sous un vernis démocratique.

Langage et pouvoir, censure et algorithmes

Dans *1984*, Orwell développe une conception politique du langage qui va au-delà des considérations traditionnelles sur la rhétorique comme compétence politique ou moyen de manipulation. *Big Brother* fait travailler ses linguistes au perfectionnement de la *Novlangue* (le *Newspeak*) dont la fonction est normative :

¹ Jean-François Amadiou, « Quand les élèves notent leurs profs : justice ou consumérisme ? », *The Conversation*, 22 mars 2022. [[Disponible en ligne](#)]. Aux Etats-Unis, une mauvaise note peut conduire au licenciement. C'est la mésaventure arrivée à un professeur chimie réputé de l'Université de New York, à la suite d'une pétition de ses étudiants accusant ses cours jugés trop « difficiles » d'être responsables de leurs mauvais résultats aux examens. (« Méthodes dépassées ou nivellement par le bas ? Un professeur d'université viré pour des corrections trop sévères », *La Libre Belgique*, 11 octobre, 2022. [[Disponible en ligne](#)] »).

Le but de la novlangue n'était pas seulement de fournir un moyen d'expression pour la vision du monde et les habitudes mentales propres aux dévots de l'angsoc [*socialisme ang-lais*], mais de rendre tous les autres modes de pensée impossibles.¹

Il s'agit de rendre sinon impensable, au moins inexprimable, toute pensée contraire à l'idéologie du régime. Le principal procédé principal consiste à supprimer du vocabulaire tous les « mots indésirables », de sorte que les idées critiques qui pourraient toujours germer dans un esprit, ne le puissent « que sous une forme vague, inexprimable en mots » (1984), impossible à élaborer ou communiquer. Cette conception est proche du mythe linguistique, popularisé sous le nom d'« hypothèse Sapir-Whorf »², d'après lequel la structure d'une langue suffirait à *elle seule* déterminer la pensée. Ce n'est pas celle d'Orwell. Dans le roman, la Novlangue résulte d'une politique planifiée. Le pouvoir est nécessaire pour l'imposer et la maintenir en chassant et réprimant les usages qui s'en écartent.

À défaut créer une novlangue, les autocrates contemporains s'emploient à traquer eux aussi l'utilisation des « mots indésirables » pour les faire disparaître des sites officiels, de la presse, des réseaux sociaux. Cette censure à grande échelle n'est possible que par le Big Data et les algorithmes d'IA.

Dans la **Russie de Poutine**, il n'y a pas de « hausse des prix », seulement une « correction des prix », pas d'« opposants » seulement des « agents étrangers », pas de « guerre » en Ukraine mais une « opération militaire spéciale », pas de « déportation de civils » mais des « opérations de sauvetage » ; on ne doit pas parler de « militaires ukrainiens » mais d'« ukronazis » ou de « terroristes », de « mercenaires russes » mais de « héros de la Grande Russie ».³ Tout accès par Internet à une pensée « dissidente » est bloqué par les algorithmes d'IA dédiés à la censure.

La **Chine de Xi Jinping** utilise des algorithmes visant des mots-clés ou des locutions afin de censurer l'expression ou l'accès à des pensées contestataires. Par exemple : « Quitter » ou « renverser le PPC » (le Parti Communiste chinois), « Tibet libre », « Dalaï-Lama », « Pour l'amour d'Allah » (et ses nombreuses variantes) ainsi que tout ce qui se rapporte aux « Ouïghours » (peuple de chinois musulmans), « dictature », « Tiananmen », « 4 juin » (date des manifestations de Tiananmen), « Ferme des animaux » et « 1984 » (pour des raisons évidentes), « Xi-Zedong »

¹ Orwell, 1984, Appendice : « les principes de la Novlangue ».

² C'était aussi un lieu commun du cours de philosophie avec l'exemple des 75, 100 ou 200 mots (supposés) pour parler de la neige chez les Inuits. Sur ce mythe, voir Geoffrey Keith Pullum, *The Great Eskimo Vocabulary Hoax and Other Irreverent Essays on the Study of Language*, University of Chicago Press, 1991, pages 159-171. [[Disponible en ligne](#)]

³ Voir Lidia Likhacheva, *De la création lexicale au « novlangue » russe : lorsque le monde d'Orwell devient réalité*, 2023. [[Disponible en ligne](#)]

IA et politique. Entre 1984 d'Orwell et le Procès de Kafka

(contraction de Xi Jinping + Mao Zedong), « Winnie l'ourson » (dont le physique est comparé à celui de Xi Jinping)¹.

Depuis la prise de fonctions de **Donald Trump**, des centaines de mots et de locutions ont été bannis des communications officielles de l'administration américaine. Le 7 mars dernier, le *New York Times*² publiait la liste d'environ 200 mots « déconseillés » par l'administration Trump pour l'ensemble des acteurs publics des États-Unis, sans distinction. La censure idéologique vise la Justice sociale (égalité sociale, égalité des chance, discrimination, stéréotypes, exclusion, sont proscrits), la diversité et les minorités (ethnique, diversité / mixité ethnique, multiculturel, racisme, ségrégation), l'égalité homme-femme, l'orientation sexuelle, l'écologie, etc.

