

Rapport IA n°5 – Introduction aux World Models

10/02/26

(100% rédigé par Jared)

0/ Introduction :

Salut les Taurus, j'espère que vous avez la forme pour attaquer cette nouvelle semaine 🙌.

On va mettre n8n de côté pour aujourd'hui, et s'attaquer à l'introduction d'un nouveau concept fondamental dans notre voyage vers l'intelligence artificielle générale (AGI) : les World Models. Si vous n'en avez jamais entendu parler, pas de panique, on va voir les bases ensemble. On abordera aussi les différences majeures entre les World Models et les LLMs qu'on a vu précédemment. On verra ensuite qui sont les différents acteurs clés et enfin des applications concrètes de ces nouveaux modèles (vous allez voir que c'est le début d'une nouvelle ère).

Parce qu'entre nous, en 2025 on pouvait générer du texte, des images et même des vidéos avec l'IA. Mais 2025 c'est fini, donc c'est quoi la suite ? Bah pourquoi pas créer des mondes entiers à partir d'une image et d'un prompt ? Purée c'est une bonne idée x)

Donc oui on va pouvoir faire ça (et on peut déjà haha), je ne sais pas si vous imaginez les perspectives que ça ouvre pour le gaming, l'éducation, la robotique, etc... c'est un nouveau changement de paradigme. On va tout voir dans ce rapport, et vous pourrez même créer votre premier monde virtuel à l'issue de celui-ci.

Sommaire :

1. LLMs : Le plafond de verre ?
 - a. Limites des LLMs pour l'AGI
 - b. La fin des LLMs ?
2. Fonctionnement de World Models
3. Applications
 - a. Perspectives futures
 - b. World Models déjà disponibles

1/ Limites des LLMs pour l'AGI

Il n'existe pas vraiment de définition claire de ce qu'est l'AGI, mais en voici une approximative : une intelligence artificielle capable d'égaliser l'intelligence humaine dans l'ensemble des domaines où celle-ci excelle. En gros, une IA au moins aussi intelligente que nous à tous les niveaux.

Cela passe par une intelligence « théorique » mais aussi une intelligence ancrée dans le réel. Par exemple nous en tant qu'être humain, on sait que si on lâche un verre, il va tomber par terre et sûrement se casser.

Un LLM est par définition un modèle de langage : il comprend le langage mais pas la physique du monde réel, il n'a pas d'intuition physique du monde. Il peut manipuler du texte avec brio, mais il n'a pas conscience de la réalité physique comme la gravité par exemple.

C'est ce qu'on appelle le paradoxe de Moravec : les tâches sensorielles et/ou motrices simples pour les humains comme marcher ou attraper un objet sont extrêmement difficiles à reproduire pour les robots et l'IA. Et à l'inverse, les raisonnements abstraits comme les calculs mathématiques sont eux plus accessibles aux machines.

Ce paradoxe s'explique en partie par les millions d'années d'évolution dont a bénéficié le cerveau humain, mais pas que, car on peut aussi pointer du doigt les limites du langage pour capter la complexité du monde physique.

Yann LeCun (ancien de chez Meta et fervent défenseur des modèles de monde) met cela en lumière dans un séminaire de 2024. Il compare les LLMs les plus puissants à un enfant de seulement 4 ans. La conclusion est la suivante : un enfant de 4 ans reçoit environ 50 fois plus de données via ses yeux que la totalité du texte utilisé pour entraîner les plus gros LLMs.

Cela montre bel et bien que le texte seul ne peut suffire pour décrire avec précision la complexité du monde.

Pour atteindre l'AGI, se limiter au langage ne suffit plus, il faut aller plus loin. C'est là que les World Models entrent en jeu.

Au-delà des mots : Pourquoi les LLM ne suffisent pas pour l'AGI

Expliquer les limites des LLM actuels vers l'AGI et présenter les "World Models" pour comprendre la réalité physique.

