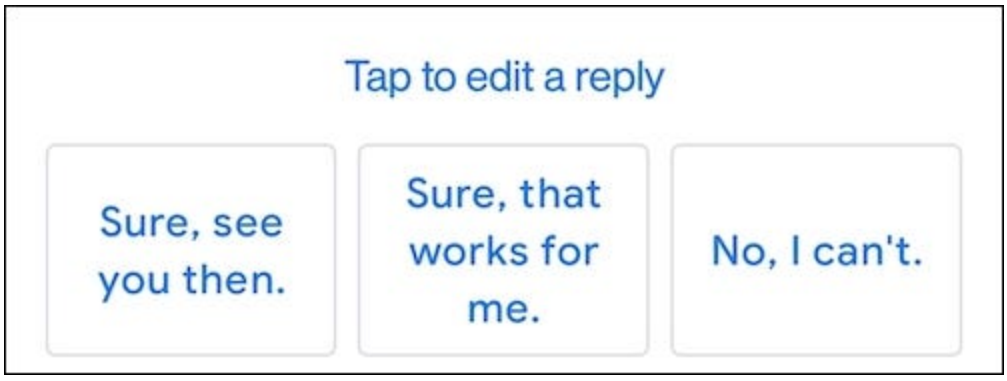
Les grands modèles de langage: moyens ou objets?

Pierre Depaz (NYU Berlin) - 26.11.2024

Je m'appelle Pierre Depaz, je suis chargé de cours à NYU Berlin. Après un cursus en sciences politiques et en design de jeux vidéos, j'ai fini une thèse sur le rôle de l'esthétique du code source, et donc plus généralement sur les questions d'épistémologie dans les textes de programme informatique.

L'idée de cette contribution est de proposer différentes perspectives sur l'usage d'un nouvelle technologie (le LLM) au sein d'un contexte des humanités numériques. Si les nouvelles technologies tendent parfois par leur immédiateté à éclipser des réflexions sur le plus long termes, il faut, je crois, rester néanmoins attentif à toutes les implications d'une nouvelle manière de faire pratique qui vient s'articuler à une manière de faire théorique. Ces possibilité d'usage sont encore largement ouvertes et indéterminées, et donc pluridirectionnelles.

Plutôt que de présenter des conclusions, je voudrais plutôt suggérer des pistes de réflexions, qui sont informées par une approche multidisciplinaire, entre sciences sociales, design et études littéraires.



Pour moi, un des premiers examples de textes générés par une IA, c'est la fonction Smart Reply de Google, publiée en 2016. Il y a deux manières dont on peut envisager Smart Reply.

D'une part, le système permet de faciliter la communication et la coordination entre individus, de supporter une certaine forme de productivité en allant droit au but, ou encore de maintenir une illusion de politesse en répondant facilement "Merci", "Vous aussi", "Bien reçu" (em somme, des petits lubrifiants sociaux).

D'autre part, on peut s'intéresser à la nature de cette nouvelle forme d'écrit, à ce que ce genre de texte signifie, à la manière dont il a été créé. Est-ce que c'est ce qu'on pourrait entendre comme agent

conversationnel, une machine qui facilite la conversation? Quels sont les étapes de l'auctorialité de ce texte? Comment les ingénieurs, designers, utilisateurs et utilisatrices qui sont impliquées dans cette chaîne vont-ils décider de ce qui va apparaître? Combien de fois ces textes ont-ils été écrits, lus, ré-écrits, relus? Quel sont les champs lexicaux inclus, ou exclus? Smart Reply, c'est la conflation d'un texte et d'un bouton: d'une interface, d'une action et du langage.

Depuis 2022, il y a une tendance à exceptionnaliser les grands modèles de langage, comme le montre la synechdoche qu'est ChatGPT. ChatGPT n'est pas un grand modèle de langage, c'est une interface commercialisée médiatisant l'accès à un grand modèle de langage. De fait, les grands modèles de langages semblent toujours médiatisés, mais de manière plus ou moins spectaculaire.

À ce sujet, il y a une histoire intéressante sur le genre de messages que peut suggérer Smart Reply. Apparemment, les premières version de la Smart Reply a du être modifiée puisque les suggestions incluaient de manière récurrentes la phrase "I love you". Trois petits mots qui insistaient à faire surface dans nos conversations d'emails, mais qui, pour des raisons de gestion de produit, ont été éffacés. Pourquoi est-ce qu'ils y sont apparus? De quels genre d'échanges est-ce qu'ils témoignent?

*Cette technologie numérique peut donc être comprise comme un **moyen** (faciliter une production de texte) ou comme un **objet** (analyser une production de texte). Et je pense que cette dichotomie est fondamentale aux humanités humériques.*

> *The digital humanities try to take account of the plasticity of digital forms and the way in which they point toward a new way of working with representation and mediation, what might be called **the digital 'folding' of reality**, whereby one is able to approach culture in a radically new way.* [1]

Cette dualité est assez bien capturée par cette analyse de David Berry.

Il considère les humanités numériques comme une attention particulière à la manière dont les formes digitales vont affecter nos procédés de représentation et de médiation, qu'il va alors qualifieer de 'plissure' de la réalité.

Cette 'plissure' de la réalité, je la vois de deux manières: est-ce que nous la provoquons (par exemple en détectant la translittération du nom Oum Kalthoum en une expression régulière), où est-ce que nous l'observons (par exemple en réfléchissant à aux possibilités et implication de la conception du texte comme un assemblage de motifs formels exclusivement syntactiques).

Les deux dernières innovations technologiques qui vont contribuer à ces plissures de la réalité sont donc les architectures Transformer et les architectures Stable Diffusion, mais je n'addresserai que la première aujourd'hui

Quelles places peuvent prendre les grands modèles de langage dans les sciences humaines?

Quelles sont les implications de les envisager comme un **moyen** de la recherche? Comme un **objet** de la recherche?

