TEXTE de l'ARTICLE de PRESSE signé
ALEXIS FAVRE PRODUCTEUR D'INFRAROUGE (TSR / CH)

Je me lève ce matin avec une pensée émue pour tous les gens qui gagnent du temps au quotidien grâce à l'intelligence artificielle en général, et les grands modèles de langage en particulier. Pour tous les enthousiastes qui ont délégué la rédaction de tout ou partie de leur récit personnel à ChatGPT, Grok, Claude, Copilot ou je ne sais quel autre robot malin. Pour tous ceux qui, grâce aux miracles du numérique, n'ont plus besoin de se fatiguer à écrire un mail, à rédiger une lettre, à mettre de l'ordre dans leurs idées. Cette semaine, ils tombent de haut, s'ils lisent encore les journaux, en l'occurrence celui-ci.
Après avoir testé dix des principaux robots conversationnels, la plate forme NewsGuard-spécialisée dans l'analyse de fiabilité des sites d'information est arrivée à une conclusion sans appel: ils répètent tous sans broncher les wagons de désinformation dont les abreuve avec application la modernité mal intentionnée. En particulier les 3,6 millions de faux articles savamment élaborés par un réseau pro-Kremlin, forcément baptisé Pravda, Bêtes à manger du foin (et donc à en recracher), les chatbots d'Open Al, d'Elon Musk, de Microsoft ou de Mistral s'avèrent donc incapables de trier le bon grain de l'ivraie informationnelle et disent n'importe quoi quand ils ont appris n'importe quoi... Ça alors! Qui aurait bien pu s'imaginer que des fonctions mathématiques conçues pour mouliner des données soient à la merci de la qualité de ces données? Les bras nous en tombent.
Très concrètement, si vous demandez par exemple à l'une de ces petites merveilles de technologie de masse si Volodymyr Zelensky a tapé dans la caisse de l'aide à l'Ukraine pour son enrichissement personnel, il y a de bonnes chances qu'elle vous le confirme, et même qu'elles vous fournissent une preuve de sa bonne foi, préalablement boutiquée par les hackers de Moscou.
Si nos amis les robots se font si facilement enfler par de zélés Moscovites et leurs réseaux, il n'est pas complètement absurde de s'interroger sur la qualité de l'output, en bon français, qu'ils délivrent par ailleurs sur tous les autres sujets. Et se poser cette question revient à y répondre sans grand risque d'erreur cette fois: les grands modèles de langage donnent parfois, et peut-être même souvent, des réponses satisfaisantes, mais pas toujours. Reformulé: ils sont toujours fiables, sauf quand ils ne le sont pas.
Tout intuitif soit-il, ce constat objectivement problématique peut inspirer deux réactions possibles aux promoteurs de ces technologies, comme à la horde grandissante de ceux qui s'y fient. La première relève de l'ingénierie, et pour tout dire de la fuite en avant. Elle consiste à identifier les erreurs quand elles apparaissent, et à corriger patiemment le système pour qu'il évite de dire des bêtises.

Cette option assurera du travail à beaucoup d'informaticiens pendant très longtemps. Très précisément le temps dont ont besoin les Danaïdes pour remplir leur tonneau. Elle supposera aussi, cette fois du côté des utilisateurs, de se méfier un chouïa, mais systématiquement, des conclusions automatiques de la machine, au risque de gagner un peu moins de temps que prévu, voire d'en perdre beaucoup.

L'alternative, que j'ose à peine esquisser ici, consisterait à prendre un peu de recul face aux développements supposément fulgurants de l'intelligence artificielle. Et à réaliser peut-être, et fût-ce dans la douleur, qu'il est un peu tôt pour lui donner les clés de la maison.

Concrètement, et par exemple toujours, cela pourrait conduire à préférer suer parfois un peu sur un problème ou sur un autre, plutôt que d'attendre une réponse toute cuite, pour ne pas dire magique. Un choix qui suppose quelques efforts, c'est vrai, et qui est assorti de risques indéniables, comme celui d'apprendre quelque chose au passage et de mourir moins idiot.

Mais il faut parfois savoir vivre dangereusement.