



TECH

Le cloud, une arme majeure dans la grande bataille de l'IA

Les Big Tech sont fournisseurs d'un service essentiel mais aussi acteurs.

INGRID VERGARA  @Vergara_j

INTELLIGENCE ARTIFICIELLE Derrière la « magie apparente » des réponses fournies par ChatGPT, il y a bien sûr ce qu'on appelle des « grands modèles de langage », intégrant des milliards de paramètres. Mais il y a, avant tout cela, de massives capacités de puissance de calcul, seules capables de permettre l'entraînement de ces algorithmes. Pas d'intelligence artificielle générative sans une infrastructure cloud ultra-performante, capable de supporter les volumes et la fréquence des requêtes exigées.

Depuis 2019, date du premier accord entre Microsoft et OpenAI, la société d'intelligence artificielle a développé tous ses modèles GPT-3, Dall-E et ChatGPT sur Azure, l'infrastructure cloud du géant américain. « Nous sommes heureux d'être leur fournisseur exclusif de services cloud. Nous déploierons leurs modèles dans nos produits, tout en continuant à faire progresser l'état de l'art de l'IA. Toutes ces innovations stimulent la croissance de nos ser-

vices Azure AI », se félicitait la semaine dernière Satya Nadella, le PDG de Microsoft en marge des résultats.

Outre une prise de participation de 10 % dans Anthropic, un concurrent d'OpenAI, Google a signé en parallèle un partenariat technologique pluriannuel sur la partie cloud, comme il l'a fait avec d'autres start-up prometteuses comme Cohere et C3.ai. De son côté, Amazon est devenu le « partenaire cloud privilégié » pour le développement des modèles de Stability AI. « Il est crucial pour les start-up de disposer de la puissance de calcul nécessaire pour construire des systèmes d'IA fiables et dignes de confiance », soulignait Sundar Pichai, le PDG de Google, dans une note de blog lundi.

De fait, ces jeunes sociétés d'intelligence artificielle n'ont pas vraiment d'autres choix que de se tourner vers ces grands acteurs pour accéder à la puissance de calcul nécessaire, à un coût supportable pour elles. Construire leur propre infrastructure de données nécessiterait

des investissements tellement considérables, qu'ils sont même hors d'atteinte pour des fournisseurs de cloud installés. « Dans l'offre actuelle des cloud providers européens, il n'y a pas les processeurs nécessaires à faire tourner de puissantes IA (les cartes GPU de Nvidia), pourtant disponibles depuis deux ans chez les Big 3. Impossible pour une société comme la mienne d'y faire tourner nos modèles - installé donc, pour nos clients - sur un hyperscaler américain », témoignait Laurent Daudet, directeur général de LightOn, en novembre dernier.

En contrepartie, les start-up d'IA concèdent un aperçu sur leur savoir-faire et de leurs technologies, et lient leur destin à des groupes, qui disposent par ailleurs de leurs propres équipes de recherche sur l'intelligence artificielle. Google Cloud et AWS ont beau se prévaloir que les accords conclus sont non exclusifs, il sera techniquement compliqué et coûteux pour ces sociétés de changer de fournisseur, une fois la construction de ces nou-