La purge des mots « inappropriés » vise un usage « politiquement correct » de la langue, c'est-à-dire conforme aux valeurs et à l'idéologie des censeurs du moment. Les régimes totalitaires semblent partager la métaphysique qui préside à la création de la novlangue : « Nous contrôlons la matière parce que nous contrôlons l'esprit. La réalité est à l'intérieur du crâne », « Tout ce qui se passe est dans l'esprit. Et tout ce se passe dans l'esprit, se passe effectivement » (1984). Cette métaphysique fonde aussi l'idée de « vérité alternative » et la pratique des « fake-news ».

II. Automatisation des décisions. Vers un monde kafkaïen ?

Être accusé sans savoir pourquoi, ni par qui

Quelqu'un a dû raconter des histoires sur Joseph K., car un matin, sans avoir rien fait de mal, il a été arrêté. (Kafka, *Le Procès*)

Je suppose que l'ordinateur s'est trompé. (Joseph K. au 21e siècle)³

Dans *Le Procès*, Joseph K. reçoit un beau matin la visite inattendue de trois agents qui l'informent de son d'arrestation. K. ne comprend pas pourquoi et compte bien se défendre. Confronté à un appareil bureaucratique aussi absurde que proliférant, il ne connaîtra jamais les motifs de l'accusation, sera jugé, puis exécuté « comme un chien ».

En janvier 2020, la police de Détroit menotte Robert Williams sur sa pelouse devant sa femme et ses deux filles terrifiées, âgées de 2 et 5 ans. Il est conduit

¹ Lotus Ruan, Jeffrey Knockel, Jason Q. Ng, and Masashi Crete-Nishihata, *One App, Two Systems How WeChat uses one censorship policy in China and another internationally*, Université de Toronto, 2016. [[Disponible en ligne](#)].

² Karen Yourish, Annie Daniel, Saurabh Datar, Isaac White, Lazaro Gamio, « The Words Federal Agencies Are Discouraged From Using Under Trump », *New York Times*, 7 mars 2025.

³ Épigraphe de l'Introduction du livre de Mark Coeckelbergh, *The Political Philosophy of AI : An Introduction*, Cambridge, UK: Polity Press, 2022. (j'ai simplement inversé les deux phrases).

en centre de détention. On prend ses empreintes digitales, un échantillon d'ADN et sa photo pour l'identité judiciaire. La police refuse de lui dire pourquoi il est arrêté. Il passe sa nuit dans une cellule crasseuse et surpeuplée. L'après-midi suivant, on lui présente des images de vidéo-surveillance montrant un Noir volant des montres dans un magasin de luxe. Aux inspecteurs qui lui demandent « *C'est vous ?* », il répond : « *Non. Vous pensez que tous les Noirs se ressemblent ?* ». Trente heures après son arrestation, il est relâché. L'IA de reconnaissance faciale a produit un faux positif. Un inspecteur reconnaît que « l'ordinateur s'est trompé ». Robert Williams commente : « *Je n'aurais jamais pensé devoir expliquer à mes filles pourquoi leur papa avait été arrêté. Comment expliquer à deux petites filles qu'un ordinateur s'est trompé, mais que la police l'a quand même suivi ?* »¹.

Pour Daniel Solove, le *Procès de Kafka* saisit mieux que *1984*, la nature et les effets des relations de pouvoir créées par les bases de données et la manière dont on traite les individus :

Ce que le *Procès* illustre, c'est que le pouvoir ne s'exerce pas uniquement sous des formes totalitaires, et que les relations de pouvoir déséquilibrées avec des bureaucraties peuvent avoir des effets mortifères sur les individus, quels que soient les objectifs de ces bureaucraties.²

Certes, aux États-Unis, la plupart des gens ne se voient pas signifier qu'ils sont arrêtés sans raison apparente et ne s'attendent pas à être exécutés de manière inattendue un soir... [Mais] *Le Procès* rend compte du sentiment d'impuissance, de frustration et de vulnérabilité que l'on éprouve lorsqu'une grande organisation bureaucratique dispose d'un dossier détaillé sur votre vie. À tout moment, quelque chose peut arriver à Joseph K. ; les décisions résultent des données du dossier, Joseph K. n'est pas informé, il n'a pas son mot à dire et n'a aucun moyen de se défendre. Il est complètement à la merci du processus bureaucratique.³

Daniel Solove souligne cependant que les « métaphores de Kafka et d'Orwell ne s'excluent pas mutuellement ». Être surveillé et s'auto-censurer n'est qu'un aspect du problème. L'autre aspect est le remplacement du jugement humain par des IA prédictives. C'est là qu'Orwell et Kafka se rencontrent.

¹ Il raconte toute l'épreuve dans un éditorial publié par le *Washington Post* (« I was wrongfully arrested because of facial recognition. Why are police allowed to use it ? » - [[En ligne](#)]) et une vidéo publiée par l'*American Civil Liberties Union* (ACLU) [[Disponible en ligne](#)]

² Daniel Solove, *The Digital Person. Technology and Privacy in the Information Age*, NYU Press, 2004, page 38.