C'est bien cette bidirectionnalité qui m'intéresse dans les humanités numériques, parce que, pour moi, l'égalité de ces deux approches est importante: en plus de ce qu'un grand modèle de langage peut apporter à la recherche en sciences humaines en termes pratiques, je suis aussi curieux de ce que les sciences humaines peuvent apporter pour mieux saisir ces nouveaux moyens de création.

1. Le LLM comme moyen de recherche
2. Le LLM comme objet de recherche
3. Pistes d'épistémologie critique

Je vais donc commencer par développer ces deux grandes approches de la génération de texte par ordinateur, comme des grandes dynamiques des rapports entre sciences humaines et technologies numériques, en distinguant entre le moyen, et l'objet. Chacune des approches sera accompagnée par quelques études de cas d'usages de LLMs afin de typologiser différentes approches, avant de conclure sur une critique de l'épistémologie de l'IA.

Le LLM comme moyen

Toute technologie est une extension de l'humain [2] , et est un *pharmakôn* [3] .

Il semble que la technique comme moyen c'est la façon la plus "naturelle", ou du moins spontanée d'y penser, principalement parce que l'objet technique a pour composante fondamentale la fonction. La technique fait des choses, elle s'applique, elle est transitive, pour ainsi dire. Elle nous permet de mieux voir, de mieux agir, de mieux penser. Il y a donc deux formes de techniques: les technologies physiques et les technologies cognitives. La technologie physique, c'est la voiture; la technologie cognitive, c'est le livre.

À la base de la philosophie de la technologie se trouve cette observation, qui devient postulat, que toute technologie est une manière dont l'humain étend ses capacités, qu'elles soient physiques ou mentales. McLuhan va part de ce principe en considérant cet aspect mental, cognitif, en considérant toute média comme technologie, qui nous interface avec le monde.

Quoi qu'il en soit, une technologie a toujours des bons et des mauvais côtés, ce qui est représenté à travers le concept de pharmakon, que Bernard Stiegler emprunte à Platon. Face à une présentation du progrès technologique qui tend à l'unilatéralité (quelque chose est soit toujours bien, soit toujours mauvais), je trouve que c'est un concept utile pour ramener un peu de nuance.

Toute technologie computationnelle est une *mnémotechnique*, et nos interactions avec elles résultent en des *assemblages cognitifs*. [4] [5]

L'évolution des technologies cognitives consiste en l'évolution des manières de lire et d'écrire.

Pour ce qui est de l'écriture, il s'agit des alphabets, chiffres, notation algébraiques et codifications. Cela implique toujours un acte de traduction, afin de faciliter l'acte de remémoration.

Pour ce qui est de la lecture, on a inventé le rouleau, puis le folio, les marges, les notes de bas de page, etc. L'ordinateur en général fait partie de ces techniques, puisque c'est toujours une machine qui lit et écrit, avec des niveaux de plus en plus avancés d'automatisation.

Donc, plutôt que d'agir physiquement sur notre environnement matériel, il s'agit d'agir cognitivement sur notre environnement mental. Et cela va également avoir des effets plus large qu'une simple facilité à se remémorer un numéro de téléphone, ce qu'a notamment montré Jack Goody avec ses travaux sur la liste, et Bruno Latour avec ses travaux sur les vues de l'esprit, ce que Katherine Hayles appelle la technogenèse, la co-construction de l'humain et du medium.

Donc si le dessin en perspective, l'écriture en liste, et la lecture réticulaire vont avoir des affordances et des conséquences particulières sur la manière dont on réfléchit, alors quelle va être cette co-construction et ces conséquences pour un medium comme les LLMs?

Le grand modèle de langage comme traversée d'espaces syntactiques.

Les grands modèles de langage sont donc les dernières nouvelles manières de lire et d'écrire: à travers des vecteurs de mots. Il s'agit de systèmes qui peuvent, à partir d'une masse de données textuelles (une sorte de mémoire collective) d'abord générer des probabilités de suites de mots (des espaces de possibilité), puis utilser ces probabilités pour synthétiser d'autres suites de mots.

L'échelle à laquelle cela se passe suggère que le saut quantitatif résulte en un saut qualitatif. Le modèle mental que j'ai pour ce procédé d'acquisition et de génération de langage, c'est la navigation d'espaces syntactiques. Ce qui est fascinant, c'est que, de ces espaces syntactiques, on peut dériver des notions sémantiques.

Et pourtant, comme toute technique d'informatique, il s'agit d'un procédé purement formel. La sémantique semblerait, a priori, juste une coïncidence?

Cette formalité va mener à notre première étude de cas.

Le LLM comme mise en forme

Étude de cas: le secrétariat de rédaction et la pose de style.

Une des utilisations, c'est donc la question de la mise en forme du texte. Ça n'a qu'un rapport indirect aux études de textes en soi, mais a un rapport très direct à la diffusion des études de textes.

En gros, on donne une entrée, et on demande de le réaligner sur des pondérations plus élevées. Les modèles de langage ont été entraînés (à différents niveaux) pour générer de la prose de qualité. Je reviendrais plus tard sur ce que ça peut être cette "prose de qualité".

Et je fais l'hypothèse que c'est cette prose de qualité, comme interface d'accès, qui a été cruciale dans l'établissement de la dynamique d'usage des grands modèles de langage. Autrement dit, c'est en partie parce que ChatGPT s'exprime bien qu'il a suscité cet engouement.

De manière triviale, donc, les grands modèles de langage sont utilisés comme des secrétaires de rédactions. Ils mettent en forme, et surtout ils mettent en forme de manière _standardisée-.

Il y a un contexte où la standardisation est bonne: les emails, et les demandes de financement, les publications dans des "moulins à article", des journaux qui n'existent que pour participer à l'industrialisation néo-libérale de la production scientifique. Lorsqu'il s'agit d'interagir avec des machines (au sens de Kafka), autant utiliser le langage de ces machines.

Plus généralement, le style permet aussi de s'intégrer dans des communautés de pensée. Les anthropologues n'écrivent pas comme les chimistes, qui n'écrivent pas comme les historiens, etc. Donc

pour des jeunes chercheur.es qui sont en quête de légitimité et dont l'écriture n'est pas forcément le point fort, ça peut être extrêmement utile.

