

# Littérature, philosophie et analyse computationnelle

Entretien avec Mark Algee-Hewitt

*par Bruno Cousin*

---

**Alors que nous vivons désormais à l'âge de l'IA générative, Mark Algee-Hewitt montre comment l'informatique nous aide à expliquer la vie des concepts, des esthétiques et des genres.**

---

**Mark Algee-Hewitt** est professeur en humanités numériques au département d'anglais de l'université de Stanford. Directeur du Centre d'analyse spatiale et textuelle (CESTA) et du Laboratoire littéraire de Stanford, ses travaux appliquent des méthodologies computationnelles à l'étude d'objets culturels et esthétiques, en se concentrant particulièrement sur les schémas linguistiques dans de grands corpus de textes. Formé principalement à l'analyse de la littérature anglaise et allemande des XVIII<sup>e</sup> et XIX<sup>e</sup> siècles, ainsi qu'en informatique, ses recherches ont évolué vers l'étude des schémas cachés de l'utilisation du langage dans l'écologie médiatique contemporaine.

Algee-Hewitt a publié de nombreux articles dans des revues de sciences humaines et d'études littéraires, à la fois en tant qu'auteur solo et en tant que collaborateur sur des projets à grande échelle. Au Literary Lab, il a mené des recherches sur des sujets aussi divers que la technique du suspense pour les lecteurs de romans, la relation entre les auteurs et les écrivains dans les communautés de fan fiction en ligne, la capacité de la fiction climatique à enseigner aux lecteurs des faits

environnementaux réels à travers des récits fictifs, et l'utilisation d'un langage racialisé dans la littérature américaine. Son livre *The Afterlife of Aesthetics*, à paraître prochainement, explore l'essor des concepts esthétiques, en particulier le sublime, au XVIIIe siècle, en identifiant les schémas persistants qui régissent leur utilisation et qui subsistent même après que ces concepts sont passés de mode.

En tant que chercheur au CASBS, Mark Algee-Hewitt travaille sur un nouveau projet qui utilise la distance sémantique pour retracer le développement de concepts composés complexes, tels que « droits de l'homme » ou « économie politique », au cours des XVIIIe et XIXe siècles. Au sein du centre, il a mis au point une nouvelle méthode visuelle qui permet aux chercheurs d'identifier à la fois les configurations de mots qui représentent des idées couronnées de succès et, plus important encore, celles qui distinguent les concepts qui ont échoué. En étudiant ces points d'échec conceptuel, Algee-Hewitt soutient que nous pouvons mieux comprendre pourquoi certaines idées réussissent et pourquoi d'autres sont reléguées aux oubliettes de l'histoire.

### ***La Vie des idées : Pourquoi analyser les textes philosophiques et littéraires avec des méthodes computationnelles ?***

**Mark Algee-Hewitt :** Il y a plusieurs raisons pour lesquelles je trouve personnellement utile d'utiliser l'analyse computationnelle pour comprendre les textes, en particulier les textes littéraires, mais aussi les textes philosophiques. En gros, c'est parce que cela me permet de poser des questions différentes de celles que je pourrais poser grâce aux approches de ma formation de critique littéraire ou en tant qu'historien de la vie intellectuelle. Et ce, de différentes façons.

Par exemple, cela offre des possibilités en termes d'échelle. L'analyse computationnelle nous permet d'absorber et d'analyser des textes à une échelle bien plus grande que nous ne pourrions jamais le faire en tant que lecteurs : non pas seulement des dizaines de textes analysés simultanément, mais des centaines, des milliers, des dizaines de milliers, voire des millions de textes.

Nous pouvons examiner un phénomène particulier à travers tous ces textes et parvenir ainsi à une compréhension beaucoup plus large de ce qui se passe, par exemple, dans le champ littéraire au cours d'un siècle donné, et dans un certain domaine intellectuel et historique. Il s'agit là d'une approche révolutionnaire de la littérature, d'une approche révolutionnaire du texte en général.

