

Communiqué de la SIF sur ChatGPT et les modèles génératifs

La mise en accès ouvert de ChatGPT par la société OpenAI a réussi le double tour de force d'amener à la une des médias le sujet des modèles génératifs ("MG" – c'est-à-dire les modèles d'IA qui génèrent du contenu) en intelligence artificielle, mais aussi de provoquer une curiosité, des réactions, et des inquiétudes parfois extrêmes du public. La Société informatique de France prend position par ce communiqué : au-delà d'une forme d'engouement massif et non raisonné pour une technologie déjà en gestation depuis des années, l'impact sociétal potentiellement considérable de celle-ci demande à être accompagné.

Que sont les modèles génératifs de texte ?

Les MGs de texte comme ChatGPT (GPT signifiant Generative Pre-trained Transformer ou Transformateur génératif pré-entraîné) sont construits par l'apprentissage d'associations de mots dans leur contexte, c'est-à-dire dans des phrases, et ce sur des quantités de données gigantesques. Leur tâche est de prévoir et de proposer les meilleurs mots et phrases suivant ceux proposés par un utilisateur. L'avancée de ChatGPT (mais aussi celle de Bard de Google, ou de tous les autres déjà annoncés) vient du modèle de langage GPT-3 (LAMBDA pour Bard), entraîné sur des volumes massifs de phrases prises sur le Web, suivi d'une phase complémentaire pour affiner les "meilleures" réponses à proposer. Ainsi, ChatGPT propose un modèle de langage complété d'une fonctionnalité d'agent conversationnel (chatbot).

Nous sommes donc ici face à un système qui s'appuie sur des probabilités pour générer la suite la plus probable au regard des mots qui précèdent et des données d'apprentissage en rapport avec la question posée. Par exemple : "je suis étudiant... à l'université" est plus probable que "je suis étudiant... à Bali". Il faut bien insister sur le fait que ceci ne résulte en rien d'un mécanisme "intelligent" ou "conscient" doté de la capacité de raisonner. Les MGs manipulent ainsi des suites de mots, mais pas des assertions logiques ou des éléments de calcul : la notion de justesse ou de vérité du résultat proposé est totalement absente du mécanisme en question. De ce fait, les MGs sont aussi "bons" (ou "mauvais") que les données ayant servi à les entraîner : ils ne savent que ce qu'ils ont appris. Par exemple, ChatGPT, entraîné sur des données jusque fin 2021, ignore tout ou presque de ce qui s'est passé depuis, même si l'agent conversationnel qui lui est associé continue à évoluer en continu grâce aux données des conversations avec les utilisateurs.

Il est également important de souligner que de tels modèles génératifs sont presque tous fournis et opérés actuellement par quelques très grandes entreprises (Microsoft, Google, Meta-Facebook, Baidu). Ils ne sont ainsi pas neutres : le commanditaire a en pratique toute latitude de les modeler sur les sujets qui lui importent ou ne lui nuisent pas, stratégiquement ou économiquement ; c'est lui qui décide ce qu'il est convenable de dire ou ne pas dire, comme c'est déjà le cas pour les moteurs de recherche. Lesquels moteurs de recherche vont d'ailleurs se trouver profondément transformés par l'incorporation de MGs, comme le montre la guerre désormais déclarée entre Microsoft-ChatGPT et Google-Bard.

Des usages qui se multiplient

Comme avec les outils génératifs d'images (type Dall-E, MidJourney, Stable Diffusion), les possibilités offertes par les MGs de texte comme ChatGPT sont nombreuses, et de nouveaux usages apparaissent quotidiennement. On peut imaginer que cela va transformer les tâches de rédaction de textes : courriels, rapports techniques, articles journalistiques, etc. L'écriture de code informatique et plus généralement le développement logiciel le serait également comme le montre l'introduction par Microsoft de Copilot dans GitHub. Certains de ces usages potentiels sont bien reçus, d'autres moins. En effet, ces outils soulèvent de nombreuses questions comme le risque de perturber l'emploi, ou des problèmes de copyright (rappelons que le modèle de langage est entraîné sur des données du Web, qui peuvent être protégées). Dans des domaines comme l'éducation ou les ressources humaines, ils questionnent aussi les formes classiques d'évaluation, depuis les notations d'examen jusqu'à l'appréciation de lettres de motivation.