Dans mon cas particulier, j'ai fait partie des communautés de pensée des Instituts d'Études Politiques, et dont la communauté stylistique consistait en….

Deux parties, deux sous-parties comme **horizon d'attente de lecture**.

par avoir le concours. Jamais je n'ai compris pourquoi il fallait faire deux parties deux sous-parties, plutôt que trois parties trois sous-parties, comme on faisait au lycée.

Donc le moment où on va utiliser un LLM afin de formatter le style d'une prose peut avoir un effet bénéfique afin d'harmoniser son style personnel afin de correspondre à l'horizon d'attente de nos lecteurs.

En particulier quand on travaille dans une langue qui n'est pas sa langue maternelle: un grand modèle de langage peut aider drastiquement une étrangère à outrepasser la barrière de la langue.

Cependant, le style comme accord collectif n'est pas le seul moyen d'envisager le style académique, ni de faciliter la diffusion de la recherche.

Les très bons travaux académiques sont aussi ceux qui sont très bien écrits. Par exemple, je me souviens très bien de la première fois que j'ai lu l'introduction de Surveiller et Punir.

> *Voilà donc un supplice, et un emploi du temps. Ils ne sanctionnent pas les mêmes crimes, ils ne punissent pas le même genre de délinquants. Mais ils définissent bien, chacun, un certain style pénal. Moins d'un siècle les séparent.* [6]

Example du style de Foucault. On m'a toujours appris qu'il fallait commencer une dissertation par un exemple, afin d'en découler une question. La première fois que j'ai lu Surveiller et Punir, qui compare donc deux jugements, j'ai compris pourquoi cette formule marchait si bien.

Je suis sûr que chacune et chacun, dans le monde académique, a déjà été marqué.e par des travaux de recherche tant par sa forme, que par son contenu. J'ai donné Foucault en example, mais je pourrais aussi citer Lorraine Daston, Friedrich Kittler, Wendy Chun ou encore Sherry Turkle. Lorraine Daston, par exemple, a écrit un papier sur l'épistémologie du "coup d'oeil", de la manière qu'on a de comprendre quand on survole, et tout au long de ce papier, elle file une métaphore sur la vision des anges, et c'est magnifique.

La forme et le contenu ne sont pas disjoints.

C'est à dire qu'une certaine forme (standardisée), va possiblement insinuer que les idées sont également standardisées. À l'inverse, une forme particulière peut aussi suggérer des perspectives particulières sur certaines idées (dans un mouvement similaire aux configurations syntactiques de LLM qui ont des conséquences sémantiques).

Les LLMs, employés comme moyen de mise en forme (de conformisme, pourrait-on dire), sont donc une arme à double-tranchant. On retrouve ici le concept du pharmakon de Stiegler: quand on met en forme, on gagne en conformité, et on perd en identité.

Le LLM comme facilité d'accès

Etude de cas: La recherche en ondes gravitationnelles [7]

La baisse des coûts d'opportunité.

Passons maintenant à des applications plus directes de LLMs à des activités de recherche, et particulièrement dans les cas des sciences dures.

André Curtis Trudel présentait à un séminaire un travail en cours qui se basait sur ses observations dans un laboratoire d'astrophysique, qui se préoccupait notamment des méthodologies de recherche en vagues gravitationnelles. La question de "qu'est-ce que le LLM fait à la recherche", il l'aborde par le prisme du coût d'opportunité.

Il part du principe, posé par C.S. Peirce, que la pursuitworthiness d'une direction de recherche est une fonction des ressources engagées dans cette poursuite. Si cela demande beaucoup de ressources (du temps, de l'argent, de l'effort), cela ne vaut pas forcément le coup de dépenser ces ressources pour explorer cette hypothèse. Cela est particulièrement vrai dans des grands corpora, et dans des corpora qui coûtent cher à explorer.

Ce qu'il observe, c'est que les grands modèles de language (et les modèles d'apprentissage machine en général) vont permettre aux chercheuses d'explorer d'avantage de pistes de recherche, des choses qui auraient été hors de portée auparavant.

Et cela peut aussi s'appliquer aux SHS.

Étude de cas:: Détection des motifs de langages dans des grands corpus (détection et contextualisation)[^do-ollion].

- large corpus bien défini et digitalisé
- requête précise (instances de sources anonymes)
- facilité d'entraînement et de vérification [8]
- utilisé pour la réponse à une problématique

[^do-ollion] Do, Sheng, Ollion, Étienne, & Shen, Rui. 2022. *The Augmented Social Scientist: Using Sequential Transfer Learning to Annotate Millions of Texts with Human-Level Accuracy*. Sociological Methods & Research.

Sheng Do et Étienne Ollion se posaient la question de la citation anonyme dans le journalisme politique: quand est-ce que c'est arrivé, dans quel contexte, sous quelles modalités?

C'est particulièrement difficile de répondre à ces questions de manière purement humaine (sans rentrer dans les détails de ce que voudrait dire "purement humaine"), mais on voit bien que les coûts d'entrée sont élevés. Donc il y a là un critère: c'est très facile pour les humains, mais très compliqué pour des ordinateurs traditionnels.

Ce genre d'exemple est, à mon sens, un bon exemple de la manière dont des grands modèles de langage peuvent être utilisés de manière pertinente:

- énorme corpus bien défini et digitalisé
- utilisations précise (instances de sources anonymes)
- facilité d'entraînement ("I know when I see it") mais aussi possibilité de vérification des performances. C'est ce que fait le papier de Stefano De Paoli, qui se focalise sur la méthodologie, et confirmant qu'un LLM peut performer aussi bien qu'une annotation humaine.