Elle nous fournit un type de preuves complètement nouveau, que nous pouvons alors utiliser. Et on oublie souvent que cet avantage en termes d'échelle existe également dans l'autre sens : au lieu de nous concentrer uniquement sur un grand nombre de textes, nous pouvons aussi utiliser l'analyse computationnelle pour prêter attention à des éléments qui se situent bien en dessous du seuil que nous, lecteurs, sommes habitués à remarquer ou même à comprendre.

Ainsi, par exemple, je peux prendre plusieurs de mes romans préférés et vous raconter de quoi il s'agit. Je les ai peut-être lus trois, quatre, dix ou vingt fois. Mais si vous me demandez combien de fois le mot « le » apparaît dans le roman, je n'en aurai absolument aucune idée. Et vous me direz peut-être : quel intérêt ? En quoi cela peut-il être important ?

Il se trouve que, en fait, c'est important. L'analyse computationnelle est notamment utilisée dans ce que l'on appelle la stylométrie et l'attribution de paternité littéraire. Il s'avère que l'on peut déterminer qui a écrit un texte en fonction de la fréquence de mots syntaxiquement très courants tels que « le », « la », « un », « une », « je », « il », « elle ».

Ce sont des choses que nous ne sommes pas formés à reconnaître en tant que lecteurs, et l'analyse computationnelle nous permet donc de prêter attention à ces échelles vraiment minuscules, ce qui serait impossible autrement. Ainsi, entre le très grand et le très petit, cela nous ouvre un tout nouveau domaine d'investigation à travers lequel aborder l'étude du texte et, encore une fois, pour moi, l'étude de la littérature.

Je peux de ce fait comprendre un phénomène tel que le sublime : comprendre l'évolution d'un concept au cours de deux siècles d'une manière qui m'était inaccessible en utilisant simplement l'approche fondée sur la lecture critique. Cela ne veut pas dire pour autant que la lecture critique n'a plus d'importance. Elle est absolument essentielle.

C'est elle qui me permet d'élaborer les questions que je veux poser. C'est par elle que j'interprète les résultats obtenus par analyse computationnelle. Elle reste donc vraiment centrale. Mais ce que l'analyse computationnelle apporte, et je tiens à faire la distinction : ce n'est pas seulement de l'analyse computationnelle, c'est aussi de l'analyse quantitative... Il s'agit de transformer le texte en chiffres qui deviennent alors calculables et susceptibles d'élaborations à différentes échelles, macroscopiques et microscopiques. En faisant tout cela, je peux désormais poser des questions telles que

« qu'est-il advenu du sublime ? » d'une façon qui m'étais impossible avec les autres méthodes ou approches que nous, critiques littéraires, avions mobilisées jusqu'ici.

À bien des égards, cela réalise les rêves les plus fous du formalisme russe. Au début du XXe siècle, la critique littéraire rêvait que nous puissions soulever ce genre de questions grâce à une approche mathématique ou quantitative de la littérature. Les humanités numériques, et en particulier les études littéraires computationnelles, nous permettent enfin de les aborder de cette manière inédite.

***La Vie des idées : Pouvez-vous nous parler de vos recherches actuelles sur les concepts qui ont échoué ?***

**Mark Algee-Hewitt :** Je m'intéresse beaucoup en ce moment à l'étude de ce que j'appelle les concepts manqués. Je me suis donc penché sur la « révolution politique » et la manière dont ces deux termes ont évolué au cours du XVIIIe siècle pour finalement fusionner en un concept qui les unit en une seule chose. C'est ce que j'appellerais un concept réussi. On part de deux voies différentes d'investissement intellectuel, qui finissent par fusionner au fil du temps.

Ils s'approprient de plus en plus l'identité de l'autre jusqu'à ne plus former plus qu'une seule et même chose : un concept complexe et composite, comme décrit par John Locke. Et ce qui m'intéresse, c'est lorsque cela ne fonctionne pas, car je pense qu'en étudiant les concepts qui ont échoué, nous pouvons réellement comprendre pourquoi ceux qui ont réussi ont réussi. Qu'avaient-ils que les autres n'avaient pas ?