³ Même ouvrage, page 41.

Les algorithmes de sécurité

Les services de police aux frontières et de sécurité aéroportuaire utilisent les techniques de reconnaissance faciale, à côté des techniques biométriques (scan de l'iris, empreintes digitales). Pour faire du profilage sécuritaire, les gouvernements exploitent leurs propres données et celles du secteur privé (collectables sur les réseaux sociaux, ou exigibles de certaines entreprises, les banques par exemple, au nom de la sécurité). L'IA élabore des modèles de comportements délictueux et criminels à partir des données agrégées des individus condamnés, même pour une simple infraction. Une fois le modèle testé, les systèmes d'IA prédictives peuvent calculer le degré *supposé* de dangerosité ou de criminalité d'une personne. Ces IA sont réellement performantes. Et quel voyageur souhaiterait la suppression du contrôle de sécurité des passagers de l'avion dans lequel il embarque ? Le problème des systèmes d'IA prédictives est qu'elles vont bien au-delà de la détection d'un criminel connu ou recherché. Elles indiquent la dangerosité *potentielle* de tout voyageur, en fonction de « *profil* » qu'elle lui attribue. Le point commun à ces IA est leur fragilité. D'abord, parce que la décision (refuser l'accès à l'embarquement par exemple), ne résulte pas d'un raisonnement causal à partir des données factuelles et individuelles du moment, mais d'une *prédiction* de comportement reposant sur des corrélations statistiques et l'algorithme utilisé (le « *profil* » n'est qu'une prédiction statistique, non le portrait de l'individu réel). Ensuite, parce qu'un grain de sable suffit à faire dérailler ces systèmes : une erreur dans les données, ou une faille du modèle comme dans le cas de Robert Williams (et toujours possible malgré leur efficacité générale). Cathy O'Neil, mathématicienne et data scientist américaine, souligne ce danger :

Si des erreurs apparaissent dans un dossier - et c'est souvent le cas -, les algorithmes même les mieux conçus prendront la mauvaise décision. Comme le disent de longue date les professionnels du traitement de données : « *garbage in, garbage out* » - si ce qui entre est faux, ce qui sort sera faux.

L'individu qui fait les frais de ce processus automatisé peut en endurer les conséquences pendant des années. Les listes d'interdiction de vol sur des compagnies aériennes pour terrorisme, générées par ordinateur, sont ainsi connues pour regorger d'erreurs. Une personne innocente dont le nom ressemble à celui d'un terroriste présumé affrontera une épreuve infernale chaque fois qu'elle devra monter dans un avion.¹

Les conséquences de ces erreurs sont graves. Des personnes peuvent se voir refuser le droit de prendre l'avion ou d'entrer sur un territoire, sans raison ni recours. D'autres, peuvent se voir détenues pour un temps indéfini, sans pouvoir joindre un avocat ni connaître la raison de leur détention. D'autres encore,

¹ Voir le chapitre « Dommages collatéraux » dans le livre de Cathy O'Neil, *Algorithmes la bombe à retardement*, Arènes, 2018.

peuvent être arrêtées ou accusées d'un délit ou d'un crime qu'elles n'ont pas commis. L'IA menace ainsi le principe constitutif des démocraties, à savoir la garantie des droits civils et libertés fondamentales de la personne. Les erreurs entraînant une violation des droits des personnes sont si fréquentes et attendues que, l'UNICRI (*United Nations Interregional Crime and Justice Research Institute*) et INTERPOL ont dû publier un guide à destination des services de police « *pour promouvoir une utilisation responsable de l'intelligence artificielle* »¹. L'euphémisme « responsable » rappelle que *l'intelligence* dite *artificielle* ne l'est pas tant cela, qu'elle commet toujours des erreurs « stupides », et que le jugement humain reste absolument nécessaire pour l'évaluation des situations.

Les algorithmes de surveillance des bénéficiaires d'aides sociales

Les algorithmes de détection des fraudes aux aides sociales sont utilisés dans de nombreux pays, et partout critiqués. Le cas français de la Caisse nationale des allocations familiales est emblématique des dérives de ces IA prédictives (j'utiliserai l'acronyme CAF plutôt que CNAF moins usuel).

En 2024, 29 millions de contrôles *automatisés* ont été réalisés. On n'a jamais fait mieux. Un tel volume de contrôle n'est réalisable que par des IA capables d'exploiter des mégadonnées (*Big Data*) et de les croiser par la technique du *data mining*. Ce succès devrait réjouir, après tout, les fraudeurs mettent en péril le système social. Ce n'est pourtant pas le cas. L'IA prédictive de la CAF souffre des mêmes biais et failles que tous les programmes de ce type. Les problèmes sont si graves et les témoignages si accablants² que quinze associations, fondations et syndicats ont porté l'affaire devant le Conseil d'État³.