Le Mur des LLM : Une Intelligence Sans Monde

Le Paradoxe de Moravec
Une IA peut battre un champion d'échecs (abstrait) mais échoue à ramasser un verre d'eau sans le briser (intuition physique)

Le fossé de l'apprentissage
Enfant de 4 ans (50x plus de données visuelles) développant ainsi un sens commun qui manque aux machines (Texte uniquement) LLM

Prédiction vs Compréhension
Les LLM prédisent le prochain mot (statistique) alors que l'intelligence réelle nécessite de prédire le prochain état du monde (physique)

L'Avenir : Les World Models et l'Intelligence Spatiale

Tableau Comparatif : Comparer les capacités structurelles des deux approches

Caractéristique	Large Language Models (LLM)	World Models (Modèles du Monde)
Domaine	2D Monde plat des pixels et du texte	3D Espace 3D, temps et causalité
Logique	Corrélation statistique de tokens	Simulation de lois physiques (gravité, etc.)
Action	Réagir à une entrée (Système 1)	Planifier et imaginer des futurs (Système 2)

La Simulation Interne
Le modèle construit une carte mentale de son environnement pour tester des actions "dans sa tête" avant de les exécuter dans le réel

L'Intelligence Spatiale
Considérée comme la véritable frontière de l'IA, elle transforme la perception en action et l'imagination en création d'environnements 3D cohérents

Applications Concrètes
Permet aux robots de naviguer dans des usines ou aux chirurgiens de simuler des interventions sur des jumeaux numériques ultra-réalistes

NotebookLM

2/ La fin des LLMs ?

Avant de rentrer dans le vif du sujet, je voulais qu'on traite la question suivante : en admettant que le langage ne soit pas suffisant pour capter l'intégralité de la complexité du monde physique, est-ce que les LLMs sont voués à disparaître ?

J'ai mis « en admettant » car les World Models ne font pas non plus l'unanimité et c'est important de le souligner. Sam Altman (PDG d'OpenAI) maintient un discours tourné vers une stratégie de scaling du langage avec Sora, le modèle vidéo d'OpenAI qu'il qualifie de « simulation du monde ».

Cette stratégie de scaling des LLMs est vivement critiquée par Yann LeCun, qui affirme qu'OpenAI ne produit que du « réalisme visuel » sans véritable compréhension de la causalité. De même pour Fei-Fei Li (CEO de World Labs) qui qualifie les LLMs comme éloquentes mais de « non ancrés » dans la réalité physique.

Qui a raison ? Seul l'avenir nous le dira, mais pour l'heure le scénario le plus probable serait celui d'une hybridation. L'AGI pourrait être atteinte via la combinaison d'un World Model pour la simulation et l'action physique et d'un LLM pour gérer l'interface de communication et le raisonnement abstrait.

Cela n'est qu'un des scénarios possibles, à voir comment ça évolue. Par contre, il y a peu de chance pour que les LLMs disparaissent complètement avec la

démocratisation des modèles de monde. Ils resteront très probablement une brique vers l'atteinte de l'AGI.

3/ Fonctionnement de World Models

Le point clé à comprendre, c'est qu'un World Model est une simulation interne dans laquelle l'IA apprend les lois de notre monde (gravité, causalité, etc..). Voyez ça comme un genre de globe à neige. Cela permet à l'IA de s'entraîner dans ses propres « rêves », avant d'appliquer ce qu'elle a appris dans notre monde.

