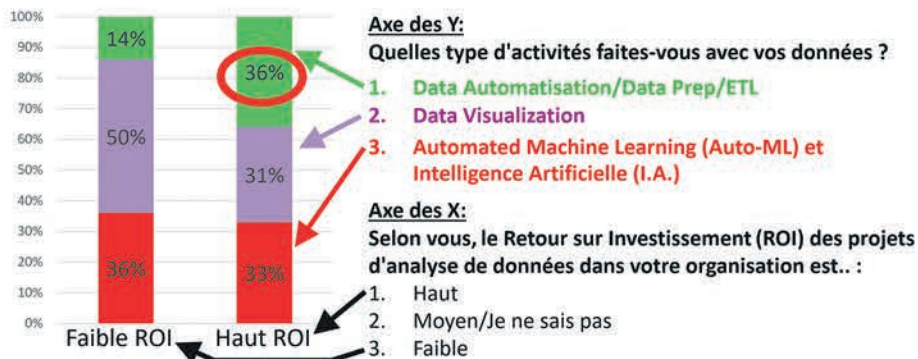


IA ET « LARGE LANGUAGE MODEL » FUTUR FLOP OU RÉVOLUTION ?



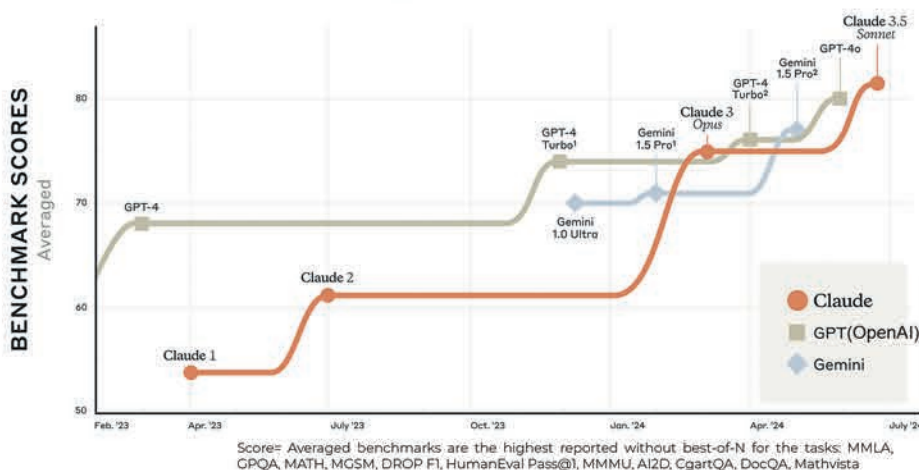
Il est incontestable qu'à long terme, l'IA et le Big Data vont fondamentalement améliorer nos vies. Actuellement, lorsque l'on évoque l'IA, on se réfère presque systématiquement à une technologie récente appelée "LLM" ("Large Language Models"). Toutefois, il convient de rappeler que les logiciels d'IA existent depuis plus de 50 ans sous divers autres noms (logiciels de statistique, data mining, machine learning, etc.). Récemment, une promesse récurrente émerge: Nous allons pouvoir tout automatiser en confiant aux LLM les tâches ennuyeuses et répétitives que personne ne veut faire.

Selon une enquête réalisée en Septembre 2023 par le bureau d'étude indépendant "Into The Minds", les sociétés qui arrivent à extraire le maximum de ROI de leurs données dédient la plus grande partie de leur temps de travail (36%) à remplir une activité particulière : Faire de l'automatisation de flux de données (aussi parfois appelé "Data Prep."). Cela est visible sur ce graphique qui provient de l'étude:



Malheureusement, pour l'instant, les LLM ne sont pas suffisamment fiables pour réaliser ces automatisations. En effet, selon Sam McAllister, un "AI Scientist" travaillant chez Claude.ai, un LLM produira en moyenne des "hallucinations" 20% du temps (1). En d'autres termes, Les LLM effectuent correctement leurs tâches 4 fois sur 5 (80% de réussites):

AI model release and capabilities timeline



Vous allez dire: « Il y a certainement des domaines d'utilisations où les modèles LLM sont plus fiables ? ». De par leur nature "orientée texte", on peut penser que les LLM sont idéaux pour répondre à des questions de type juridique. Et, effectivement, OpenAI (la société derrière le fameux chatGPT) rapporte que son LLM excelle dans ce domaine, avec un score de 90% à un Examen du Barreau. Sauf que, si on examine attentivement les résultats de cet LLM, on constate que son score est de seulement 15%. (2).

Des Démonstrations truquées?

OpenAI n'est pas la seule société à déformer "légèrement" la réalité au sujet des performances de ses IA. Par exemple, on citera les cas des 3 démos IA truquées de Google(3), de la fausse technologie « Just walk out » de Amazon (4), de la démo truquée de pilote autonome de tesla (5) ou de GM (6), du chatbot de facebook (7) (qui était en fait des humains, 70% du temps). Dans la Silicon Valley, les fraudes à l'IA sont omniprésentes. En effet, il faut absolument satisfaire les investisseurs qui ont tout misé sur l'AI pour tenter de reproduire le ROI des FAANG (Facebook, Amazon, Apple, Netflix, Google). Et pour ce faire, il faut vendre à tout prix, en offrant

des "promesses" toujours plus grandes. Ce n'est pas la première fois que la Silicon Valley veut absolument nous vendre à vil prix des technologies "révolutionnaires" qui ne délivrent pas les promesses attendues: blockchain, Web 3.0, IoT, Cryptomonnaies, réalité augmentée, Hadoop/Spark, etc. Après autant de promesses brisées, nous sommes en droit de nous poser des questions sur la légitimité des LLMs.

