

On l'a vu, la recommandation que fait Platon de ne pas tolérer dans la cité un art qui repose sur l'imitation, la *mimésis*, ouvre la porte à bien des **abus**, dont on a vu de piteux exemples quand l'État s'est mêlé d'imposer une **utilité sociale** aux artistes : tableaux allégoriques et mélodramatiques de la Révolution française, réalisme socialiste dans la peinture et la littérature en Union soviétique.

Mais si cette recommandation n'était pas finalement dénué de sens, si on l'appliquait à un domaine que Platon n'avait pas anticipé, **l'Intelligence artificielle** ?

Il n'est pas douteux, en effet, que **l'IA repose sur l'imitation**. C'est l'accès aux immenses réserves de documents disponibles sur internet qui a permis d'entraîner les LLM : *Large language models* (**Grands Modèles de Langage** en français) ou agents conversationnels comme **ChatGPT, Claude ou Mistral**. On a pu dire ainsi que les GML n'étaient rien d'autre que des « **perroquets probabilistes**¹ » qui, ayant à disposition une banque de donnée immense, pouvaient utiliser les probabilités pour fournir des réponses existantes ou même former des phrases adaptées au contexte.



L'IA n'est ainsi rien d'autre qu'un miroir qui nous renvoie nos propres idées, sans rien y ajouter véritablement. On a pu observer ainsi que les propos sexistes et racistes fréquemment tenus par l'IA provenaient tout simplement des documents qu'on lui avait fournis !

Mais là où l'on rejoint la préoccupation centrale de Platon dans *La République*, à savoir la **justice**, c'est tout d'abord le fait que les entreprises qui ont entraîné ces GML ont procédé en exploitant des données qui ne leur appartenaient pas ! Il y a eu un véritable **pillage de données**, les unes sous copy-



Emily Bender

right (les ouvrages publiés par des éditeurs), les autres dans le domaine public (Wikipedia, par exemple), au profit d'entreprises privées qui essaient maintenant de **tirer profit** de ces données transformées. Et cela a eu lieu parce que le travail des entreprises a été **souterrain**, d'une part, et d'autres part parce que la situation est **inédite**, et que les pouvoirs publics ne savent pas comment réagir dans un tel cas.

D'autres part, il faut comprendre que les agents conversationnels issus des GML sont bien des **agents**, c'est-à-dire une instance capable de prendre des décisions. Il y a eu plusieurs cas où des personnes ont demandé des conseils à ChatGPT, par exemple, sur des choix à faire concernant leurs finances, ou leur santé mentale. Quand l'agent leur conseille de vendre ou d'acheter telle valeur, quand il leur donne des conseils sur la façon de mener leur vie (dans un cas, Chat GPT a **conseillé le suicide** à une personne dépressive !), qui est responsable des conséquences ? Et de plus en plus, les agents ne se contentent pas de conseiller, mais peuvent se voir confier des **actions directes**, de gestion de finances avec pouvoir d'achat ou de vente, et de déclenchement de dispositifs dans le monde réel, comme **les rideaux d'un apparte-**



¹ selon l'expression de la linguiste Emily Bender (2021) ; on trouve aussi la traduction « perroquet stochastique »



ment ou les freins d'une voiture... Et c'est bien là le point fondamental de la justice : si une action négative, nuisible à un individu ou à la société, est commise, **il faut que la personne qui en est responsable en subisse les conséquences**. Aucun droit n'est dépourvu de responsabilité correspondante, les droits existent en parallèle avec les devoirs.

Or, **les sociétés qui développent ces agents refusent d'être tenues pour responsables** des conséquences qui peuvent découler de pareilles actions, et **les agents eux-mêmes ne sont pas des personnes morales**, ce qui veut dire qu'on ne peut pas les traîner en justice.

On se retrouve ainsi dans une **zone grise** de la justice et de la philosophie, avec des individus qui n'en sont pas vraiment, avec des personnes dénuées de personnalité. Cela explique pourquoi les réactions face à l'IA sont si tranchées : **les uns l'adoptent sans retenue, les autres s'insurgent contre une imposture**. Cela nous ramène à un phénomène que Freud avait déjà identifié : le **unheimlich** ou inconfort que nous ressentons face à un phénomène perturbant, et en particulier lorsque **nous ne pouvons pas décider si ce que nous avons en face de nous est humain ou pas**. L'évolution nous a mis sans doute en garde, par ce biais, contre **les cadavres**, en apparence humains mais dénués de vie.



Ainsi, Platon attire notre attention sur un fait qui a davantage de conséquences que nous ne pouvions l'imaginer (et qu'il l'a imaginé lui-même, sans doute !) **La parole n'est pas neutre**. Quand quelqu'un s'exprime, il doit pouvoir être identifié : **qui parle ? et il doit pouvoir répondre de ses paroles**. L'art, on le pose comme principe, est **un cas particulier**. Déjà dans l'Antiquité les représentations théâtrales étaient considérées comme des cérémonies en l'honneur de Dionysos, le dieu des excès et de la marginalité. Il y avait une sorte de



licence, de **franchise pour les paroles prononcées dans ce qui était une sorte de carnaval**, de moment où les lois étaient suspendues. Mais aujourd'hui, **si un artiste crée un roman ou une peinture où il appelle à la haine raciale, il court le risque d'être condamné...** Les tribunaux sont appelés à se prononcer sur ces cas, parfois complexes. Par exemple, si je produis une sculpture sur laquelle je place le nom d'une marque de luxe connue, je peux être condamné pour contrefaçon, sauf s'il apparaît que je tiens un discours critique sur cette marque...



On l'a vu, ceci nous ramène à **la définition de l'art**, qui n'est pas évidente et laisse une grande part à l'interprétation.