Donc ici, on peut voir une utilisation possiblement assez fructueuse et pertinente d'un LLM pour des études de textes, et je pense que les questions méthodologiques que se posent les sciences sociales peuvent aussi informer les sciences humaines lorsqu'il s'agit de vérifier des résultats de techniques computationnelles. Ce qui m'intéresse là, c'est que cette étude considère complètement que le LLM est une machine computationnelle, en aucun cas un agent conversationnel. C'est une version plus subtile, plus flexible, d'une requête SQL

Étude de cas: Interagir avec des stéréotypes dans une étude de psychologie [9]

- personification d'un LLM
- problème d'interprétation des résultats

L'approche inverse, ce serait de traiter les sorties d'un LLM de manière plus... qualitative? C'est à utiliser la malléabilité de ces modèles pour façonner des entités anthropomorphiques (des personnalités) et les interroger.

Un autre example, ce serait l'utilisation d'un LLM comme agent de conversation pour des patients of psychologie. Il s'agissait par exemple de créer des personnalités et les faire interagir avec des patients pour étudier leurs réactions. Les psychologues disent que c'est dur d'évaluer un modèle car on ne sait pas ce qu'il y a dedans. Ils disent surtout que le problème, c'est qu'il est très difficile d'interpréter les résultats de l'étude: est-ce que la réaction des patients est symptôme de la pathologie, où est-ce que c'est un symptôme de l'inefficacité du LLM.

Ce qu'ils insinuent ici, c'est que pour une science dite "dure", on ne veut pas avoir à interpréter la mesure du phénomène, on veut avoir à interpréter le phénomène en lui-même. Si on a un outil, il faut qu'il soit ***bien calibré****.*

De la baisse des coûts d'opportunité:

- accès à un SQL extrêmement performant [10]
- glissmement vers le sentier de moindre résistance [11] .

Donc d'un côté, si on considère le problème comme étant une limitation des langages SQL actuels, ou des limitations de temps qui seraient impossible à surmonter sans une équipe de plusieurs personnes à plein-temps, le LLM semble être un très bon outil.

De l'autre côté, il y a aussi une tendance, ou du moins cette expérimentation à vouloir appliquer une nouvelle technologie, je trouve que c'est un glissement qui peut être problématique, et qui résonne avec l'adage "si vous avez un marteau, tout ressemble à un clou".

Et je pense que c'est une composante qu'il ne faut absolument pas sous-estimer quand on parle de nouvelles technologies: la propension que les humains ont a prioriser le comfort et la facilité face à la lenteur et la difficulté, qu'on peut nommer le sentier de moindre résistance. Cette facilité, on peut la voir dans la description des jeunes hackers du MIT par Joseph Weizenbaum (ou dans l'example de la modélisation hydrologique, ou des mathématiques en primaire).

Cette tendance peut ensuite se traduire par une dépendance au sentier: si on investit dans la formation et l'utilisation des LLMs, on se sent obligé de les utiliser (Jacques Ellul).

La question qu'il faudrait se poser c'est: "est-ce qu'il y aurait un moyen de faire ça sans LLMs? comment varierait la qualité de ces résultats? quelles sont les conditions qui font que je vais utiliser ces LLMs, et est-ce que ces conditions sont valides épistémologiquement?" ou, en d'autres termes, "est-ce que je le fais parce que je suis sous des pressions externes de production, ou d'adhésion à un zeitgeist technologique, ou est-ce que je suis face à des contraintes plus ou moins valides d'accès à, et d'études de, mon matériau source?"

Encore une fois, qu'est-ce que j'y gagne, et qu'est que j'y perds?

Le LLM comme partenaire de conversation

Discuter et raisonner avec un LLM. [12]

Le dernier cas d'utilisation des LLMs dans la recherche, c'est la phase exploratoire de la recherche, que ce soit l'idéation ou la revue de l'état de l'art.

A priori, ce n'est pas foncièrement une mauvaise idée: je souscris à la théorie de David Chalmers et Andy Clark sur l'extension mentale, qui dit en deux mots que notre réflexion n'est pas uniquement introsomatique, mais aussi extrasomatique. La technologie nous aide toujours à penser: sous quelles conditions est-ce qu'un LLM nous aide à penser?

Le problème de l'interface du dialogue, c'est qu'il va y avoir un procédé d'antropomorphisation, et ce procédé est extrêmement puissant, puisqu'il a contribué très fortement à la popularisation de ChatGPT. Donc notre pensée se trouve aussi guidée par des normativités subtiles, et donc reformulée, sans qu'on sache très bien pourquoi, ou sans forcément qu'on ne le remarque.

Je pense que la distance que l'on prend par rapport à un texte quand on l'écrit sur une feuille de papier ou un logiciel classique n'est pas la même que l'on prend lorsqu'on entre dans un échange discursif avec un LLM. Et on n'est toujours pas sûr de la nature du sens des phrases de ces modèles, et donc de la qualité de la réflexion qu'on peut mener avec ces modèles.

Étude de cas: Programmer avec un LLM.

Donc bon, on utilise quand même des outils en permanence pour nous aider à réfléchir, et mon outil préféré, chez les programmeurs, c'est le canard en plastique.

De manière très crue, les programmeurs utilisent un canard en plastique pour résoudre leurs problèmes: on explique son problème au canard, et en formulant son problème à voix haute, cela nous aide à y réfléchir. Dans le cas des canards, cela nous permet de projeter, de reformuler notre pensée.

Implications de la programmation assistée par LLM [13] :

- le court-terme prévaut sur le long-terme.
- l'individualisé prévaut sur le commun.
- la disponibilité et la patience prévalent sur les accrocs sociaux.

Résumer l'article sur le passage de Stack Overflow aux LLMs.

Et puis il y a un dernier aspect de l'utilisation d'un LLM pour la recherche, c'est celui du ***résumé d'article lors de la revue de l'état de l'art****.*

> *"Les grands modèles de langage sont la preuve que lire rend intelligent, et nous les utilisons pour arrêter de lire."* [14]

Et puis, un des grands aspects des sciences humaines, c'est que la lecture est essentielle à la formation des idées. On lit, on annote, on croise des références, on gribouille dans des marges de manière assez exploratoire. Je ne suis pas certain que ce genre d'exploration soit possible avec un LLM (notamment à cause du cadre d'action que constitue l'interface).

Le cerveau est un muscle, et quand on l'utilise moins, il se relâche.