L'algorithme est discriminatoire, il cible les personnes pauvres et vulnérables. Les paramètres de l'algorithme conduisent à présumer systématiquement comme fraudeurs certaines *catégories* d'allocataires : les personnes à revenus faibles, au chômage, au RSA, les personnes handicapées et les familles monoparentales⁴. L'algorithme est neutre : une équation mathématique n'a pas

¹ La présentation du *Manuel à l'usage des services chargés de l'application de la loi pour une innovation responsable en matière d'intelligence artificielle* en a été faite en 2023 [Accès disponible en ligne à la [présentation du manuel \(2023\)](#) et à sa [version révisée \(2024\)](#)].

² L'association *Changer le Cap* en a publié un certain nombre sur son site, [disponible en ligne](#).

³ Le Dossier de presse de ces associations est [disponible en ligne](#). Pour un résumé substantiel voir l'article de Didier Dubasque et le résumé des arguments juridiques sur son site « Quinze organisations dont l'ANAS attaquent l'algorithme de notation de la CAF devant le Conseil d'État » (22 octobre 2024). [[Disponible en ligne](#)].

⁴ La démonstration en a été faite en 2023 par l'association *La Quadrature du Net*. Elle a révélé le 27 novembre 2023 les résultats de son travail dans un article « *Notation des allocataires : l'indécence des pratiques de la CAF désormais indéniable* » qui secoué les institutions et le monde politique. [[Disponible en ligne](#)]. Une enquête conjointe *Le Monde-Lighthouse Report* a confirmé « les dérives de l'algorithme » des CAF, le profilage et les

d'intention, et la machine (le hardware) n'a pas d'âme ; on peut aussi créditer les ingénieurs de ne pas avoir intentionnellement *biaisé* le programme pour persécuter les plus faibles. La limite de cette « neutralité » algorithmique est qu'un programme prédictif de ce genre ne peut être que « discriminatoire », parce que statistiquement les bénéficiaires de la CAF sont majoritairement précaires, et que plus la précarité est grande plus le risque statistique de fraude augmente. L'accroissement des contrôles des catégories ciblées crée aussi une boucle de rétroaction pernicieuse. Elle aboutit à détecter plus de fraudes dans ces populations, et les nouvelles données nourrissent le système, ce qui « justifie » un renforcement de la surveillance de ces mêmes populations. Fabriquer un système plus juste, impliquerait d'introduire des règles correctives, c'est-à-dire des *biais* pour corriger les *biais* résultant des corrélations statistiques du modèle¹. Mais est-ce vraiment ce que veulent les gouvernements ? Quoi qu'il en soit, c'est à raison que de nombreuses associations ont dénoncé « la chasse aux pauvres » qu'entraînent ces « machines à soupçon », et l'assimilation entre précarité et fraude.

Décisions automatisées, inhumanité des contrôles. L'algorithme de la CAF décide donc qui sera contrôlé. Les personnes « à risque » subissent une surveillance accrue, une multiplication des contrôles « sur pièces » et « sur place » (à domicile). L'automatisation des décisions, des notifications, l'impossibilité d'avoir un agent – et non pas un « agent conversationnel » ou une réponse « rédigée par une IA », l'impossibilité d'avoir des réponses claires et rapides, génèrent une angoisse permanente. Des allocataires perdus dans la complexité des démarches préfèrent renoncer à leurs droits sociaux ; d'autres, traumatisés par les contrôles à répétition et une incrimination abusive, abandonnent tout recours.

L'enfer des dettes artificiellement créées par le robot de la CAF. Les personnes incriminées à tort voient souvent leurs prestations suspendues, et

discriminations : « Profilage et discriminations : enquête sur les dérives de l'algorithme des caisses d'allocations familiales », *Le Monde-Lighthouse Report*, 4 décembre 2023 [[En ligne](#)] Voir aussi l'animation numérique « Comment l'algorithme de la CAF prédit si vous êtes « à risque » de frauder », *Le Monde-Lighthouse Report*, 4 décembre 2023 [[Disponible en ligne](#)]

¹ Mais, l'idée d'une IA « éthique » est de l'ordre du mythe. Elle soulève des problèmes insolubles. Aucun principe éthique, parce qu'il est un *principe*, n'est traductible en ligne de code ou instruction d'un programme informatique ; les principes renvoient à des notions bien trop générales (autonomie, bienfaisance, dignité, égalité, justice). Ensuite, il faudrait ensuite établir la liste de ces principes et leur interprétation : qui décidera ? L'IA éthique sert d'argument à ceux qui ne veulent aucune régulation étatique - les GAFAM objectent qu'ils ont déjà tous leurs comités d'éthiques internes. Si on ne peut coder un principe éthique, on peut facilement *vérifier* si un programme aboutit à des injustices, des discriminations, des atteintes aux droits des personnes, etc. C'est pourquoi la régulation politique des usages de l'IA est en revanche tout à fait possible. Sur l'éthique et l'IA voir l'analyse critique de Gabriel Ganascia dans son livre *Servitudes virtuelles*, Seuil, 2022.

l'exécution automatisée de la demande de remboursement, par saisie-arrêt sur leur compte bancaire¹. Des allocataires brutalement privés du RSA ou d'autres prestations, accumulent des dettes. S'ils saisissent le Tribunal administratif, il faudra deux années avant que la CAF ne soit contrainte à les rétablir dans leurs droits. Agnès Callamard, secrétaire générale d'*Amnesty International*, alerte sur les conséquences déjà observées :

