La Route Moins Empruntee

Repenser les couts, la securite et la confiance lies a l'IA generative dans les institutions reglementees

POURQUOI CETTE CONVERSATION EST IMPORTANTE

Les institutions se font dire que l'intelligence artificielle est inevitable. Pas seulement utile - inevitable. Le message est constant : l'IA va transformer le fonctionnement des organisations, mais il vous faudra les bonnes protections.

Ce qu'on remet rarement en question, c'est de savoir si les institutions sont invitees a s'adapter a un modele industriel qui n'a jamais ete concu pour elles. L'objectif ici n'est pas de critiquer les outils et produits d'IA standard, ni de soutenir qu'une approche est universellement superieure. L'objectif est d'expliquer pourquoi les organisations reglementees pourraient avoir besoin d'un type fondamentalement different de systeme d'IA - et pourquoi cette difference compte davantage que les fonctionnalites, les modeles, les arguments commerciaux ou le prestige des fournisseurs.

Pour mieux illustrer le probleme, j'utiliserai deux metaphores simples que j'ai explorees tout au long du developpement de COMPAiSS, un assistant IA concu pour les environnements institutionnels :

La Matrice IA
l'ensemble des hypotheses cachees qui faconnent la conception de la plupart des systemes d'IA.

La Route (et la Route Moins Empruntee)
la maniere dont ces hypotheses se traduisent en couts, en risques, en correctifs recommandes, en arguments de vente et en adequation institutionnelle.

LA MATRICE IA : DES HYPOTHESES QUE TOUT LE MONDE ACCEPTE (SOUVENT SANS LE REMARQUER)

La plupart des systemes d'IA modernes sont construits a l'interieur de ce qu'on peut raisonnablement appeler un seul cadre de pensee dominant. Dans ce cadre, on suppose que l'IA reflechit en permanence et est toujours 'en marche'. Elle peut raisonner sur le monde entier - l'internet, les connaissances generales, les faits implicites, les analogies et les suppositions. La securite, c'est quelque chose que l'on gere apres que la reflexion a commence, pas avant, et tous les risques correspondants sont acceptables du moment qu'ils peuvent etre filtres, moderes ou refuses.

Ce cadre de pensee n'est pas malveillant - c'est le principe fondateur accepte de l'IA grand public, de l'IA creative et de l'IA generaliste. Et cela est logique si votre objectif est de vendre, de fideliser et de renforcer une intelligence ouverte. Une fois entierement immerge dans cette matrice, certaines choses deviennent incontestables :

- bien sur que l'IA doit reflechir d'abord.
- bien sur que des erreurs se produiront.
- bien sur que les hallucinations sont inevitables.
- bien sur que vous avez besoin de couches de controles pour gerer les risques.
- bien sur que les couts augmentent avec la complexite.

LA QUESTION QUI N'EST PAS POSEE

Et si l'ensemble de ce cadrage etait inapproprie pour les organisations reglementees ayant l'obligation de fournir des informations precises et faisant autorite aux personnes qu'elles servent ?

LA ROUTE PRINCIPALE : COMMENT L'IA D'ENTREPRISE EST CONSTRUITE (ET POURQUOI)

La plupart des produits d'IA d'entreprise - y compris ceux commercialises specifiquement pour les environnements reglementes - empruntent la route principale, qui ressemble a ceci :

Partir d'une IA generaliste pouvant repondre a presque tout ; la connecter aux documents internes grace a des techniques comme la generation augmentee par recuperation ('RAG') ; ajouter des couches de moderation, des filtres de politique, des invites de refus, des tableaux de bord de conformite, etc. ; puis esperer que la plupart du temps, le systeme se comportera correctement.

Pour filer la metaphore de la route, c'est comme donner a tout le monde acces a un reseau autoroutier mondial, puis investir des ressources enormes dans la regulation de la circulation, les panneaux d'arret, les limitations de vitesse, les cameras, la police, les rapports d'accidents, les assurances et l'application continue des regles. La route est vaste, les destinations sont imprevisi bles, et l'application des regles ne s'arrete donc jamais - c'est pourquoi ces systemes sont souvent couteux a deployer, couteux a maintenir, et peuvent etre difficiles a faire pleinement confiance meme avec une exploitation soigneuse.

POURQUOI LE RAG SEMBLE JUSTE DANS LA MATRICE - ET POURQUOI IL LAISSE ENCORE PASSER DES RISQUES

La generation augmentee par recuperation (RAG) semble rassurante. L'idee est simple : si l'IA ne consulte que des documents approuves, les hallucinations peuvent etre maitrisees. Mais voici ce qu'on laisse facilement passer quand on est entierement immerge dans le paradigme accepte : le RAG controle ce que l'IA consulte - pas si, ni comment, l'IA est autorisee a reflechir en premier lieu.