Donc, pour conclure cette partie sur le LLM en tant que moyen, je pense qu'il est utile de garder en tête qu'il y a toujours des bons et des mauvais côtés à l'utilisation de toute technologie, et qu'il n'est jamais donné que les bons côtés soient meilleurs que les mauvais côtés. Lorsque le résultat du LLM, utilisé en tant que moyen, demande un effort d'interprétation conséquent, ce n'est donc peut-être pas le meilleur des outils.

Un grand modèle de langage est-il une bibliothèque, ou une bibliothécaire? [15]

Jusqu'ici, considérer le LLM en tant que moyen, c'est le considérer uniquement en tant que bibliothèque: sa signification est dérivée de son jeu de données, et c'est une manière d'accès neutre au jeu de données.

En revanche, considérer le LLM comme un objet, c'est aussi le considérer comme un ou une bibliothécaire, qui serait professionnelle mais aurait néanmoins une opnion et une personnalité particulière. Dans ce cas, la signification des résultats du LLM est nouvelle, et a donc une qualité particulière à laquelle il faudrait faire bien attention.

Le LLM comme objet

Analyser du langage naturel par des méthodes de sciences informatiques,

ou analyser du langage machine par des méthodes de sciences humaines.

Si on considère qu'un LLM c'est une bibliothécaire plutot qu'une bibliothèque, alors on peut se poser la question de la sorte de bibliothécaire que c'est.

Un LLM, c'est de la génération de texte, et on a vu que le problème qu'ils avaient en sciences sociales et en psychologie, c'est que c'était compliqué d'interpréter les textes. En revanche, la littérature est bien placée pour s'enquérir d'un texte (critiquement, spéculativement, interprétativement, comparativement).

L'interprétation des textes numériques est au coeur des *software studies*, l'étude du logiciel comme artefact culturel [16] [17] .

Les software studies, c'est le champ d'études du logiciel comme objet culturel, qui va l'analyser:

- d'un point de vue socio-économique
- d'un point de vue technique
- d'un point de vue critique
- d'un point de vue herméneutique

Ce n'est donc pas le logiciel, l'algorithme, le programme, comme moyen de la connaissance, mais plutôt le programme comme objet de la connaissance. Qu'est-ce que le logiciel représente, qu'est-ce qu'il traduit, et comment, dans quel contexte est-il créé, utilisé?

C'est dans ce champ que je me suis inscrit en écrivant une thèse sur l'esthétique des codes sources. Plutôt que les considérer comme strict artefact techniques, c'était de considérer le code source comme une production culturelle, comme une sorte de patrimoine qui sort du fonctionnel pur.

> *One thing that foreigners, computers and poets have in common is that they make unexpected linguistic associations* [18] .

Le langage numérique est une façon particulière d'exprimer le monde

D'un point de vue l'étude de texte, on peut donc considérer les textes de LLMs comme méritant d'être étudiés. Donc je vais vous proposer de considérer ce que ça veut dire d'étudier un LLM, d'un point de vue

formaliste, autorial et critique.

Une approche formaliste

L'étude de textes des LLMs peut impliquer des théories de l'épistolaire, des analyses du registre formel, ou encore des approches comparatives, dans une optique d'extension du domaine de la littérature [19] .

Qu'est-ce qu'une analyse de texte peut nous apprendre des LLMs? Est-ce que cela peut mettre en exergue les formes de littératures conversationnelles? Par exemple, est-ce qu'il y a des théories venant de l'analyse de genre épistolaires qui pourrait d'avantage qualifier ces textes?

Quel est le registre de langue utilisé par GPT? Quel est le registre par défaut, et quels sont les registres qui sont plus facilement copiables que d'autres? Qu'est-ce que ça peut suggérer de notre rapport à la surface et à la profondeur?

Il semble que le "bon style" dont je parlais précédemment est le style "efficace", qui va droit au but, et qui minimise le nombre de propositions et de clauses par phrase, et qui va donc perpétuer une évolution du langage par sa réification technique.

The Breville Smart Oven Air Fryer is a countertop oven that can air fry, toast, bake, and more, while the Ottoman Empire was a vast territory that controlled Constantinople and was a center of culture, art, and science:

	Breville Smart Oven Air Fryer	Ottoman Empire
Features	11 functions, including air fry, toast, bake, broil, roast, warm, pizza, reheat, cookies, and slow cook	Controlled a vast expanse of territory, conquered Constantinople, and was a center of culture, art, and science

Par exemple, la version beta de Gemini intégrée au moteur de recherche de Google a une fâcheuse tendance à vouloir tout comparer!

Une autre piste d'approche, ce pourrait être de développer autour de la notion de discours policé, de discours formatté, comme on le mentionnait plus tôt. Pourquoi est-ce que tout doit exister sous forme de

tableau? Ou, plus généralement, sous forme de listes? Est-ce que cela participe aussi de cette productivité du discours et de la compréhension?

(exemple workshop NYU) Ici, il serait intéressant de donner des entrées variées au LLM afin de mettre en exergue l'acte de mise en forme, indépendamment des contenus.

On pourrait aussi considérer, une autre approche comparative entre grands modèles commerciaux pourrait faire apparaître des décalages, des différences, ou des cohérences. Est-ce qu'on peut dire des choses avec un modèle du CNRS qu'on ne peut pas dire avec un modèle d'Anthropic ou de Meta? Quelles sont les motifs de textes qui reviennent en permanence, et pourquoi? Y a-t-il des motifs que l'on observe jamais? Qu'est-ce qui est dicible et qu'est-ce qui est indicible avec des LLMs?

> *So, it seems that at minimum, **ChatGPT is a soft bullshitter**: if we take it not to have intentions, there isn't any attempt to mislead about the attitude towards truth, but it is nonetheless engaged in the business of outputting utterances that look as if they're truth-apt* [20] *.*

Par exemple, Hicks et ses collègues considèrent les textes de LLM à travers la perspective des bullshit studies, l'étude du baratin. Ce prisme-là permet de qualifier le texte, le rapport aux faits et à la vérité, mais aussi de mieux saisir l'effet de persuasion qui peut prendre place au sein du lectorat des grands modèles de langage.