Nous avons vu ce qui s'est passé aux Pays-Bas en 2021, où le fisc a utilisé un algorithme discriminatoire de notation des risques pour détecter la fraude aux allocations familiales. Des milliers de familles ont été plongées dans la dette et la pauvreté. Une situation similaire risque de se produire en France, et ce risque doit être atténué de toute urgence puisque des préjudices sont déjà survenus.²

La violation des règles de Droit garantissant les droits et libertés des citoyens. La CAF ignore les exigences de *transparence* et de *motivation* qui voudraient qu'une décision de suspension d'allocation soit *toujours expliquée*, avec mention *explicite de l'article de loi* appliqué. La CAF use le plus souvent de notifications vagues comme « Vos droits ont changé » ou « Nous avons recalculé vos prestations ». Comme dans le *Procès* ou le *Château* de Kafka, la décision tombe sur l'allocataire qui cherche en vain à comprendre où est sa faute ou l'erreur de l'administration. Didier Minot, ancien Haut fonctionnaire, président du *Changer de cap* témoigne :

Dans le meilleur des cas, la personne reçoit une notification d'indu assez sommaire. On a étudié par exemple le cas d'une femme qui a reçu un document dans lequel on lui disait : "Suite au contrôle de votre dossier, nous avons constaté sur vos relevés bancaires que vos grands-parents vous aident régulièrement depuis octobre 2020, or ces sommes n'ont pas été déclarées. **Par conséquent nous vous inscrivons comme fraudeuse.**" [...] Il est indiqué ensuite que la charge de la preuve lui incombe et, en caractères gras, que la loi rend passible d'amende ou d'emprisonnement quiconque se rend coupable de fraude ou de fausse déclaration. Les neuf-dixième des gens sont démunis face à ces documents et n'ont pas les moyens de les contester.³

Saisie, la Défenseure des droits a rendu un avis très sévère contre la CAF qu'elle enjoint de « se conformer aux dispositions du code des relations entre le public et

¹ Voir le témoignage sur son Blog de la journaliste Lucie Inland, « Comment la CAF a traité mon dossier avec un « robot à dettes », avril 2021 [[Disponible en ligne](#)]. Publié initialement par *Algorithm Watch* (« How French welfare services are creating "robo-debt" ») qui précise que « *Au cours des cinq dernières années, plusieurs scandales ont montré l'ampleur du problème. En Australie, 400.000 personnes se sont retrouvées en « robot-dette», 40.000 dans le Michigan et 26.000 aux Pays-Bas.* » [[Disponible en ligne](#)]

² *Amnesty International*, « France : l'algorithme de la Caisse nationale des allocations familiales cible les plus précaires », 16 octobre 2024 [[Disponible en ligne](#)]. Et suite à l'audit gouvernemental, la ville de Rotterdam a suspendu le système en 2021.

³ Voir l'enquête très précise de Radio-France « Quand des algorithmes notent les allocataires de la Caf », 9 décembre 2022. [[Disponible en ligne](#)]

l'administration en motivant *en fait* et *en droit* les décisions qu'elle notifie aux allocataires », à « ne pas suspendre les prestations en l'absence d'obstacle au contrôle caractérisé de la part de l'allocataire », à lui fournir « les informations nécessaires afin de pouvoir répondre à un tel contrôle » en « notifiant une décision administrative motivée à l'allocataire » (décision du 26 juin 2024, [disponible en ligne](#)). En clair, la CAF est sommée de respecter l'État de Droit.

Joseph K. se plaignant de la CAF à l'ère de *Services Publics+*

La plateforme numérique *Services Publics+*, a été lancée par le Gouvernement pour rendre les services publics plus proches, plus simples et plus efficaces, pour faciliter le contact avec un agent et obtenir des réponses rapides. La plateforme est utile pour obtenir des informations de base. Mais dès qu'un usager rencontre un problème précis, elle devient inutile et même nuisible. Tenons-nous au cas de l'*Agence de recouvrement et d'intermédiation des pensions alimentaires* (Aripa), liée à la CAF. Un parent qui doit verser une pension doit le faire directement à la CAF. Le service Aripa s'occupe ensuite du versement au bénéficiaire ainsi que de la récupération des pensions impayées. Formidable ! Sauf qu'en réalité, cela se passe souvent mal. Commence alors un dialogue « 2.0 ».

Une allocataire, mère de deux enfants, se plaint d'être restée 9 mois sans pension alimentaire à cause des délais et de l'oubli du rattrapage lors de la régularisation. Elle décide d'écrire à « Services publics+ » :

On fait face à un parent qui ne paye pas et à un service d'intermédiation qui ne paye pas, mais nous on élève nos enfants et on doit les nourrir tous les jours ! C'est inadmissible ! Pourquoi un dossier déjà en place met des mois à être "transférer" ? Je m'interroge. Cette situation est très chronophage et anxiogène. C'est une violence administrative. Depuis que l'Aripa gère on est laissé à l'abandon !