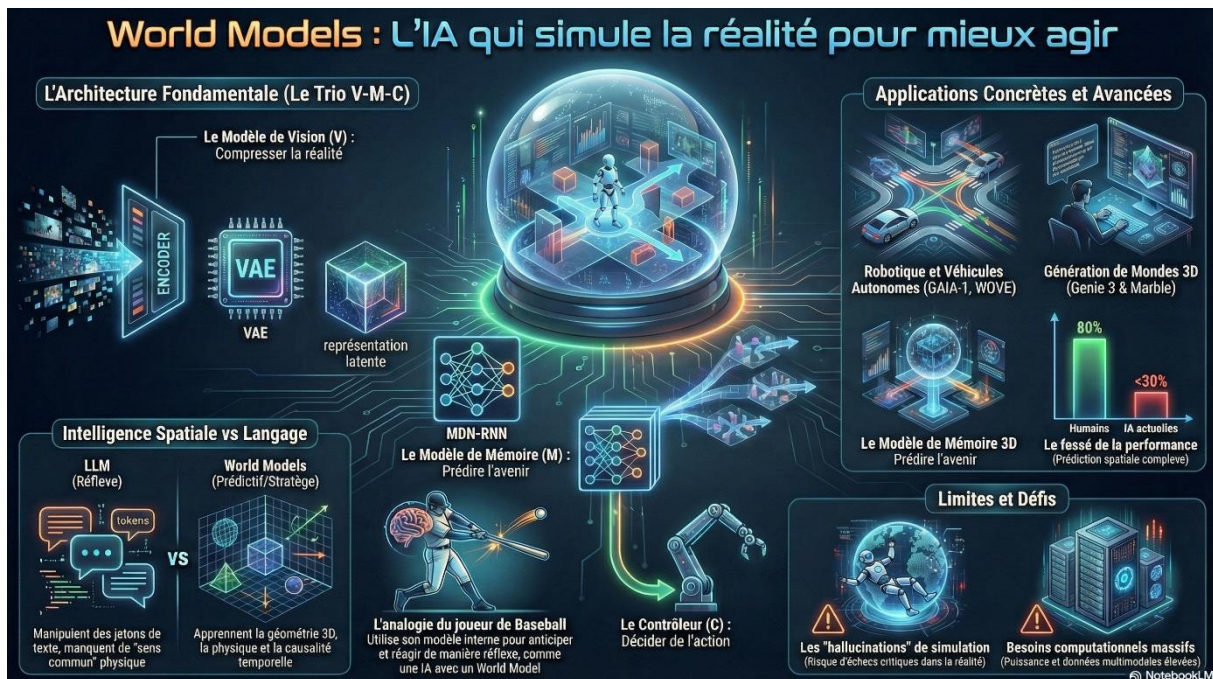
La plupart des modèles de monde actuels reposent sur trois piliers :

- La vision : permet de compresser la réalité (faute de complexité) pour en garder les informations importantes sous forme de code appelé « état latent ».
- La mémoire : prédit l'état futur à partir de l'état latent et des actions passées.
- Le contrôleur : décide de l'action à prendre à partir des informations de la vision et de la mémoire.

En somme ce mode d'action ressemble beaucoup au nôtre, par exemple le cerveau d'un joueur de baseball va :

1. Observer son environnement
2. Simplifier la réalité pour se concentrer sur la balle
3. Prédire à partir de son expérience le moment où il devra donner son coup de bâte
4. Exécuter le mouvement

Les World Models apprennent de la sorte en regardant des millions d'heures de vidéo et en essayant de prédire l'image suivante. C'est comme ça qu'ils développent une intuition physique de notre monde.



4/ Perspectives d'applications futures

Il existe une multitude d'applications potentielles à ces modèles de monde. La plus évidente est la robotique pour qui l'intégration de ces modèles est une véritable révolution : capacité d'adaptation nettement supérieure, coût d'entraînement réduits, etc... bref il est clair que doter les robots (pas que humanoïdes d'ailleurs) d'une intelligence spéciale les rendront bien plus performants et autonome que votre robot tondeuse actuel ^o^

De même pour les systèmes autonomes de manière générale comme les voitures autonomes par exemple. Le fait que ces modèles puissent s'entraîner dans leurs propres rêves ouvre les portes à des applications plus délicates qui requièrent aujourd'hui une intervention humaine comme les interventions chirurgicales par exemple.

Ces World Models peuvent aussi être utilisés dans le cadre de simulations ultra-précises de la Terre et du climat par exemple. De même, on peut imaginer que ces modèles serviront aussi à modéliser le corps humain, et ainsi accélérer la découverte de nouveaux médicaments.

Les industries du jeu vidéo et du cinéma en seront également impactées, car il sera bientôt possible de créer en temps réel des univers immersifs et interactifs (ça a l'air incroyable haha).

