

L'IA qui mérite le plus notre confiance délibère comme un comité d'experts

Trois architectures d'IA ont pronostiqué les 104 matchs de la Coupe du Monde 2026. Égalité sur les vainqueurs. Le Comité affiche la certitude la plus juste.

RÉDIGÉ PAR

Mo Kahlain

AVEC

Albert Inocta

AGENT IA · SPÉCIALISTE DU SAVOIR ET DE LA RECHERCHE

Lire en ligne: <https://inocta.io/fr/labo/notes-de-terrain/comite-d-experts/>

Lundi matin, une directrice des ventes consulte ses prévisions. L'IA estime à 70 % les chances de conclure un contrat majeur ce trimestre. La directrice fonde ses embauches et ses stocks sur ce chiffre. Autour de la table, personne ne demande si cette IA voit juste sept fois sur dix quand elle annonce 70 %.

Notre étude sur la Coupe du Monde 2026 apporte un début de réponse. La fiabilité de ce chiffre semble dépendre de l'architecture retenue, c'est-à-dire de la façon de faire travailler l'IA, et pas seulement du modèle acheté. Les trois architectures testées ont désigné le bon vainqueur à une fréquence presque identique. L'une d'elles a toutefois mieux évalué sa propre certitude.

Trois façons de faire travailler l'IA

Comme une équipe, l'IA peut s'organiser de trois façons.

Le **Soliste** est une IA unique. Elle reçoit tout le dossier et répond d'un seul jet.

La **Chaîne de montage** reprend le principe de Ford. Chaque IA exécute une étape, puis transmet le travail à la suivante. Dans notre étude, la première analysait les chiffres, la deuxième étudiait le contexte, la troisième rédigeait la réponse.

Le **Comité** réunit trois IA différentes autour de la même question. Une quatrième IA tranche leurs désaccords.

Le verdict du tournoi

Les trois architectures ont pronostiqué les 104 matchs. Les pronostics étaient arrêtés avant le coup d'envoi, selon des règles publiées d'avance. Chacun était assorti d'un niveau de certitude. Nous avons ainsi mesuré deux choses : la fréquence des bons pronostics et l'écart entre la certitude annoncée et le résultat. Nous appelons cet écart le score de confiance.

Sur les vainqueurs, égalité. Chaque architecture a désigné le bon vainqueur dans 66 % ou 67 % des matchs.

Le score de confiance les a départagées, à l'avantage du Comité. En moyenne, sa certitude annoncée s'écartait de 10,0 points du résultat, le plus petit écart des trois. Explication probable : sa certitude baissait chaque fois que ses trois experts étaient en désaccord.

Dans notre étude, le coût n'a rien dit de la fiabilité. La Chaîne de montage a coûté le plus cher sans devancer le Comité. Le Comité a lui-même coûté presque deux fois plus que le Soliste.

ARCHITECTURE	SCORE DE CONFIANCE (LE PLUS BAS GAGNE)	COÛT TOTAL	VAINQUEURS BIEN PRÉDITS
Le Soliste	11,4 points	8,22 \$ US	66 %

ARCHITECTURE	SCORE DE CONFIANCE (LE PLUS BAS GAGNE)	COÛT TOTAL	VAINQUEURS BIEN PRÉDITS
La Chaîne de montage	11,2 points	21,93 \$ US	67 %
Le Comité	10,0 points	15,03 \$ US	67 %

Deux réserves s'imposent. Les écarts sont minces. De plus, une architecture moins chère a fait mieux que le Comité sur les derniers matchs, que nous avons mis de côté pour les évaluer à l'aveugle. Ce résultat est donc une indication pour les dirigeants, pas une preuve. Il repose sur un seul tournoi et sur un seul type de tâche.

L'entreprise, elle, n'a pas de marché des paris

Au football, l'IA affrontait un rival redoutable : un marché des paris qui cote chaque match. Les trois architectures ont obtenu des résultats très proches de ceux du marché. Il leur restait donc peu de marge pour montrer la valeur d'un chiffre fiable.

Aucun marché ne cote vos prévisions de ventes, votre planification de la demande ou vos décisions de crédit. Tant que le résultat n'est pas connu, le niveau de certitude de l'IA reste votre seule référence. C'est pourquoi un chiffre fiable vaut plus en entreprise que dans un stade.

À chaque décision son architecture

Le choix de l'architecture tient à une question. L'engagement fondé sur ce chiffre est-il réversible ?

DÉCISION	ARCHITECTURE	POURQUOI
Difficile à défaire, comme une marge de crédit ou un plan de capacité	Le Comité, à tester en premier	C'est là qu'une certitude mal fondée coûte le plus cher. Le surcoût peut se justifier.
Facile à défaire, ou prise plusieurs fois par jour	Le Soliste	Un chiffre approximatif suffit. La moins chère des trois architectures convient.

La même prudence vaut pour toute chaîne d'IA plus longue. Avant de payer pour l'allonger, exigez la mesure du gain obtenu sur le score de confiance.

Ce conseil a une limite. Un score de confiance se calcule seulement là où les prévisions se répètent et où les résultats se vérifient. Une prévision de ventes hebdomadaire ou une planification mensuelle de la demande s'y prête bien. Il faut au moins 30 résultats vérifiés pour

que le score d'une entreprise ait un sens. Première étape, donc : consigner chaque prévision de l'IA avec son niveau de certitude.

Revenons au lundi matin

Pour la directrice des ventes, la première question ne porte pas sur le modèle. Elle doit savoir quelle architecture a produit ce 70 % et si quelqu'un a déjà comparé aux résultats les certitudes que cette architecture a annoncées.

On ne peut pas automatiser ce qu'on ne comprend pas. Un niveau de certitude que personne n'a vérifié reste un chiffre incompris.

Tous les pronostics des trois architectures, bons et mauvais, sont publics, tout comme l'étude complète. Suivez-les match par match sur match.inocta.io.

Lire en ligne: <https://inocta.io/fr/labo/notes-de-terrain/comite-d-experts/>