Elle signe « Maman de deux ados ». La réponse est la suivante :

Bonjour Maman de 2 ados, [sic]

Nous avons pris connaissance de votre expérience concernant l'intermédiation des pensions alimentaires et nous sommes désolés pour les difficultés que vous avez rencontrées. (...) Il est important de signaler tout impayé directement à l'Aripa pour enclencher le processus de recouvrement (...) vous pouvez vérifier le suivi avec l'Aripa (...) Nous vous encourageons à vous connecter sur leur plateforme en ligne (...).

Nous vous remercions d'avoir partagé votre situation, cela nous préoccupe et nous permet d'améliorer constamment nos services.

Cordialement.

Ibtissem

Cette réponse a été générée par une IA à 99 % et vérifiée par un agent.¹

¹ Réponse du 09/09/25. [[Disponible en ligne](#)]

La réponse rédigée à 99% de l'IA renvoie cette mère au service défaillant et la remercie de contribuer « à améliorer constamment [leurs] services ». La bêtise et l'inutilité de la réponse sont patentes. Les réponses de ce type sont légion¹. Ce qui n'est pas étonnant puisqu'une IA n'a aucune compréhension des questions qu'elle reçoit, ni des réponses qu'elle renvoie. La réponse est automatiquement générée par un algorithme de complétion statistique². D'où le renvoi des allocataires à des démarches qu'ils ont, pour la plupart, déjà effectuées sans succès.

Une autre allocataire se plaint qu'il faille « pratiquement un an pour faire réviser la pension en cas de changement de situation », alors que « dans l'autre sens, il suffit juste d'un courrier pour prélever aujourd'hui plus de 50% de [ses] revenus », que « la CAF ne répond pas aux mails », etc. Elle reçoit elle aussi une pseudo-réponse comprenant la mention « cette réponse a été **générée par une IA à 89 %** et vérifiée par un agent ». Mécontente, elle livre son « retour d'expérience » :

Une réponse de leur part rédigée à 89 % par l'IA et vérifiée par un agent ?! Là, ils ont le temps de s'en occuper. Tout ça pour dire qu'il faut les contacter en reconnaissant que c'est long. Que je peux m'opposer à la procédure en allant devant une juridiction...

Bref... autant annoncer : "Réponse rédigée à 100 % par l'IA et non vérifiée par un agent."

On comprend la demande implicite : avoir affaire à « un vrai agent » qui considère sa requête avec intelligence et compréhension de sa situation. Là où un agent humain aurait porté plus d'attention à sa demande initiale et à son mécontentement face à la réponse « automatisée », le système génère la réponse suivante :

Nous vous conseillons de persévérer dans vos démarches via les canaux disponibles. Nous apprécions votre retour qui nous aide à améliorer nos services, et vous remercions de votre compréhension.

Cordialement,

Nicolas, pour Caf de l'AISNE

Cette réponse a été **générée par une IA à 100 %** et vérifiée par un agent.³

Ce qui s'appelle se moquer de monde. Quelle est donc la finalité d'une Plateforme comme « Services Publics+ », qui se vante sur sa page d'accueil de posséder toutes les qualités qui lui font défaut ? La plateforme apparaît comme écran supplémentaire entre l'administration et l'utilisateur, elle complique inutilement les

¹ Voir par exemple, la réponse du 29/09/2025 [[Disponible en ligne](#)]. Il suffit de taper dans le moteur de recherche du site le nom de ce seul service « Aripa », pour constater le nombre d'allocataire dans des situations dramatiques qui en reçoivent en réalité aucune réponse.

² Pour savoir « Comment fonctionne un agent conversationnel ? », voir article « Melanie Mitchell, L'avenir de l'intelligence artificielle (II-1) ». [[Disponible en ligne](#)]

³ Réponse du 18/04/2025. [[Disponible en ligne](#)]

démarches et suscite de la frustration puisqu'elle est inefficace. Dans le même temps, l'enquête du Défenseur des droits, publiée le lundi 13 octobre 2025, montre que le taux de Français en difficulté dans les démarches administratives à cause du passage au numérique a explosé en 8 ans, passant de 39% à 61%.¹

L'IA de la CAF, un exemple de monde orwello-kafkaïen

Les bénéficiaires d'aides sociale sont l'objet d'une *surveillance constante* couplée à un *enfer bureaucratique*. La surveillance est disproportionnée. L'algorithme de la CAF analyse chaque mois les données personnelles de plus de 32 millions de personnes, soit près de la moitié de la population française. Suspectés de fraude, les allocataires doivent remplir des montagnes de papiers, répondre à d'innombrables questions, fournir 10 fois le même document, et subir des contrôles inquisitoires, numériques ou à domicile². La CAF refuse dans le même temps de communiquer le code source de son algorithme actuel, empêchant tout contrôle indépendant et tout débat public informé sur cette utilisation.