Ainsi, par exemple, « moral inadequacy » est l'un de mes exemples préférés de concept raté, parmi ceux que j'ai trouvés. Et encore une fois, cela n'est possible que parce que je peux mesurer la distance sémantique au fil du temps, en utilisant ces méthodes computationnelles à la surface de l'histoire intellectuelle. Je peux suivre l'évolution de ces deux mots qui, de notre point de vue, semblent sans rapport. Et je constate qu'en fait, vers le milieu du XVIIIe siècle, ils ont commencé à se rapprocher.

Ils ne faisaient plus qu'un : l'inadéquation morale. Pouvons-nous juger de l'éthique d'une action en nous fondant simplement, en quelque sorte, sur la quantité de bien qu'elle produit ? Pendant un temps, il a semblé que oui. Mais si l'on retrace l'évolution sémantique, vers 1750 cette théorie s'effondre à nouveau. Tout semblait converger, puis, tout à coup, les mots divergent et le concept s'effondre. Plusieurs explications différentes sont aujourd'hui avancées pour expliquer ce qui s'est passé.

L'impératif kantien, critique ou catégorique, entre en jeu et tout d'un coup nous pouvons comprendre la moralité d'une action non pas en recherchant le bien qu'elle fait, mais en posant – par exemple – une question ontologique telle que « Est-ce que cela est bien si cela s'applique à tous les cas ? ». Ainsi, ces changements philosophiques, sociaux et même technologiques qui se sont produits entre le milieu et la fin du XVIIIe siècle sont à l'origine d'une grande partie de ces évolutions conceptuelles, de sorte que l'analyse de celles qui échouent nous permet de mieux comprendre pourquoi elles échouent et, par conséquent, de mieux comprendre pourquoi d'autres réussissent.

***La vie des idées : Que nous apprend l'évolution de la catégorie du sublime sur les transformations du champ littéraire ?***

**Mark Algee-Hewitt :** Dans un sens très concret – et c'est la thèse que je défends dans ce projet –, le sublime nous donne au XIXe siècle le champ littéraire tel que nous le concevons aujourd'hui. C'est l'affirmation que je soumets à la discussion. Ce que je veux dire, c'est que le sublime devient une catégorie critique sur laquelle repose la différenciation entre les différents types d'écriture, quelque chose qui sépare du reste ce que nous pourrions considérer comme de la littérature : le type d'écriture culturellement concurrentielle que tout le monde veut nous faire considérer, que nous enseignons dans les écoles, que nous encourageons les gens à lire pour s'améliorer.

Et puis il y a le simple divertissement, les sottises qui sont publiées en quantités astronomiques, qui sont en quelque sorte destructrices sur le plan culturel, les choses auxquelles nous ne voulons pas que nos enfants soient exposés. Et bien sûr, au XIXe siècle, sans télévision, sans jeux vidéo, sans cinéma, cette différenciation s'est opérée dans le domaine de la littérature. Et le sublime m'a toujours intéressé parce que le sublime est cette figure ultime de l'irreprésentable.

Selon Kant, par exemple, le sublime est ce qui se produit lorsque l'imagination échoue, que l'entendement échoue et que la raison doit venir à la rescousse. En d'autres termes, c'est la façon dont nous décrivons quelque chose que nous ne pouvons pas décrire, lorsque nous pouvons décrire notre incapacité à le décrire et ainsi atteindre quelque chose d'essentiel à son sujet.

Au début du XVIIIe siècle, un changement technologique majeur s'est produit : ce que l'on appelle en littérature, en histoire et en études des médias, l'explosion de l'imprimerie.