Hadoop/Databricks: le flop

J'aimerais revenir en particulier sur la promesse de "Hadoop/Spark/Databricks" qui était de pouvoir exploiter plus facilement et à moindre coût le Big Data. À ce sujet, je cite Loïc, spécialiste Big Data indépendant: « Les outils disponibles dans Hadoop/Databricks tels que Spark ont des performances inférieures aux bases de données des années '80 sur du hardware d'époque. » Du côté de TIMi, on partage la même conclusion, celle-ci étant avalisée par des tests effectués sur les outils Databricks/Spark/Hadoop et fondée sur des benchmarks universitaires neutres(8) (TCP-H). Le résultat est sans appel : une inaptitude presque totale à traiter de grands volumes de données. Pire, les bénéfices générés par le Big Data sont insuffisants pour équilibrer les coûts liés aux licences Hadoop/Databricks et au hardware correspondant (sans parler des frais RH!). De nombreuses startups basées sur le Big Data font donc faillites ou accumulent des dettes colossales. Un exemple est Uber avec \$1823 millions de dettes en 2022. En réaction, certains vont jusqu'à qualifier tout le big data en général de "hoax" (9). Non, avec les outils appropriés, le Big Data n'est pas un Hoax. Le potentiel est immense. La police peut prévenir les crimes avant qu'ils ne se produisent, les chercheurs peuvent détecter les cancers avant qu'ils ne se propagent et les entreprises peuvent améliorer leurs produits, prendre des décisions optimales et acquérir un avantage commercial considérable. Bref, Hadoop/Databricks est un échec retentissant, une arnaque pure, mais c'était, malgré tout, le centre de toute la "Hype" jusqu'en 2020. Or, les mêmes mécanismes de marché qui ont présidé à l'essor de Hadoop se retrouvent aujourd'hui dans le domaine des LLM, chatbot et C°.

Des LLM: Pour faire quoi?

Les LLMs sont utiles pour générer rapidement des textes marketing accrocheurs, ou pour améliorer les résultats de recherches sur des corpus de documents. Le standard en termes de moteur de recherche est "Elastic Search"(en abrégé: ES). Au prix d'efforts considérables, il est possible (avec la technique RAG-10), de combiner ES et LLM pour obtenir des résultats très légèrement supérieurs à un simple ES gratuit. Reste à voir si ces efforts coûteux pour améliorer les résultats de recherche de ES sont justifiés en termes de ROI ("press X to doubt"). Par contre, en termes d'automatisation de processus data, il faut oublier les LLM sous peine de voir tous les processus s'effondrer à la première "hallucination". Or, c'est par l'automatisation que vous pouvez extraire le ROI caché dans vos données.

TIMI: une alternative sérieuse

Depuis 2007, TIMi a lancé la plateforme data la plus puissante du marché pour repousser les limites de l'analytics, de l'advanced analytics, de l'IA et du Big Data, tout en offrant un environnement convivial, rapide et productif. TIMi est une société de haute technologie qui a inventé le 1er outil de Auto-ML (automated machine learning) de la planète. En 2011, TIMi a créé Anatella: une plateforme qui permet de faire toute "automatisation data" en self-service en quelques clics. Anatella est aussi la solution Big Data financièrement la plus rentable au monde. En place dans de nombreux secteurs d'activités, nos solutions manipulent et analysent chaque jour des exabytes de data sans faillir, et ce dans le monde entier.

Mon conseil aux entrepreneurs: Évitez les services AI basés sur des LLM. De toute façon, dans le cas où les LLM prendraient une trajectoire vertigineuse, la création de chatbots aléatoires ne vous préparera pas à l'avenir. Privilégiez les outils "data" qui sont fiables et rentables, tels que les nôtres.

Infos et téléchargement sur timi.eu

(1) <https://x.com/sammcallister/status/1803791750856634873>
 (2) <https://link.springer.com/article/10.1007/s10506-024-09396-9#Sec11>
 (3) Google faked the "Duplex" AI demo (<https://www.axios.com/2018/05/17/google-ai-demo-questions-faked-the-gemini-ai-demo>) and faked the "Gemini" AI demo (<https://techcrunch.com/2023/12/07/googles-best-gemini-demo-was-faked/>) and faked the "Overview" AI demo (<https://www.nytimes.com/2024/05/24/technology/google-ai-overview-search.html>)
 (4) <https://www.businessinsider.com/amazons-just-walk-out-pullback-shows-ai-way-to-go>
 (5) https://www.theregister.com/2023/01/18/tesla_selfdriving_video_faked/
 (6) <https://www.nytimes.com/2023/11/03/technology/cruise-general-motors-self-driving-cars.html>
 (7) <https://www.technologyreview.com/2017/04/14/152563/facebook-perfect-impossible-chatbot/>
 (8) <https://github.com/Kran199/TPC-H-Benchmark-Anatella-Spark>
 (9) The video "Why AI is Tech's Latest Hoax" available at <https://youtu.be/pOubCK8XMC8>
 (10) <https://www.elastic.co/what-is/retrieval-augmented-generation>