La généralisation des décisions algorithmiques

Les décisions algorithmiques concernent un nombre croissant de secteurs, très importants dans la vie des individus. Par exemple, les entreprises confient désormais à des algorithmes l'analyse automatisée :

- de votre demande de prêt bancaire ou d'assurance, pour décider de sa recevabilité selon votre profil ;
- du CV qui accompagne votre demande d'emploi, complété par toutes les données numériques disponibles sur Internet, pour vous retenir ou écarter votre candidature

Tous ces algorithmes d'IA posent les mêmes problèmes : de fiabilité et de pertinence du traitement (tous produisent des corrélations statistiques réductrices et non pas des analyses causales) ; d'impartialité et de neutralité en raison des biais ; d'opacité des décisions puisque le lien entre les données d'entrée et les résultats est impossible à connaître et de quasi-impossibilité de les contester ; de déresponsabilisation des employés ou agents qui, incapables de comprendre et d'expliquer le processus qui conduit à la décision, ne peuvent la justifier que par une phrase du type « c'est l'ordinateur qui a décidé ».

¹ *Enquête sur l'accès aux droits*, 2^{ème} édition, octobre 2025. [[Disponible en ligne](#)]

² En témoignage, par exemple, l'article d'Isabelle Rey-Lefebvre dans *le Monde* du 20 février 2021 titré « "L'entretien avec l'agent de la CAF a été une humiliation" : les bénéficiaires du RSA dans l'enfer des contrôles », qui relate, *entre autres*, les interrogations d'un contrôleur sur l'usage de 20 € retirés en espèces.

On trouvera de nombreux exemples du danger des systèmes d'IA prédictives dans divers domaines, notamment celui de la justice, dans le livre de Cathy O'Neil, *Algorithmes la bombe à retardement* (Arènes, 2018).

Conclusion

Il est tout fait possible d'aborder l'IA dans le cadre d'un cours de philosophie politique. Les programmes de notions sont tellement vagues que l'enseignant a l'embarras du choix : La Justice, L'État, La Liberté, La Raison, La Technique, La Science. Même si l'IA a renouvelé les manières de surveiller et de punir, les analyses de Foucault et d'Orwell restent pertinentes.

On peut aussi se servir de références plus classiques. Par exemple, passer diverses utilisations de l'IA au crible des notions de « liberté négative » et de « liberté positive », telles qu'Isaiah Berlin les définit, mais indépendamment de ses positions à ce sujet¹ :

[Liberté négative] Je suis libre, dit-on généralement, dans la mesure où personne ne vient gêner mon action. En ce sens, la liberté politique n'est que l'espace à l'intérieur duquel un homme peut agir sans que d'autres l'en empêchent. Si d'autres m'empêchent de faire ce qu'autrement j'aurais fait, je ne suis pas entièrement libre ; et si cet espace se trouve réduit en dessous d'un certain minimum, on peut dire que je suis contraint, opprimé et peut-être même asservi. [...] Plus vaste est cette aire de non-ingérence, plus étendue est ma liberté.

Isaiah Berlin, *Éloge de la liberté* (1969), Calmann-Lévy, 1988, pages 171-172.

Le sens « positif » du mot liberté découle du désir d'un individu d'être son propre maître. Je souhaite que ma vie et mes décisions dépendent de moi, et non de forces extérieures quelles qu'elles soient. Je désire être l'instrument de ma propre volonté, et non celui de la volonté des autres ; je désire être un sujet et non un objet ; être mû par des raisons et des mobiles conscients qui soient les miens, et non par des causes, pour ainsi dire extérieures. Je désire être une personne et non personne, décider et non qu'on décide à ma place, agir et non être agi par la nature ou par d'autres hommes, comme si j'étais une chose, un animal ou un esclave, incapable d'assumer sa condition d'homme, autrement dit, incapable de concevoir et de réaliser des fins et des stratégies qui lui soient propres.

Isaiah Berlin, *Éloge de la liberté* (1969), Calmann-Lévy, 1988, page 204.

¹ Voir dans Mark Coeckelbergh, déjà cité, le chapitre 2, « Freedom : manipulation by AI and robot ».

Le paternalisme est despotique, non pas parce qu'il opprime davantage que la tyrannie à l'état pur, brutale et obscurantiste, non parce qu'il ignore la raison transcendante qui réside en moi, mais parce qu'il est une insulte à la conception que j'ai de moi-même comme homme qui entend conduire sa vie selon ses fins propres (lesquelles ne sont pas forcément rationnelles ou bien intentionnées) et qui, surtout a le droit d'être reconnu comme tel par les autres. [...]

Manipuler les hommes, les pousser vers des fins que moi – le réformateur social je discerne – mais eux pas forcément, revient à nier leur essence, à les traiter comme des objets sans volonté propre, en un mot à les avilir. En effet, leur mentir, les tromper, les utiliser comme des moyens pour mes propres fins, même si c'est pour leur bien, revient en fait à les traiter comme des sous hommes, comme si leurs fins étaient moins ultimes et moins sacrées que les miennes. Au nom de quoi puis-je les contraindre à faire ce qu'ils ne veulent pas ? Seulement au nom d'une valeur plus sublime qu'eux.

Isaiah Berlin, *Éloge de la liberté* (1969), Calmann-Lévy, 1988, pages 204-205.

