

Évolutions récentes de l'IA (génération)

MLA – ITI 5

Introduction & Contexte

- Définitions : IA générative, LLM, modèles de diffusion
- Historique : Transformers (2017), GPT-2, GPT-3, ChatGPT (2022)
- Importance académique, industrielle et open source

Écosystème industriel

- Startups : Mistral, HuggingFace, Perplexity, Cohere
- Big Tech : Google Gemini, Microsoft Copilot, Meta LLaMA, Alibaba Qwen
- Rapport de force : modèles open source vs propriétaires
- Analyse des gagnants et perdants

Percées académiques et industrielles

- Chain-of-Thought (CoT) – Wei et al. (2022)
- RLHF & alignement – Bai et al. (2022)
- RLVR (2024), DeepSeek R1 (2025)
- Protocole MCP (Anthropic, 2024)
- Benchmarks : ARC-AGI, REST, etc.

Nouvelles capacités – Raisonnement

- Techniques : CoT, ReAct, Tree of Thoughts, Test-time reasoning
- DeepSeek R1 : mixture of experts, raisonnement par renforcement
- Vers des IA plus efficaces et locales

Nouvelles capacités – Agents

- Agentic AI : IA capables d'agir et d'interagir
- Exemples : Manus, Lovable, Replit, Cursor, ChatGPT Agent
- Applications : génération de code, exécution autonome de tâches
- Défis : hallucinations, sécurité, robustesse

IA et Robotique

- Embodied AI : IA incarnées dans des robots
- Exemples : Waymo (projet EMMA), Nvidia Jetson Thor
- World models : apprentissage d'un modèle du monde physique
- Applications : manipulation, navigation, robotique humanoïde

LLM locaux et Edge AI

- Petits modèles (SLM) : moins de 20B paramètres
- Exemples : Mistral Small, Google Edge AI Gallery, LLMStudio
- Avantages : faible coût, fonctionnement hors-ligne, personnalisation
- Cas d'usage : smartphones, montres, lunettes, assistants privés

Matériel et infrastructures

- AI GigaFactories : entraînement à grande échelle
- GPU, TPU, NPU : Nvidia GH200, AMD MI300A, Apple M4
- Défis : consommation, mémoire, interconnexions
- Co-conception hardware/software pour l'efficacité énergétique

Sécurité, Alignement et Régulation

- Risques : hallucinations, comportements émergents
- Exemples : Claude Opus 4, GPT-4o
- Méthodes : CoT monitoring, alignment hiérarchique, Scientist AI
- Géopolitique : US, Chine, Europe régulatrice

Perspectives et Frontières

- Vers l'AGI : limites structurelles des Transformers
- Nouveaux paradigmes : diffusion discrète, modèles autorégressifs alternatifs
- Personnalisation : modèles de langage personnalisés (PLLMs)
- Questions ouvertes : raisonnement, action, mémoire, sécurité

Conclusion

- L'IA générative évolue rapidement sur 3 axes : académique, industriel, open source
- Nouvelles capacités : raisonnement, agentic AI, robotique, personnalisation
- Défis : matériel, sécurité, régulation, souveraineté numérique
- Impacts pour les ingénieurs : outils, recherche, métiers à venir