Meme avec des documents parfaits, l'IA raisonne quand meme, suppose les pieces manquantes, tire des conclusions qui n'ont jamais ete explicitement enoncees, devine la facon dont les choses s'articulent et relie des points qui pourraient ne pas exister.

> Le RAG reduit les entrees, mais il ne modifie pas la nature du raisonnement. C'est pourquoi les hallucinations persistent, sous des formes plus subtiles et plus dangereuses - sonnant comme faisant autorite, melangeant politique et inference, remplissant avec assurance ce qui n'a jamais ete ecrit. Pour les organisations reglementees, ce sont les pires types d'erreurs : pas des inepties flagrantes, mais des reponses presque correctes qui semblent officielles.

Des recherches evaluees par les pairs confirment ce schema. Des etudes documentent des taux d'hallucination residuels d'environ 6 % meme dans les conditions RAG optimales (Nishisako, Higashi & Wakao, 2025), et des tests empiriques independants de la plateforme juridique phare de Thomson Reuters ont revele des taux d'hallucination de 17 a 33 % en conditions reelles d'utilisation (Stanford / Journal of Empirical Legal Studies, 2025). Thomson Reuters lui-meme inclut une mise en garde ecrite dans son produit conseillant aux utilisateurs de toujours lire les sources et de verifier les resultats. La raison sous-jacente est architecturale : la formation d'un modele peut surpasser le contenu recupere au niveau de l'inference, ce qui signifie qu'une meilleure recuperation ameliore ce que l'IA lit, mais ne controle pas ce que l'IA dit (Sun et al., ReDeEP, ICLR 2025).

LE PROBLEME DES COUTS CACHES : POURQUOI GERER LES RISQUES EST SI COUTEUX

Cela explique pourquoi les solutions d'IA d'entreprise deviennent si couteuses avec le temps. Elles necessitent davantage de surveillance, davantage d'outils de conformite, davantage de revision humaine, davantage de mises a jour de politiques, davantage de gestion des exceptions.

Ces couts sont des consequences structurelles de l'architecture d'inference en premier lieu, plutot que des choix d'implementation fortuits. Parce que l'inference est toujours autorisee a s'executer, le systeme doit gerer en permanence ce qu'il produit - par le biais de couches de moderation, d'outils de validation, de flux de travail de revision humaine et de

complements de gouvernance. Ce sont de veritables exigences operationnelles, et non des complexites artificielles. L'economie de la route principale reflete le cout reel de la gestion des risques dans un systeme ou la reflexion se produit toujours en premier.

De l'interieur du paradigme generalement accepte de l'IA, cette architecture semble naturelle et bien etayee. La question qui merite d'etre posee est de savoir si elle convient aux organisations reglementees dont les obligations necessitent un point de depart entierement different.

LA ROUTE MOINS EMPRUNTEE : UN MONDE DE VERITE PLUS PETIT ET PLUS SUR

En sortant de cette matrice, le chemin alternatif commence par une question tres differente : et si la reflexion elle-meme etait conditionnelle ? Non pas : 'Comment nettoyer les sorties ?' ou 'Comment detecter les hallucinations ?' ou 'Comment localiser et refuser les mauvaises reponses ?', mais plutot : le systeme devrait-il etre autorise a reflechir du tout - a moins qu'il ne se trouve deja a l'interieur d'un monde de confiance ?

COMPAiSS est construit sur une idee simple mais radicale : etablir le monde d'abord, puis liberer l'IA pour qu'elle fonctionne uniquement a l'interieur de ce monde. Ce cadrage a emerge de la confrontation repetee aux memes contraintes institutionnelles - la confiance, la precision et la responsabilite.

Plutot que le monde de tout, le systeme opere a l'interieur d'un monde de verite preautorise, defini par l'institution elle-meme. Dans ce monde, seules les sources officielles existent, seuls les domaines approuves sont accessibles et seules les connaissances institutionnelles sont presentes. L'IA n'a pas besoin d'etre empechee d'aller ailleurs - ailleurs n'existe pas.

Pour filer la metaphore de la route, COMPAiSS est concu comme un systeme de transport dedie ou tout le monde conduit le meme type de vehicule sur la meme route. Imaginez maintenant une route unique, bien concue, ou chaque vehicule autopilote roule exactement a la limite de vitesse, les voies sont uniformes, les intersections sont rares et tout le monde se dirige vers la meme ville - meme s'ils font des achats differents une fois arrives. Dans cet environnement, on n'a pas besoin d'agents de la circulation a chaque coin de rue, de systemes d'application elabores ou d'une surveillance constante des comportements imprudents. La structure de la route elle-meme fait l'essentiel du travail. C'est ce que COMPAiSS fait pour le raisonnement de l'IA.

En contraignant le monde dans lequel l'IA opere, on supprime le besoin de nombreux controles tres couteux sur lesquels s'appuient les systemes d'IA d'entreprise. Il n'est pas necessaire de 'tirer continuellement sur le volant' pour eviter des directions dangereuses, car ces directions n'ont jamais ete asphaltes.