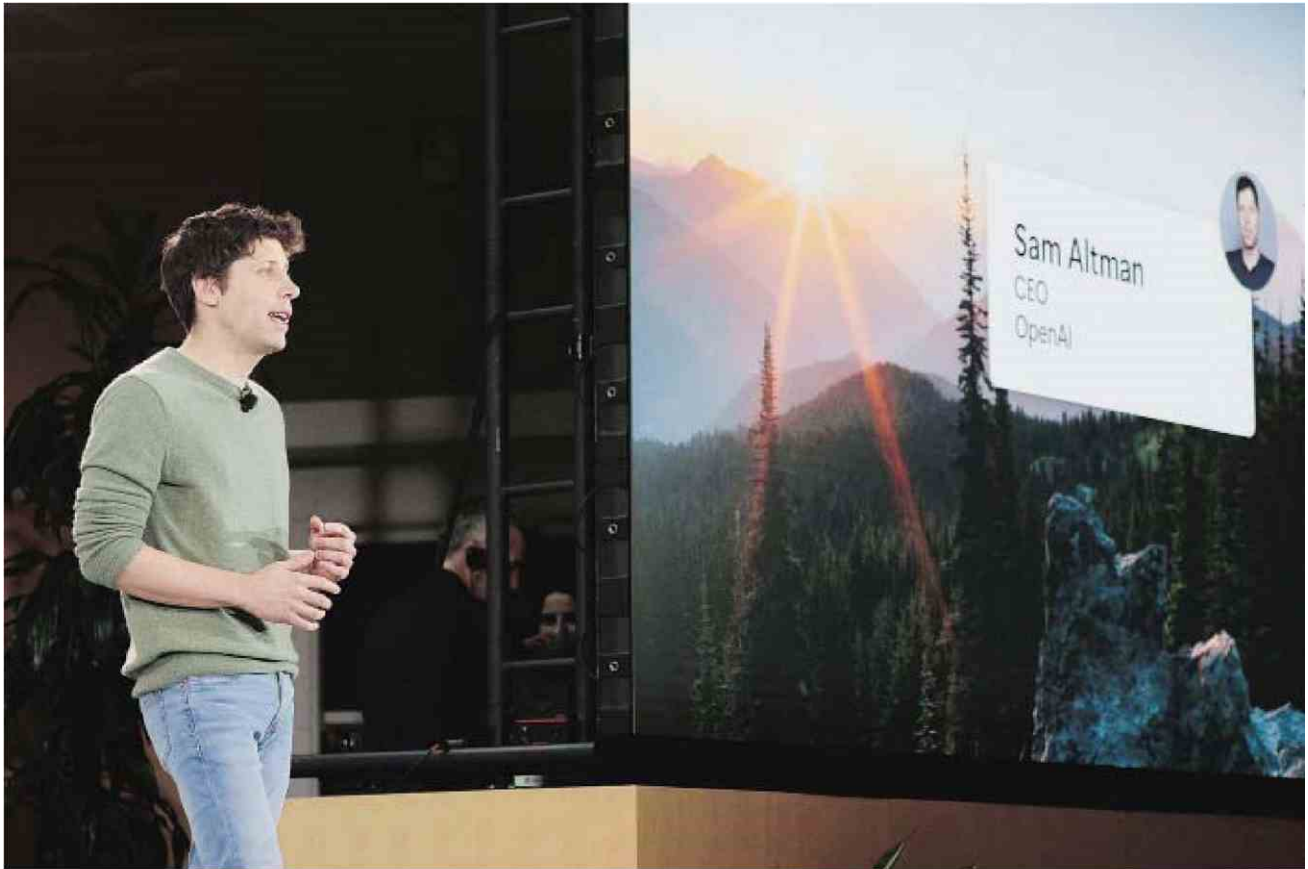


veaux méga modèles de langage engagée.

Enquête des régulateurs

Des start-up à la pointe de l'innovation, dépendantes d'un contrat de fourniture cloud auprès de géants qui sont aussi des concurrents... Ce scénario aux airs de « déjà-vu » dans le secteur technologique alimente les inquiétudes sur la façon dont ces grandes sociétés pourraient contrôler l'émergence d'une nouvelle génération de concurrents dans les mois à venir. C'est ce contre quoi veulent lutter les régulateurs des deux côtés de l'Atlantique, qui scrutent attentivement depuis quelque temps ce qui se passe sur le marché du cloud. En France, l'Autorité de la concurrence doit rendre prochainement ses conclusions après un an d'enquête sur les pratiques dans le secteur. La Commission européenne a été saisie de plaintes d'acteurs européens et scrute certains accords cloud signés par des géants. Le nouveau règlement européen DMA (Digital Market Act) doit aussi en théorie soumettre les services cloud à une obligation « d'interopérabilité », autrement dit faciliter le passage d'un acteur à un autre. Beaucoup d'acteurs demandent à voir... Pour éviter de se retrouver poings liés, d'autres start-up d'intelligence artificielle, à l'instar du français NuKKai, explorent d'autres voies pour l'avenir, avec de nouvelles générations d'IA moins gourmande en puissance de calcul pour se développer. ■





Sam Altman, le PDG d'OpenAI, lors d'un discours annonçant l'intégration de ChatGPT pour Bing, mardi, au siège de Microsoft à Redmond, Washington.

JASON REDMOND/AFP