Pour tenter de dédramatiser le débat, une analogie peut cependant être faite entre l'apparition de ces outils et l'invention des dictionnaires à partir du XVII^e siècle : l'apparition du dictionnaire simple a été suivie par celle du dictionnaire des synonymes, puis de celui des citations – le plus proche des outils que nous discutons ici. Jugée comme disruptive au départ, l'utilisation de ces dictionnaires a été progressivement intégrée au processus rédactionnel courant. Une différence majeure avec les MGs de texte est la nature aléatoire de ces derniers, qui permet des variations significatives dans les réponses lorsque des questions similaires sont posées. Si les textes produits n'ont aucun caractère de vérité, ils peuvent néanmoins constituer une excellente source d'inspiration, ou permettre d'étoffer et de mettre en forme rapidement une trame grossière donnée par l'utilisateur, laissant ce dernier se concentrer sur l'aspect le plus créatif de son propos. Ces MGs doivent ainsi être vus comme de simples assistants de rédaction de textes qui doivent toujours être relus, vérifiés, complétés par l'utilisateur.

Une telle assistance a cependant le travers de pouvoir se faire passer pour ce qu'elle n'est pas. Une classe d'élèves pourrait s'en servir pour répondre au même sujet de dissertation, sans que le professeur n'ait l'impression de corriger deux fois la même copie (en effet, il est encore impossible de détecter automatiquement qu'un texte a été produit par un MG, même si des recherches dans ce sens sont en cours). Des enseignants pourraient ainsi se trouver en difficulté face à ces outils, dont ils soupçonneraient l'usage sans pouvoir véritablement le prouver. De même, ces technologies pourraient permettre à un lobby de créer automatiquement un grand nombre de pages web défendant – mais de manière légèrement différente – la même thèse. Plutôt que de vouloir interdire les MGs, la SIF recommande un accompagnement de leur usage pédagogique, qui passe bien sûr par une phase de formation et de démystification : une charte d'usage devrait alors être proposée pour réguler le recours aux MGs.

Conclusion

De nombreuses questions sont toujours en suspens. Tout d'abord, ces outils nécessitent des moyens considérables pour leur entraînement puis leur utilisation, qu'il faut questionner à l'heure d'une nécessaire sobriété numérique (empreinte carbone tout particulièrement). Les avancées de la recherche en informatique permettront-elles à terme de réduire ces coûts ? De plus, ils sont pour le moment aux mains de grands industriels non européens. Au-delà de la question des biais dans les résultats fournis, l'avantage compétitif, mais surtout les impacts géopolitiques attendus de l'accroissement de leur efficacité et de leur usage devraient pousser l'Europe à prendre les mesures nécessaires pour maîtriser ses propres outils. Pourrait-on alors imaginer de concevoir et de mettre en ligne, au niveau européen, des modèles génératifs au service de tous, fondés sur la recherche publique, le partage, la mise en commun de logiciels et de données d'apprentissage ? La Commission européenne comme l'Agence Nationale de la Recherche viennent d'ailleurs de lancer des appels à projet de recherche sur les MGs.

Pour la génération de texte, le génie est sorti de sa boîte : ces outils sont une opportunité créative, permettant de réduire les tâches répétitives et les coûts, et on voit mal pourquoi on devrait s'en passer. L'esprit critique s'impose plus que jamais sans bloquer pour autant la créativité que ces MGs peuvent stimuler. On devra sans doute en réguler les utilisations problématiques, mais surtout, il convient aux institutions d'accompagner les usagers, et de faciliter la compréhension des objectifs et des limites de ces systèmes. La SIF plaide donc pour la formation des jeunes aux outils numériques et à l'informatique, et ce dès l'école et le collège, et plus généralement pour un accompagnement éclairé de la population, tout au long de la vie et pour le plus grand nombre.