Certes, l'imprimerie est alors une technologie qui existe depuis un certain temps déjà, depuis deux siècles environ, mais elle reste relativement rare. En effet, même si l'on imprime de plus en plus, c'est vraiment au XVIIIe siècle que l'on observe, si l'on représente graphiquement le nombre d'ouvrages publiés, une courbe en forme de crosse de hockey qui s'envole littéralement, de manière exponentielle. Auparavant, le mécanisme de distinction reposait sur cette différenciation technologique : l'impression, qui coûtait cher, parce que la reliure coûtait cher, parce que les livres étaient beaux et reliés en cuir, et que seuls les très riches pouvaient se les offrir. Les bibliothèques de prêt, par exemple, qui ont vu le jour un peu avant la période que j'étudie, ont beaucoup contribué à différencier les bonnes œuvres, parce qu'elles étaient imprimées, des mauvaises œuvres, parce qu'elles n'étaient pas imprimées, qu'elles circulaient sous forme de manuscrits, ou qu'elles étaient imprimées à moindre coût et diffusées sous forme de tracts et de journaux.

Mais à mesure que les textes imprimés deviennent de plus en plus accessibles à un nombre croissant de personnes, grâce à l'évolution de la technologie et à l'augmentation du volume imprimé, ça ne marche plus. Les rues sont inondées, comme le décrit Dryden, de romans bon marché, de littérature bon marché.

On ne comprend plus de la même manière le partage entre ce qu'il faut lire et ce qu'il ne faut pas lire. Et donc, une grande partie du XVIIIe siècle est en quelque sorte consacrée à la recherche de critères de différenciation, à comment distinguer le bon du mauvais. Et c'est ce qui finit par donner naissance à la discipline de la critique, la discipline des études littéraires.

C'est là son origine. C'est à cela qu'elle est destinée à l'origine. Et en recherchant le sublime, qui est le concept esthétique clé du XVIIIe siècle, elle prend son essor de la même manière que l'imprimerie. Les gens parlent beaucoup de la beauté au XVIIe siècle, et ils continuent à en parler beaucoup à la fin du XVIIIe siècle.

Mais le terme de sublime, rarement utilisé au XVIIe siècle, apparaît – au moins une fois - dans un tiers des livres publiés en Grande-Bretagne en 1798. Il est partout. Mon projet tente donc de comprendre d'où il vient. Pourquoi est-il devenu si populaire ? Et que devient-il ? Quel est son rôle, en tant que concept esthétique qui existe jusqu'au début du XIXe siècle, puis disparaît assez rapidement ? Je soutiens qu'il s'est dissous dans la critique, dans cette tentative de distinguer la bonne littérature et la mauvaise. C'est ce que faisait déjà la technologie du sublime. La bonne littérature représente alors cette sorte d'irreprésentable, en utilisant les techniques du sublime esthétique, en utilisant les techniques qui procurent au lecteur ce type de catharsis

aristotélicienne, qui lui enseignent l'irreprésentable et lui font ressentir cette incroyable passion ou transformation – la « transcendance » est en fait le mot qui est généralement utilisé pour la décrire. Alors que la mauvaise littérature peut vous procurer des sensations fortes de bas étage, mais elle n'est pas vraiment sublime.

Ainsi, en retraçant l'évolution de ce concept esthétique, comment il se développe, comment il change, comment il s'intègre dans différents types d'écrits, dans différents types de contextes, et en examinant ce qu'il conserve à mesure qu'il entre dans le XIXe siècle, à savoir cet attachement à la poétique, cet attachement à la représentabilité, mais aussi cet attachement à l'affect, à la transcendance... et puis, en examinant ce qu'il abandonne, c'est-à-dire une grande partie de ses connotations religieuses et rhétoriques, on peut vraiment voir comment cela devient le concept fondateur de la critique littéraire et, par extension, le concept fondateur de la littérature au sens moderne du terme.

***La Vie des idées : À propos maintenant d'une autre de vos recherches, comment les romans sur le changement climatique parviennent-ils à allier rigueur scientifique et fiction sans induire le lecteur en erreur ?***

**Mark Algee-Hewitt :** La clé, ici, réside dans la compréhension de ce qu'est la fiction, concept que la critique littéraire comprend d'une manière différente que la conscience commune, qui a tendance à utiliser le terme « fiction » de manière informelle pour parler de choses qui ne sont pas vraies. Ainsi, je m'intéresse beaucoup aux représentations fictives de faits réels, par opposition à la désinformation climatique, qui est omniprésente et que je qualifierais de représentation réelle de faits fictifs.