Et à partir de la caractérisation de la nature des textes de LLMs, on pourrait alors développer sur les aspects uniques de productions textuelles humaines, et donc mieux déterminer ce qui fait une bonne oeuvre textuelle: qu'est-ce qui fait un bon roman, un bon poème, un bon essai? Quelle est la place de la standardisation et de l'improvisation? Quelle est la place de quel genre de dialogue dans l'expression littéraire?

Une approche auctoriale

La figure de l'auteur peut-être reconsidérée comme une chaîne d'écritures [21] , dont il faut aussi déterminer le but.

Une autre approche serait l'approche auctoriale. Est-ce qu'il y a un auteur? Qui en est l'auteur? La mort de l'auteur se traduit-elle par une focalisation exclusive sur la réception par le lectorat, ou bien est-ce qu'il s'agit plutôt de repenser ce qu'est un auteur?

Plutôt qu'un seul acte d'écriture, avec une intention claire et transmise jusqu'à la publication, est-ce qu'il y en a plusieurs? Ici encore, je trouve qu'il est utile de référer à des travaux en sciences sociales, notamment les chaînes d'écriture, de Béatrice Fraenkel. Dans énormément de matière textuelle qui nous

entoure, il y a très peu de clarté sur l'auteur initial (et encore moins sur l'auteur unique) d'un texte. Un document est le résultats d'une suite d'écritures, ancrées dans des contextes socio-économiques particuliers.

*Si on pousse la question de la chaîne d'écriture, il faut aussi s'intéresser au **but** de cette chaîne! Qu'est-ce que ce texte est sensé accomplir? C'est une question qu'on avait abordé précédemment à travers la forme, et qu'on peut aborder maintenant à travers l'intention.*

Quelles sont les *chaînes d'écriture* des LLMs [22] ?

- les auteurs des articles scientifiques?
- les personnes qui implémentent les algorithmes?
- les personnes qui ont écrit les données d'entraînement?
- les personnes qui ont récolté les données d'entraînement?
- les personnes qui raffinent le modèle? [23]
- les personnes qui *promptent* le modèle?

Qui est l'auteur? Les auteurs du papier sur l'attention et leur formules mathématiques pour déterminer le sens des mots? Quelle serait leur part de contribution?

Une fois que cette formule est publiée, il faut qu'elle soit implémentée. Ce qui se fait généralement dans des entreprises privées de l'ouest nord-américain. Alors il faut faire attention aux gestionnaires de projet de Microsoft, Google et Meta qui vont déterminer ce qui peut être dit, peut-être en se penchant sur des paratextes qui vont dicter ce qui peut être dit, et ce qui ne peut pas être dit.

Enfin, peut-on considérer que les données d'entraînement sont elles-même une forme d'auctorialité? Est-ce que, quand on lit les résultats d'un LLM, on lit dans la voix de Reddit, ou de Wikipedia? Peut-on attribuer une voix au processus de collecte des ces textes, de la même manière qu'un commissariat d'exposition peut s'insérer dans le discours d'une exposition? On pourrait aussi s'intéresser aux pondérations relatives des modèles de fondations, et des modèles raffinés (fine-tuned)

Ce qui vient alors reposer la question de l'intention: l'intention est-elle toujours individuelle? Peut-elle résulter d'un concert de voix dont la totalité vaudrait plus que la somme des parties?

Enfin, quel est le rôle relatif du prompt dans cette chaîne? Quelle est la latitude du choix de l'expression de l'usager, particulièrement dans un acte d'écriture qui va être médiatisée à plusieurs niveaux (interface graphique, interface programmatique)? Est-ce un acte d'écriture, ou un acte d'actualisation d'une écriture latente?

Étude de cas: Les examens scolaires comme preuve de l'intelligence. [24]

Comment est-ce qu'on fait sens de ce que nous racontent les LLMs? Comment est-ce qu'on établit ce qui est "correct" ou non? On fait comme avec les humains, on leur fait passer des examens, et l'utilisations de ces paratextes peut nous permettre d'analyser la notion d'intelligence telle qu'elle est utilisée dans le contexte sur terme "intelligence artificielle".

Le focus était plutôt sur Foucault que sur Deleuze...

Une approche critique

Mettre à jour ce qui est dicible par les LLMs, et ce qui ne l'est pas.

La question de l'autorialité va donc mettre à un jour tout un système de création qui va permettre au texte généré par la machine d'exister. Il devient alors possible d'examiner de plus près et ce texte, et cette chaîne d'écriture.

Dans ce cas, le texte des LLMs devient alors un résultat, un exemplaire, de l'environnement techno-économique qui lui ont permis de voir le jour, on pourrait mobiliser ici les travaux de Michel Foucault sur l'ordre du discours, ou le rapport au sens d'un LLM à travers l'analyse du postmodernisme proposée par Frederic Jameson.

Étude de cas: Tracer la conception de la "toxicité" le long de ces chaînes d'écriture comme exemple de *critical code studies* [25] [26] .

Étudier l'évolution d'une notion à travers différents textes. Qu'est-ce qu'un écrit "toxique"?

- Textes de questionnaires
- Rapports d'entreprises
- Textes de programmes
- Interfaces de visualization
- Jeu de données de test

Question du rapport entre fond et forme, exemple aussi du jeu de données: par Jigsaw, sur Kaggle. Qui vient originellement de Wikipedia. Le concept de "toxicité" est aussi quelque chose d'historiquement situé.

Cette approche là, se focalisant principalement sur les textes de programme, fait partie du champ émergent des critical code studies. Là, c'est aussi un exemple de la manière dont les études littéraires peuvent se développer d'avantage, en considérant d'avantages de textes comme pouvant faire partie d'un corpus.

Ce problème a également été addressé du point de vue de l'anthropologie linguistique, via les travaux de Gabriella Chronis, et qui pose la question de l'idéologie du langage.

Le logiciel est un langage exécutable qui réifie des idées, et ainsi incarne une sorte d'idéologie [27] [28] .