La « **liberté négative** » s'oppose à l'oppression, à la contrainte, ainsi qu'à tout paternalisme (au sens classique du terme), à toutes les ingérences et immixtions dans la vie privée ou sa conduite de la part des États et, par extension aujourd'hui, des géants du numérique (Isaiah Berlin ne connaissait pas les GAFAM !). Elle est aussi impliquée dans tous les cas de violations des droits fondamentaux : discriminations, privations illégitimes de droits, restrictions indues de circulation, enfermement arbitraire, condamnations sans procès et impossibilité de se défendre.

La « **liberté positive** » s'oppose au paternalisme du « Nudge » (« coup de pouce »). Le Nudge est une technique qui utilise les biais cognitifs mis en évidence par la psychologie comportementale, pour « aider » les citoyens à adopter les bons comportements. Le Nudge est la plupart du temps bien intentionné¹. Son utilisation en politique fait question. D'abord, parce qu'il repose sur l'idée qu'il vaut mieux manipuler psychologiquement les citoyens plutôt que de perdre son temps à discuter ou argumenter pour les convaincre des bienfaits d'un changement de comportement, ce qui revient à renoncer à la politique, ou à l'idéal d'une démocratie participative fondée sur le débat et la délibération rationnelle. Deuxièmement, parce qu'il ne traite pas l'individu comme un sujet capable d'autonomie critique, mais comme un être perclus de biais cognitifs, un être irrationnel et

¹ Par exemple, à Nantes, des poubelles de tri ont été peintes par des street-artistes pour sensibiliser les habitants à la nécessité de trier leurs ordures ménagères. L'exemple canonique est celui de la mouche dessinée au fond d'un urinoir, pour éviter les éclaboussures à l'extérieur de la cuvette en incitant à viser juste.

infantile, ce qui justifie de le traiter comme un enfant¹. Enfin, parce que les gouvernements comprennent si bien l'intérêt du Nudge qu'ils l'utilisent pour forcer « gentiment » l'acceptation de certaines réformes (voir le livre d'Audrey Chabal cité ci-dessous), ou fabriquent des « nudges » négatifs (les « sludges »²), par exemple, pour dissuader des usagers d'accomplir certaines démarches en augmentant le coût cognitif et psychologique.

Sur le Nudge en politique, on pourra lire :

1° L'article important que Gerald Dworkin consacre au « Paternalisme » dans la *Stanford Encyclopedia of Philosophy*³. Il y discute notamment le « paternalisme libertaire » (ou « soft ») dont se réclament Cass Sunstein et Richard Thaler, les théoriciens du Nudge⁴.

2° L'enquête d'Audrey Chabal, *Souriez, vous êtes nudgé. Comment le marketing infiltre l'État*, Éditions du Faubourg, 2021. Petit livre pédagogique avec de nombreux cas documentés (le paternalisme libertarien pour gérer la crise Covid) ou la face sombre du Nudge politique (chapitre 3. Sur la « pente savonneuse des nudges »).

3° « Paternalisme digital ? L'apothéose de la gouvernementalité », cinquième chapitre du livre de Philippe Huneman, *Les sociétés du profilage. Évaluer, optimiser, prédire*, Payot, 2023. Le livre est stimulant et bien documenté, même si son propos parfois un peu rapide mérite discussion.

Serge Cospérec

¹ À l'exemple du monstre « poubellator » des trains Ouigo et son adresse aux passagers : « j'adore vos déchets », ou de la fausse mouche d'urinoir (cf. note précédente) censée déclencher chez l'homme une irrésistible impulsion à viser juste pour l'anéantir.

² Pour une explication philosophico-pédagogique du sludge, voir le bref article de Nicolas Duffau, « Kézako. Sludge, ces obstacles qui nous engluent » sur *Philonomist* [[Disponible en ligne](#)]. Dans la lignée des « sludges », il y a les « dark patterns », ces interfaces conçues rendre particulièrement infernale l'accès de l'utilisateur à ce qu'il cherche. Une excellente vidéo sur Youtube donne comme exemple le parcours de combattant pour fermer son compte Amazon « How Dark Patterns Trick You Online » [[Disponible en ligne](#)]

³ Gerald Dworkin, « Paternalism », *The Stanford Encyclopedia of Philosophy*, 2020. [[Disponible en ligne](#)]

⁴ Cass Sunstein et Richard H. Thaler, *Nudge : la méthode douce pour inspirer la bonne décision*, Vuibert, 2^e éd., augmentée, 2022.

Au fil des numéros, Côté Philo aborde divers aspects de la culture et du métier de professeur de philosophie ; le journal constitue ainsi un instrument d'information et de réflexion régulièrement alimenté et renouvelé. Selon les livraisons, nous proposons ainsi :

- Des *Dossiers* sur des questions intéressant l'enseignement de la philosophie
- Des *Notes de lecture* à vocation pédagogique
- Des synthèses sur un champ ou un philosophe
- Des pratiques pédagogiques
- Des articles sur l'enseignement de la philosophie à l'étranger
- Des informations institutionnelles et l'éclairage qu'elles nécessitent
- Ainsi que des *Humeurs* qui parfois s'imposent...

ISSN 2971-3773