Et en fait, tous les romans ne correspondent pas à ce cas de figure de faits réels dans une représentation fictive. Par exemple, *État d'urgence* de Michael Crichton est l'un des rares exemples de roman niant le changement climatique. En d'autres termes, l'objectif de Crichton est d'essayer de nous convaincre que le changement climatique n'existe pas, ou du moins que le changement climatique anthropique n'existe pas.

Il mobilise donc toutes les ressources de la fiction au service de cette thèse. Néanmoins, la plupart des romans que nous examinons présentent effectivement le changement climatique comme anthropique et tentent d'enseigner aux lecteurs des faits réels à l'aide d'univers fictifs. Là encore, la clé réside dans l'univers fictif, car celui-

ci permet au narrateur du livre, et par extension à l'auteur du livre, d'intégrer ces faits dans le texte de manière à faciliter leur assimilation par les lecteurs.

Et peut-être pas seulement leur assimilation, mais aussi la confiance en ces faits, le fait de les intérioriser d'une manière qui ne mette pas nécessairement les gens sur la défensive, en particulier à propos d'un sujet aussi sensible que le changement climatique. Car beaucoup de gens ont déjà des opinions très tranchées et très fortes à ce sujet. Nous pouvons ainsi analyser comment différents livres utilisent ces faits et les intègrent différemment dans leurs récits, comment ils s'y prennent pour contourner avec succès les barrières mentales que les gens leur opposent en suivant leurs idées préconçues.

L'un des moyens d'y parvenir est de passer par la construction d'un univers. On le voit tout le temps dans les médias populaires. On a un jeune protagoniste naïf qui, à un moment donné, rencontre un mentor plus âgé et plus sage. Et ce mentor plus âgé et plus sage enseigne au jeune protagoniste, et – par extension – au lecteur, au spectateur, au public, toutes les règles du monde dans lequel ils vivent.

Par exemple, dans la série des *Star Wars*, dans le tout premier film, Luke Skywalker va voir Obi-Wan Kenobi. C'est un jeune fermier naïf. Obi-Wan Kenobi lui apprend ce qu'est la Force, ce qu'est l'Empire, ce qu'est la Rébellion, et par extension, nous, le public, apprenons tout cela en même temps que le protagoniste. C'est un mécanisme narratif très efficace.

Et on le retrouve dans beaucoup de ces romans sur le changement climatique, qui commencent avec un protagoniste naïf qui rencontre un mentor plus âgé qui lui apprend comment le monde est en train de changer, comment le changement climatique a affecté ceci et cela, et comment tous ces changements sont liés à une série d'activités humaines. C'est un exemple assez manifeste de la façon dont ce mécanisme fonctionne.

Mais en l'intégrant dans ces conventions narratives que nous reconnaissons, que nous, lecteurs, spectateurs et consommateurs de médias, savons saisir, interpréter et intérioriser... en l'intégrant dans cette technique, un roman peut nous enseigner ces faits sans nécessairement nous faire la morale et sans avoir à se doter, par exemple, d'une annexe contenant des faits simples et arides que nous serions censés lire afin de comprendre ce qui se passe dans le livre. Et c'est ce que j'entends par cette compréhension de ce qu'est la fiction ; car il ne s'agit pas simplement de créer quelque chose qui serait faux.



On peut considérer la fiction comme une sorte d'expérience de pensée. Elle permet aux auteurs – et bon nombre des livres que nous étudions sont des œuvres de fiction spéculative, de science-fiction, etc. – de créer ce terrain de jeu où ils peuvent développer et imaginer les conséquences des faits dont nous avons connaissance aujourd'hui. En fait, d'après le travail que nous avons accompli sur ces romans et les modélisations qui en découlent, je dirais que les plus réussis, ceux les plus à même d'enseigner aux lecteurs le changement climatique, ceux qui parviennent le mieux à transmettre ces faits, sont ceux qui ne situent pas leur univers dans le présent.