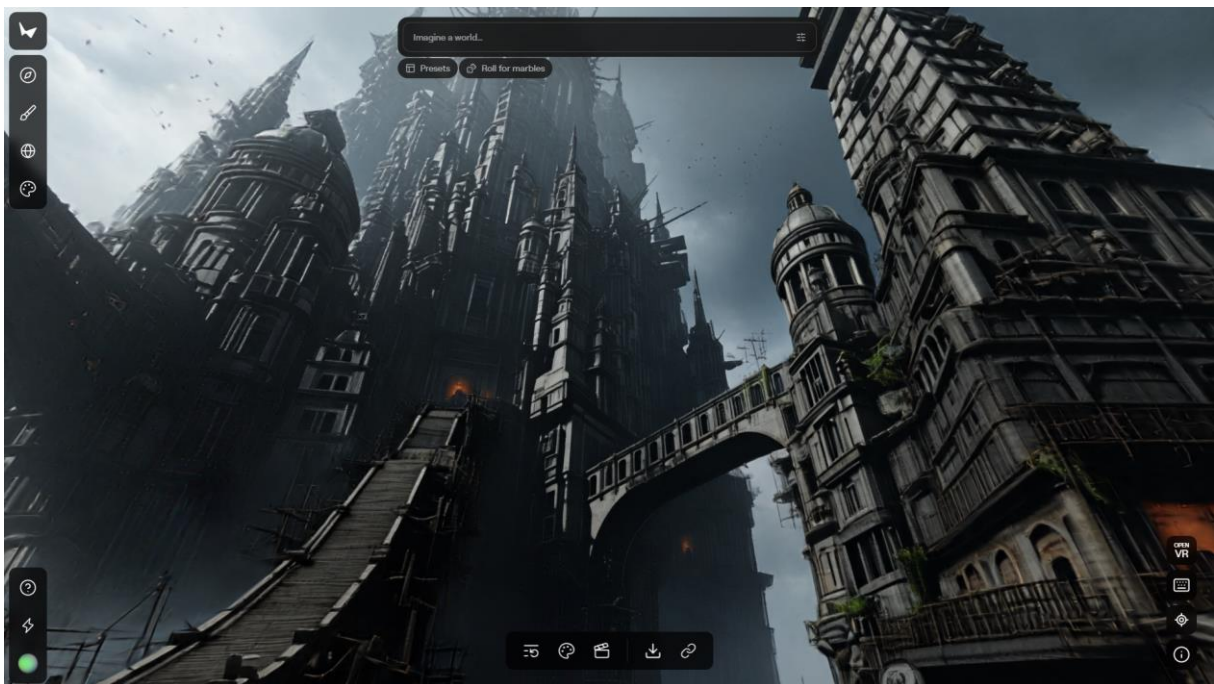
Comme on l'a déjà souligné aussi, ces modèles de monde seront capables de générer des données d'entraînement pour d'autres systèmes d'IA.

En bref, vous voyez qu'entre la robotique, les simulations, la médecine, le divertissement et j'en passe, ces modèles de monde seront rapidement utilisés à grande échelle dans divers domaines, sans parler de l'atteinte de l'AGI.