Le logiciel comme forme d'idéologie mérite aussi que l'on s'y attarde. Le double processus d'implémentation et d'éxecution fait que le logiciel est à même de matérialiser des idées, dans des actions de silicone. Une étude critique en amont permettrait alors d'identifier les idéaux et influences qui entrent en jeu lors du processus de conception et d'implémentation d'un LLM pour ensuite mettre à jour les différentes manières dont le langage est orienté.

Une approche de la réception

La dynamique et les conditions d'assignation de sens sont encore à élucider.

Et enfin, pour conclure, il serait intéressant d'appliquer la théorie de la réception. Quelles vont-être les relations de croyance par rapport au texte qu'on lit? Quels sont les imaginaires, les raisons, les désirs que l'on projette sur cet objet? Comment est-ce que cela va affecter notre propre perception de nous-même?

Il semble y a voir une corrélation entre anthropomorphisation et déshumanisation [29] .

Par exemple, Michael Burgess a mené une étude sur les processus d'humanisation (anthropomorphisation) des LLMs lors d'interactions discursives. Et il y a une forte corrélation etre l'attribution d'intelligence au LLM et la déshumanisation de soi. En d'autres termes, plus on anthropomorphise, moins on se sent compétent. Je pense que, là, ça peut avoir des compétences dans le processus d'enseignement des études littéraires, plus que dnas la recherche.

***TODO**: résumer du papier (de)humanization of AI*

Si il y a anthropomorphisation, quel est ce nouveau personnage?

Une des pistes pour étudier cette figure du personnage peut-être celle des stéréotypes, ou des archétypes.

- une version de l'"IA" (avec beaucoup de guillemets), ou d'un robot?
- une version d'un compagnon?
- une version de Dieu?

Et comment ces archétypes vont-ils se comparer à travers les cultures et les époques?

L'application des théories littéraires peuvent être fructueuses pour nous aider à comprendre ce que sont vraiment ces nouveaux écrits, ces nouvelles manières d'écrire, et ces nouvelles manières de lire.

Plutôt qu'utiliser cette nouvelle technique pour mieux comprendre le passé, il devrait être envisageable de mobiliser le passé pour mieux penser le nouveau, et par la même occasion, peut-être même repenser les outils conceptuels des humanités.

Épistémologie du LLM

Je vais maintenant conclure sur quelques éléments à prendre en compte pour analyser l'impact des grands modèles de langage sur les études littéraires.

La question de la littératie socio-technique. [30]

Il s'agit de comprendre pourquoi ca marche, d'abord pour éviter toute sorte de fétichisme magique. Cela n'implique pas uniquement de comprendre les maths, ou bien les processus d'entraînement, mais aussi les grandes dynamiques de présentation et d'utilisation de la technologie. Le problème de la récompense à court-terme vs. la récompense à long-terme (plus agréable de régler un problème technique, plutôt que de ramer sur un problème théorique) (exemple de la modélisation hydrologique)

Il s'agit aussi de comprendre l'idée du solutionnisme technologique: si les LLMs sont la réponse, quelle était la question?

Sur les questions d'éthiques de l'IA, Valérie Beaudouin et Julia Velkovska ont montré que ce débat a été principalement mis en place par des entreprises privée, et que le secteur public a suivi en toute hâte (spécifiquement au sujet de l'éthique). Il y a un courant de recherche, les critical hype studies qui vont chercher à mettre à jour les mécanismes qui vont susciter des fantasmes et des imaginaires autour de l'IA.

La question de la rigueur du processus de recherche. [31]

De manière plus prosaïque, il y a aussi la question de la reproducibilité.

D'abord, la reproductibilité? on n'arrive même pas à le faire avec des études normales, ni avec du code normal, alors avec des trucs qu'on ne comprend pas?? et il y a déjà une crise de la reproducibilité en général.

Ensuite, la durabilité dans le temps: de la recherche jusqu'à quand? dans 10 ans? dans 50 ans?

Si la bulle économique explose et on perd accès à certains de ces systèmes, on n'aura plus d'accès à certains services (tout comme les sciences sociales computationnelles se retrouvent un peu perdues depuis le rachat de Twitter), mais aussi on n'aura aucun moyen de vérifier.

Ce n'est pas une nouveauté des humanités numériques, mais ça continue dans une direction qui n'est pas souhaitable.

La question de la justice épistémologique.

Et puis il y a la question de ce qui est juste, à deux niveaux.

Premièrement, la justice épistémologique: un grand modèle de langage, c'est encore assez cher. Donc soit il y a le scénario où les coûts ne se démocratisent pas, donc c'est plus facile pour l'université Paris-Saclay de s'acheter des cartes graphiques que pour l'université de Port-au-Prince, et si en plus l'utilisation de grands modèles de langage est valorisée pour elle-même dans des publications académiques, alors on assistera encore plus au décrochage entre les universités riches et moins riches.

L'autre scénario, c'est la démocratisation, et une des conséquences, c'est l'empreinte carbone: celle-ci augmente très très fortement depuis la popularisation des modèles.

Il est difficile de quantifier **exactement** l'emprunte carbone du cycle de vie d'un LLM mais il est clair qu'elle **augmente** relative aux techniques précédentes.

Relativement parlant, ce n'est pas grand chose lorsqu'on compare à l'empreinte des déplacements en avion pour aller à des conférences. Mais, alors qu'on est entrain de foncer vers un monde à +3 degrés, ce n'est pas forcément la meilleure des directions à prendre, ce qui est pour moi un argument fort pour établir l'utilité épistémologique de ces modèles par rapport au confort qu'apporte son utilisation et ne l'utiliser que dans des cas où cela fait parfaitement sens.

Et on retrouve ici le problème de la dépendance au sentier déjà mentionné. Si Microsoft et Google construisent des réacteurs nucléaires pour leurs modèles de langage, il faut peut-être se poser la question: est-ce que les bénéfices des LLMs compensent vraiment leurs problèmes?