Ils n'essaient pas de nous expliquer ce qui se passe actuellement. Ils sont tournés vers l'avenir et inventent une sorte de monde contrefactuel, un monde fantastique, où ils peuvent figurer les conséquences inévitables, selon eux, de la situation climatique réelle dans laquelle nous nous trouvons aujourd'hui. Si l'on compare avec un ouvrage comme *État d'urgence* de Michael Crichton, qui se déroule largement dans le présent et s'attache à la récitation aride de faits : Crichton n'utilise pas ce type de techniques narratives ; à la place, il fait intervenir des personnages qui monologuent sur le fait que le changement climatique est une invention et que le terrorisme environnemental est en train de ruiner le monde. Cela n'est pas vraiment captivant, surtout en comparaison de quelque chose comme la trilogie de *La Terre fracturée* de N.K. Jamin, qui relève davantage de la science-fantastique que de la science-fiction. Celle-ci se déroule dans un monde complètement différent du nôtre, où les lois de la physique fonctionnent quelque peu différemment.

Mais ce qu'elle sait très bien faire, c'est prendre une grande partie de la situation climatologique dans laquelle nous nous trouvons actuellement – les faits sur le climat – et les transposer dans le monde qu'elle a créé : de telle sorte que nous pouvons en voir les conséquences, et que nous pouvons voir ce qui se passe. Mais comme c'est tellement décalé, comme cela ne se situe pas dans notre monde, nous ne réalisons pas immédiatement qu'il s'agit de notre situation.

Cela vient plus tard, et c'est une manière beaucoup plus efficace d'enseigner ces faits au public et aux lecteurs, par opposition au type de récitation que l'on trouve dans les romans les moins réussis.

***La Vie des idées : Les chercheurs en humanités numériques peuvent-ils aider les développeurs d'IA générative ?***

**Mark Algee-Hewitt :** La question de l'IA, ou des grands modèles linguistiques sur lesquels repose la technologie de l'IA de la génération actuelle, est en fait assez délicate dans le monde des humanités numériques. Il s'agit d'une technologie que nous utilisons sous une forme très rudimentaire depuis plusieurs années : les plongements lexicaux, les modèles linguistiques, la technologie sous-jacente aux moteurs d'IA générative actuels, sont exactement, par exemple, dans mon projet sur l'évolution des concepts, ce qui me permet de mesurer la distance sémantique en fonction de la similarité contextuelle.

Aujourd'hui, les modèles d'IA contemporains le font de manière beaucoup plus sophistiquée, mais le modèle sous-jacent reste relativement similaire. Et pourtant, ce qui s'est passé dans le monde de l'IA, en particulier depuis l'avènement de ChatGPT et la façon dont il a conquis le public il y a trois ou quatre ans, c'est qu'une grande partie de l'énergie a été consacrée au développement des modèles, en leur fournissant de plus en plus de données d'entraînement à utiliser, car cela s'est avéré être un excellent retour sur investissement. Plus on leur fournit de données, plus ces modèles deviennent sophistiqués. Mais le problème auquel nous sommes confrontés est que ces données ne sont pas toutes bonnes. Pendant un certain temps, plus il y en avait, mieux c'était. Mais il s'avère que cela entraîne avec le temps une forte diminution des rendements.

Et c'est là que se déroule actuellement une grande partie de la recherche de pointe dans le domaine des humanités numériques. Au lieu de simplement prendre l'un de ces modèles d'IA et de l'entraîner en y introduisant toutes sortes de données, pourquoi ne pas adopter une approche légèrement plus sophistiquée en matière de données d'entraînement ? Pourquoi ne pas lui fournir des données issues de domaines particuliers afin de créer un modèle spécialisé qui serait vraiment très performant sur une tâche ou une chose précise ?