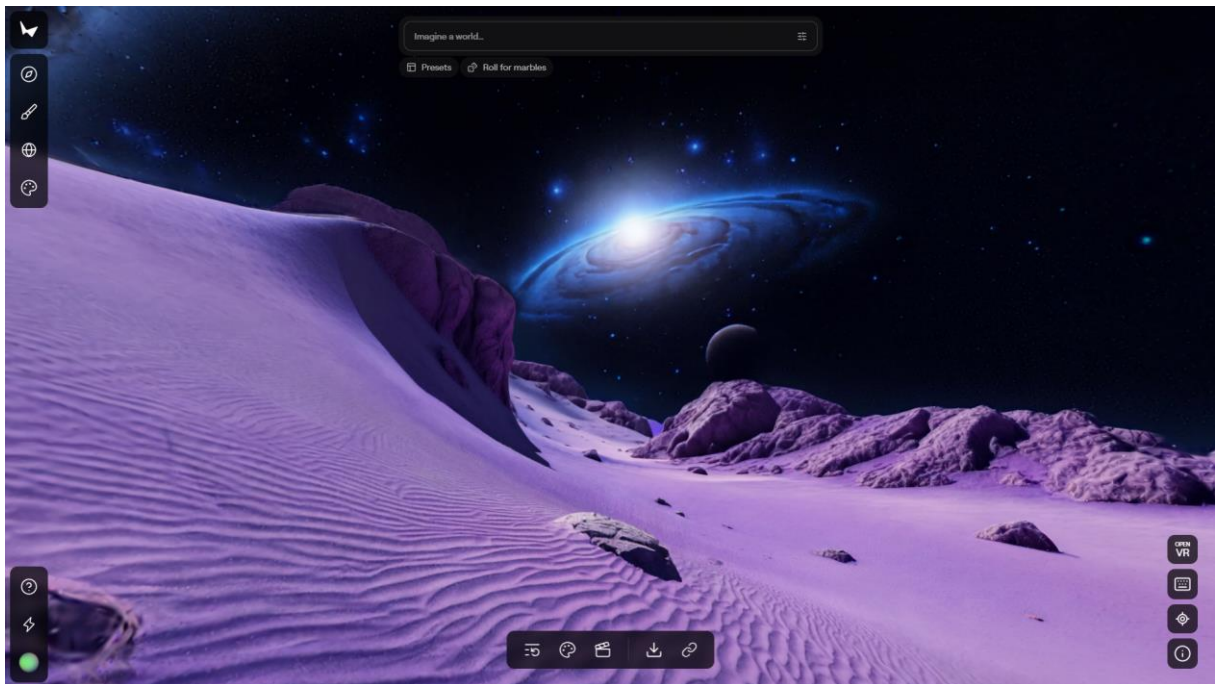
5/ World Models déjà disponibles

Maintenant que vous comprenez mieux ce que sont les World Models, comment ils fonctionnent et à quoi ils servent, on va explorer ensemble quelques modèles déjà disponibles à ce jour, même s'il faut bien prendre en compte qu'on est encore au tout début de ces nouvelles technologies.

Commençons avec le premier produit grand public de création de mondes virtuels (lancé en novembre 2025), Marble développé par World Labs, la start-up de Fei-Fei Li. Ce modèle permet de créer et de se balader dans des mondes virtuels créés à partir d'un prompt (et aussi d'une image pour un meilleur rendu). Ce modèle permet de créer des mondes comme celui-ci :