En guise de conclusion, je trouve que les sciences humaines sont peut-être plus adaptées à traiter ces grands modèles de langage comme **objet** d'études plutôt que comme **moyen** d'études.

Si toute technologie présente du négatif et du positif, il me semble qu'il y a plus de choses à gagner dans le premier cas de figure, plutôt que dans le deuxième cas de figure.

Néanmoins, puisque ce sont juste des réflexions, et pas vraiment des conclusions, je serais très curieux de savoir ce que vous en pensez, par exemple sur les aspects suivants:

- Pourquoi est-ce que ChatGPT a eu une adhésion aussi fulgurante?
- Qu'est-ce qu'on perd à ne *pas* utiliser un LLM?
- Qu'est-ce qu'on ne pouvait pas faire *réalistiquement* avant les grands modèles de langage?
- Qu'est-ce que peut faire l'humain, que la machine ne peut pas faire (ou ne *devrait* pas faire)?

- Quels sont les aspects essentiels du processus de recherche, et quels sont les aspects annexes?

Berry, D. (2011), *The Computational Turn* , Culture Machine.

MacLuhan, Marshall. 1964. *Understanding Media. The Extensions of Man* . MacGraw Hill.

Stiegler, Bernard. 1998. *Technics and Time, 1: The Fault of Epimetheus* . Stanford University Press.

Herrenschmidt, Clarisse. 2007. *Les Trois Écritures. Langue, Nombre, Code* . Bibliothèque Des Sciences Humaines. Gallimard.

Hayles, N. Katherine. 2012. *How We Think: Digital Media and Contemporary Technogenesis* . Chicago, IL: University of Chicago Press.

Foucault, Michel. 1975. *Surveiller et Punir* , Gallimard.

Curtis-Trudel, André. 2024. *On Finding What You're Not Looking For: Prospects and Challenges for AI-Driven Discovery* , TU Dortmund.

De Paoli, Stefano. 2024. *Performing an Inductive Thematic Analysis of Semi-Structured Interviews With a Large Language Model: An Exploration and Provocation on the Limits of the Approach* . Social Science Computer Review 42 (4): 997–1019.

Demszky, Dorottya, Diyi Yang, David S. Yeager, Christopher J. Bryan, Margarett Clapper, Susannah Chandhok, Johannes C. Eichstaedt, et al. 2023. *Using Large Language Models in Psychology* . Nature Reviews Psychology 2 (11): 688–701.

Rieder, Bernhard. 2024. *Les Modèles de Fondation sont des Plateformes* . Séminaire MetSem. Science Po.

Weizenbaum, Joseph. *Computer Power and Human Reason* , WT Books, 1973.

Clark, Andy, and David Chalmers. 1998. *The Extended Mind* . Analysis 58 (1): 7–19. https://doi.org/10.1093/analys/58.1.7.

Depaz, P. (2024). *Commons-based memories: Programming practices and large language models* . Memory Studies Review, 2(1). Birll Publishing.

Raphaël Gaillard, *L'homme augmenté: futurs de nos cerveaux* , Grasset, 2024.

Lederman, Harvey & Mahowald, Kyle. 2024. *Are Language Models More Like Libraries or Like Librarians? Bibliotechnism, the Novel Reference Problem, and the Attitudes of LLMs* . Transactions of the Association for Computational Linguistics 12:1087-1103.

Montfort, Nick, Patsy Baudoin, John Bell, Ian Bogost, and Jeremy Douglass. 2014. *10 PRINT CHR$(205.5+RND(1));: GOTO 10* . Illustrated edition. The MIT Press.

Fuller, Matthew, ed. 2008. *Software Studies: A Lexicon* . Cambridge, Mass: The MIT Press.

Reichardt, Jasia. 1968. *Cybernetic Serendipity* , Londres.

Gefen, Alexandre, et Claude Pierre Perez. 2019. *Extension Du Domaine de La Littérature.* Elfe XX-XXI Études de La Littérature Française Des XXe et XXIe Siècles.

Hicks, Michael Townsen, James Humphries, and Joe Slater. 2024. *ChatGPT Is Bullshit.* Ethics and Information Technology 26 (2): 38.

Fraenkel, Béatrice. 2006. *Actes écrits, actes oraux : la performativité à l'épreuve de l'écriture* . Études de communication 29 (1): 69–93.

Kittler, Friedrich A. 1997. "There Is No Software." In Literature, Media, Information Systems: Essays, John Johnston, 147–55. Amsterdam: Amsterdam Overseas Publishers Association.

Mu, Tong, Alec Helyar, Johannes Heidecke, Joshua Achiam, Andrea Vallone, Ian Kivlichan, Molly Lin, Alex Beutel, John Schulman, and Lilian Weng. 2024. "Rule Based Rewards for Language Model Safety." arXiv. https://doi.org/10.48550/arXiv.2411.01111.

Depaz, Pierre. 2024. *Shaping Vectors: Discipline and Control in Word Embeddings.* A Peer-Reviewed Journal About 13 (1).

David M. Berry, *Tracing "Toxicity" Through Code: Towards a Method of Explainability and Interpretability in Software* . 2023. DHQ, Volume 17 Number 2, 2023.

NLP as Language Ideology: Discursive and Algorithmic Constructions of 'Toxic' Language

Chun, Wendy Hui Kyong. 2005. *On Software, or the Persistence of Visual Knowledge.* Grey Room 18 (January):26–51.

Galloway, Alexander R. 2006. *Language Wants To Be Overlooked: On Software and Ideology.* Journal of Visual Culture 5 (3): 315–31.

Burgess, Michael. 2024. *Deceptive AI Dehumanizes: The Ethics of Misattributed Intelligence in the Design of Generative AI Interfaces.* In 2024 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), 96–108.

Beaudouin, Valérie, and Julia Velkovska. 2023. "Enquêter sur l'« éthique de l'IA »." Réseaux 240 (4): 9–27.

Ollion, Étienne, Rubing Shen, Ana Macanovic, and Arnault Chatelain. 2024. "The Dangers of Using Proprietary LLMs for Research." Nature Machine Intelligence 6 (1): 4–5.