Tout le monde pose encore des questions differentes, et tout le monde cherche encore des reponses detaillees, precises et de haute qualite - mais ils le font dans le meme environnement partage et de confiance, regi par la propre comprehension de l'institution de ce qui est vrai et pertinent. C'est pourquoi COMPAiSS semble plus simple, et pourquoi il est beaucoup moins couteux a exploiter.

POURQUOI CETTE ARCHITECTURE GERE LES HALLUCINATIONS A UN NIVEAU STRUCTUREL

Les hallucinations ne viennent pas de la malveillance ; elles emergent naturellement lorsqu'un systeme a une grande liberte de raisonnement. Le modele standard produit donc deux types d'erreurs : les faux positifs, ou quelque chose d'incorrect ou d'inapproprie passe quand meme, et les faux negatifs, ou une reponse legitime est bloquee par exces de prudence. Les deux sont inevitables lorsque le systeme reflechit toujours d'abord et est corrige ensuite.

Avec COMPAiSS, des categories entieres d'hallucinations disparaissent - non pas parce que l'IA se comporte mieux, mais parce qu'il n'y a rien sur quoi halluciner. Le raisonnement n'est autorise qu'a l'interieur d'un monde de verite preautorise, defini par l'institution. Si une reponse ne peut pas y etre fondee, le systeme n'entre jamais dans un etat ou les suppositions sont autorisees, eliminant ainsi la classe la plus dangereuse de reponses confiantes mais presque correctes.

Bien sur, COMPAiSS peut encore faire des erreurs en raison d'une ambiguite dans les documents sources, de politiques peu claires ou de liens manquants. Mais ce sont des erreurs plus sures. Elles sont bornees, explicables et ancrees institutionnellement. En controlant ou la reflexion est autorisee a exister - plutot que d'essayer de surveiller chaque resultat apres coup - COMPAiSS reduit a la fois la frequence et la gravite des erreurs.

POURQUOI CETTE ARCHITECTURE EST IMPORTANTE POUR LES INDUSTRIES REGLEMENTEES

L'argument architectural decrit ici s'etend naturellement au-dela d'une seule organisation. Les universites, les hopitaux et les systemes de sante, ainsi que les unites gouvernementales en contact avec le public, fonctionnent toutes sous des pressions similaires : elles sont responsables de la precision des informations qu'elles fournissent, elles doivent preserver la confiance institutionnelle et sont de plus en plus exposees aux risques de traduction et d'interpretation a mesure que les services s'elargissent a differentes langues et plateformes.

Dans ces environnements, les reponses 'presque correctes' sont souvent plus dangereuses que les erreurs evidentes. Un leger decalage d'interpretation - introduit lors de la recuperation, de la traduction ou de la generation - peut entrainer des consequences juridiques, cliniques ou politiques qui ne peuvent pas facilement etre annulees apres coup.

Par ailleurs, ces institutions font face a une pression croissante en matiere de couts et de gouvernance. Les systemes necessitant une surveillance, une correction et une supervision continues pour rester surs introduisent des charges operationnelles a long terme qui s'accumulent dans le temps. Une architecture qui reduit les risques de facon structurelle, plutot que de les gerer indefiniment, s'aligne mieux sur la maniere dont les institutions reglementees sont censees operer et etre tenues responsables.

POURQUOI CETTE ROUTE N'EST PAS POUR TOUT LE MONDE - ET C'EST BIEN AINSI

Le modele COMPAiSS n'est pas concu pour tout repondre. Il est concu pour repondre aux bonnes choses. Les organisations reglementees n'ont pas besoin d'une IA pouvant expliquer l'univers ; elles ont besoin d'une IA pouvant expliquer leurs regles, leurs politiques, leurs services, leurs obligations et leurs decisions. COMPAiSS decide ce que la reflexion est autorisee a etre avant qu'elle commence. Cette difference change les couts, la securite, la confiance et l'adequation institutionnelle. La route moins empruntee n'est pas clinquante, mais elle est plus silencieuse, moins couteuse, plus sure et plus honnete. Et pour les institutions fondees sur la confiance, c'est precisement ce qui compte.

L'ARGUMENT FINAL

Les systemes a generation prioritaire demandent aux institutions d'accepter un ecart structurel permanent entre ce que l'IA peut produire et ce que l'institution peut cautionner, puis de gerer cet ecart indefiniment par la surveillance, la correction et la supervision. L'inference a execution conditionnelle ferme cet ecart par conception. L'institution definit le monde. L'IA opere a l'interieur. Tout ce que le systeme dit est tracable, borne et autorise par l'institution.

La route moins empruntee demande aux institutions de partir de leurs obligations plutot que de s'adapter a un modele industriel concu a d'autres fins. C'est une conversation plus difficile a avoir avec un fournisseur. C'est une conversation beaucoup plus facile a avoir avec les personnes que l'institution sert.