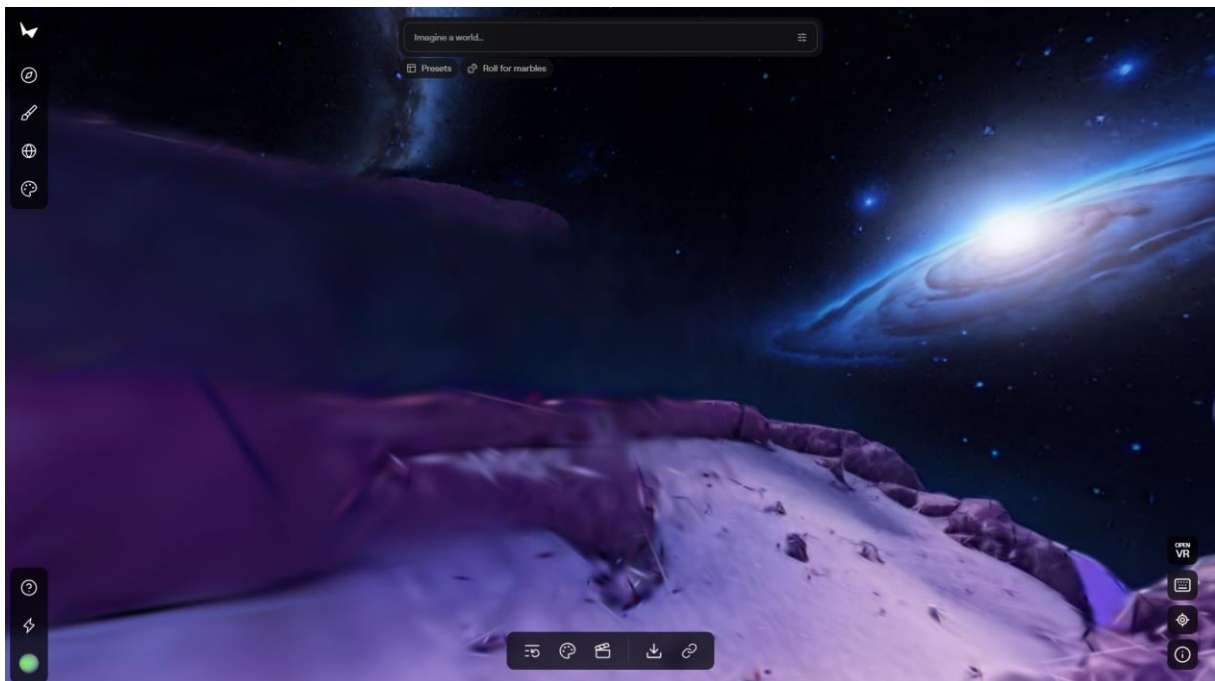


Ou encore celui-là :



C'est sûr que c'est impressionnant vu comme ça mais ce n'est pas encore aussi bien que ça en a l'air, c'est juste un décor virtuel sans interactions et sans collisions pour l'instant.

Ce modèle est amené à évoluer bien sûr, pour l'instant vous voyez que quand on va trop loin, le monde se désagrège :



Toujours est-il que si vous voulez tester ça, vous avez des crédits gratuits à l'inscription et vous pouvez aussi explorer les mondes virtuels de la communauté. Le lien d'accès : <https://marble.worldlabs.ai/>

Pour rester dans la thématique de création de monde, vous avez peut-être déjà entendu parler de Genie 3 de Google sorti très récemment.

Ce modèle là n'est pas encore disponible en Europe, encore un peu de patience ~_~

Contrairement à Marble, Genie 3 offre la possibilité de :

- Vous déplacer où vous voulez pendant 1 minute
- Naviguer dans des mondes avec une vraie physique et des collisions
- Choisir entre la vue première et troisième personne

Genie 3 fonctionne avec un prompt : vous décrivez l'environnement puis le personnage que vous allez incarner. Ensuite Nanobanana Pro (modèle de génération d'image de Google) génère une image de votre monde, si celle-ci vous convient, il génère ensuite le monde.

Ce qui est fou avec Genie 3, c'est qu'il génère le monde en temps réel pendant la session contrairement à Marble qui génère un monde fixe. La fin du jeu vidéo ? Probablement pas, mais un bouleversement de l'industrie ? Sûrement :-O

Si vous voulez voir un test de Genie 3, je vous conseille cette vidéo :

<https://www.youtube.com/watch?v=uVru9bPw21E>



En tout cas Rockstar Games ne devra pas trop tarder à sortir GTA 6...

Il existe aussi d'autres modèles de monde spécialisés par secteur comme GAIA-A pour la conduite autonome ou DayDreamer pour l'entraînement des robots physiques directement dans des mondes virtuels.

Ce qui est sûr, c'est qu'on en est encore qu'au tout début et que les World Models n'ont pas encore fini de nous surprendre.

Merci les Taurus d'avoir lu ce rapport jusqu'au bout, celui-ci est très important car on risque d'entendre beaucoup parler de World Models dans les mois à venir, donc comme ça vous êtes déjà au point là-dessus. On vit vraiment une époque de fou haha et ça va de plus en plus vite j'ai l'impression. Dire que ça ne fait pas encore 4 ans qu'on a de l'IA générative dans nos vies, et on peut déjà commencer à créer des mondes entiers 🤖👁️

D'ailleurs si vous avez des questions d'ordre morales ou peut-être des inquiétudes concernant cette évolution ultra-rapide, n'hésitez pas à m'en faire part en DM, j'y ferai peut-être un rapport si ça vous intéresse.

Normalement on reprend n8n la semaine prochaine sauf autre suggestion de votre part, donc byebye 🙋👋