Frontières de sens, frontières de langue :

les IA génératives et la question de la diversité des réponses pertinentes

Proposé par Alexandre Joux (Aix-Marseille Université, IMSIC)

En février 2023, trois mois seulement après la mise à disposition de ChatGPT pour le grand public, nous avons initié un enseignement reposant sur l'utilisation de ce chatbot.

Nous sommes partis d'un constat. La réponse unique proposée par ChatGPT à chaque question posée, aussi pertinente soit-elle, ce qui a expliqué son succès, semble en fait restreindre le champ des possibles. En effet, pour une question donnée, les réponses peuvent souvent être multiples, complexes, et donc difficiles à donner en une fois, de manière assertive.

Il ne s'agit pas là d'une limite évidente de l'IA générative car il est possible de lui faire produire plusieurs réponses à une question donnée. Il s'agit là des limites propres à une utilisation naïve, peu raisonnée, de l'IA générative, qui considère une réponse unique et pertinente comme une réponse suffisante, parce que l'on délègue à l'IA générative le soin de synthétiser « l'existant ». C'est cette attitude naïve qu'il s'agissait de révéler à travers notre expérimentation pédagogique.

Le cours de master en question, dédié aux théories des SIC, a donc été adapté. Une liste d'auteurs et de notions a été présentée aux étudiants, ce qui a permis de dresser une histoire sommaire mais représentative des grands moments de l'histoire des SIC *lato sensu*. Par exemple, le modèle de la piqûre hypodermique chez Lasswell, la théorie des leaders d'opinion chez Lazarsfeld, le déterminisme technique chez McLuhan, le codage / décodage chez Stuart Hall, la notion de société de l'information, la notion d'ethnoscape chez Arjun Appadurai, la notion de diversité culturelle.

Nous avons ensuite constitué des groupes d'étudiants (*a minima* un binôme) chargés chacun d'explorer une notion grâce à ChatGPT. Ensemble, nous avons établi la question à poser à ChatGPT selon l'auteur et / ou la notion choisis. Cette même question a ensuite été posée à ChatGPT à partir de deux comptes distincts, en anglais et en français. Les réponses, selon les auteurs et les notions, étaient souvent proches, parfois très différentes. Nous avons alors demandé aux étudiants de comparer les réponses en anglais et en français fournies par ChatGPT et d'expliquer pourquoi elles sont similaires ou non, ceci à partir d'une compréhension de la manière dont fonctionnent les IA génératives. Il a donc fallu introduire dans le cours une explication du fonctionnement global des IA génératives.

Les IA génératives « apprennent » à partir d'immense jeux de données. Mais leurs réponses, dans un premier temps, si elles partent de traces écrites numérique humaines, sont souvent inadaptées. Il faut qu'elles deviennent « pertinentes » pour l'humain. C'est pourquoi leur entraînement repose aussi sur l'injection de « feddbacks » humains (des gens sont payés pour corriger et / ou valider les réponses des IA jusqu'à ce que ces réponses finissent par être jugées globalement pertinents et correctes).

Cette « pertinence pour l'humain » comme objectif assigné aux IA génératives, a des conséquences. Elle invite les IA à privilégier, quand une question leur est posée, les corpus de données dans la langue de l'utilisateur, ceci afin de coller au mieux, dans la fabrique de la réponse, à son univers de référence. Quand les ressources sont insuffisantes dans la langue de l'utilisateur, alors la réponse tirée de l'exploitation des corpus anglosaxons l'emporte et est traduite.

C’est ce qu’a permis de constater notre expérimentation pédagogique de 2023. Pour certains auteurs, les réponses en anglais et en français varient très fortement car ce que permettent les données varient fortement d’une langue à l’autre, d’un univers de référence à l’autre.

Il ne s’agit pas ici de « biais », ce que peuvent aussi proposer les IA génératives quand elles reproduisent des stéréotypes, des jugements de valeurs, des associations d’idées problématiques. En effet, nous avons testé des notions d’auteurs scientifiques qui ont fait l’objet d’une discussion argumentée de la part de la communauté des chercheurs.

Si nous avons des réponses différentes, c’est donc parce qu’il y a, dans la communauté scientifique, une forme de polysémie, une diversité d’interprétations, de « lectures » possibles pour un même auteur. En l’occurrence, cette différence s’est fortement manifestée sur la notion d’ethnoscape et son lien avec celle d’appropriation culturelle chez Arjun Appadurai. Dans l’univers anglo-saxon, les *cultural studies* inscrivent la question de l’appropriation dans un un rapport de domination quand, dans l’univers francophone, l’importance de la notion de diversité culturelle conduit à valoriser l’idée de cultures créolisées, sans souligner trop fortement les logiques inégales dans la capacité d’emprunts et d’appropriations d’une culture à l’autre.

Autant dire que, même sur des questions scientifiques, il y a une sorte de « bulle culturelle » propre aux IA génératives. Ces dernières nous « enferment » dans des frontières de sens qui reposent sur la langue utilisée … et sur les données disponibles dans cette langue. En effet, évident sur Arjun Appadurai où le conflit d’interprétation s’explique par des corpus scientifiques en français et en anglais de nature différente, cet « enfermement » prend potentiellement un visage différent quand les corpus de données utilisés par les IA génératives ne permettent pas d’établir une réponse à partir de la langue de l’utilisateur. Dans ce cas, une traduction de la réponse en anglais risque d’être proposée car les IA génératives ont été entraînées d’abord à partir de données en anglais. L’enfermement culturel prend alors soit la forme d’une ouverture, soit la forme d’une domination (là encore, comme chez Arjun Appadurai, il y a plusieurs lectures possibles) : il peut être considéré comme une ouverture à l’univers de référence anglo-saxon, mais cette ouverture est « à sens unique » (l’enferment est paradoxal), ou il peut être considéré comme une forme de domination de l’univers de référence anglo-saxon (enfermement strict).

Si, dans l’expérimentation conduite en 2023, nous avons pu mettre en évidence la polysémie des réponses des IA génératives selon la langue posée, parce que les données existaient soit en anglais, soit en français, nous ne sommes pas véritablement parvenus à identifier, alors, des réponses qui, de toute évidence, imposent un contexte interprétatif étranger à l’univers de référence du locuteur.

Ces conflits-là nous ont été signalés par des utilisateurs de l’IA générative, collègues par ailleurs, qui travaillent au Liban. En arabe, les IA génératives produiraient régulièrement des réponses « hors de propos », en dehors donc des frontières de sens au sein desquelles l’utilisateur considère que les réponses proposées sont effectivement pertinentes pour lui. Ici, la transgression de la frontière de sens serait trop importante pour que les IA soient considérées comme véritablement utiles. Elles iraient au-delà de l’ouverture, jusqu’à l’incompréhension. Les IA génératives de type ChatGPT auraient du mal à maîtriser le sens / la signification de nombreux mots en arabe – elles se méprennent donc sur les questions posées et dans la manière de répondre. Les données en arabe sont moins nombreuses et moins exploitées par les IA, ce qui favoriserait des traductions maladroites de réponses issues de corpus anglais.

Pour explorer cette possibilité, et dans le cadre d'une masterclass organisée en avril 2025 avec des étudiants français et libanais, le même exercice que celui réalisé en 2023 sera reproduit mais cette fois-ci en français, en anglais et en arabe, sur des notions en SIC et sur des actualités internationales récentes.

Nous comptons présenter les résultats de cette seconde expérimentation, en lien avec les conclusions établies à partir de l'expérimentation réalisée en 2023, lors de la 6ème édition du colloque Frontières numériques.