Ainsi, dans mon domaine, par exemple, les modèles historiques suscitent beaucoup d'intérêt. L'un des défis liés à l'utilisation de modèles d'IA, comme ChatGPT, pour effectuer des recherches historiques, réside dans le fait qu'ils ne peuvent pas oublier. Vous pouvez interroger un modèle sur des événements qui se sont déroulés au XIXe siècle. Dans l'ensemble, ses réponses seront correctes. Mais si vous continuez à le questionner, par exemple sur la Lune et la façon dont elle était comprise au XIXe siècle, à un moment donné il glissera quelque chose qui provient des alunissages Apollo, car cela fait partie de ses données d'entraînement et il n'a pas de bon moyen

de faire la différence, de se dire : attendez, en 1899, les gens ne savaient pas à quoi ressemblait la Lune, car nous n'y étions pas encore allés.

Il y a donc un réel intérêt pour l'entraînement de modèles historiques, c'est-à-dire à entraîner les modèles à partir de données historiques afin de pouvoir simuler un lecteur ou un penseur de 1899. Mais le défi réside dans le fait que l'entraînement de ces modèles nécessite une telle quantité de données... alors que nous ne disposons pas d'un volume de textes historiques équivalent à celui des textes contemporains.

Ainsi, une grande partie de la recherche de pointe sur la question de savoir comment mener à bien ce travail de formation de l'IA consiste à déterminer si nous pouvons entraîner un modèle plus intelligemment à l'aide de textes choisis de manière beaucoup plus judicieuse et hautement sélective, tels que ceux dont nous disposons pour les modèles historiques. Ou bien devons-nous plutôt prendre un modèle contemporain et essayer de l'affiner à l'aide de données historiques ?

Cette question reste encore sans réponse, mais le fait que nous la posions, le fait que nous expérimentions cette alternative, est étroitement lié à la manière dont les chercheurs et les développeurs en IA collaborent avec les chercheurs en humanités numériques, qui sont ceux qui posent ce genre de questions et qui veulent utiliser l'IA générative comme outil de recherche, car elle représente une sorte de réservoir inexploité de lecteurs naïfs.

Une grande partie du travail opérationnel dans le domaine des humanités numériques consiste à annoter des textes. Je mène actuellement un projet visant à identifier les espaces domestiques dans les romans, et afin de voir si nous pouvons former un modèle capable de déterminer si une scène se déroule ou non dans un espace domestique. Pour ce faire, nous avons dû étiqueter des milliers et des milliers d'extraits comme étant domestiques ou non, puis les intégrer dans un modèle d'apprentissage automatique, qui s'est alors avéré très efficace pour faire la différence.

Cela a dû nous prendre près d'un an. Maintenant, imaginez si nous pouvions simplement fournir tous ces passages à ChatGPT, lui demander de les annoter, puis utiliser cela pour construire le modèle d'apprentissage automatique que nous voulions créer au départ. Ce serait formidable, cela révolutionnerait le type de travail que nous pouvons faire. Mais nous n'en sommes pas encore tout à fait là. Et l'écart entre ce dont il est capable aujourd'hui et ce dont il pourrait être capable dans deux, trois, cinq ou même dix ans : une grande partie de cette recherche-là, qui est en cours, ne consiste pas seulement à gaver le modèle avec de plus en plus de données, mais plutôt à

réfléchir de manière intelligente et critique à la façon dont nous entraînons les modèles, à la façon dont nous leur apprenons à mieux lire et, par extension, à mieux écrire.

Tout cela découle en grande partie du type de travail effectué dans mon domaine. De nombreuses collaborations ont vu le jour à travers le pays et dans le monde entier, entre des chercheurs en humanités numériques et des développeurs d'IA, dans l'espoir de pouvoir utiliser à la fois les techniques que nous avons affinées et perfectionnées grâce à des études computationnelles de la littérature, de l'histoire, de la géographie, etc., et des textes en général, et ces nouveaux modèles d'IA que nous voulons rendre un peu plus intelligents à mesure qu'ils évoluent vers davantage d'agentivité et une utilisation par la recherche.

Traduit de l'anglais par A. Suhamy

Publié dans [lavedesidees.fr](http://lavedesidees.fr), le 3 octobre 2025.