

Vers une architecture IA globale et adaptative

Pour faire face à des situations variées et imprévisibles, une IA doit être conçue comme un **système polyvalent** capable d'intégrer plusieurs modalités de données, de s'adapter à de nouveaux domaines et de combiner diverses stratégies. Les travaux récents soulignent en effet que les progrès vers une IA générale (AGI) passent par des modèles **multimodaux** unifiés et modulaires ¹ ². Un système idéal combinerait ainsi apprentissage statistique, raisonnement symbolique et composants spécialisés pour exploiter « toutes les options possibles » selon le contexte. Voici un cadre théorique (template conceptuel) détaillant les principes et composants clés d'une telle IA globale :

- **Contexte multi-domaine.** À l'instar d'un AGI, le système doit pouvoir être **exposé à de nouveaux domaines inconnus** et apprendre à y raisonner ³ ⁴. Le théorème du *no-free-lunch* affirme qu'« aucun algorithme unique n'excellera sur tous les problèmes » ⁴ : il faut donc prévoir une palette d'approches (réseaux profonds, systèmes experts, règles symboliques, optimisations). Par exemple, un mécanisme hybride peut combiner les points forts d'un système expert (symbolique) avec ceux des méthodes d'apprentissage (neuronales ou génératives) pour créer et affiner automatiquement des heuristiques dans un nouveau domaine ⁵.
- **Multi-modalité et unification.** Le modèle doit traiter de façon intégrée **plusieurs types de données** (texte, image, audio, capteurs, etc.). De grands « modèles de base » actuels sont souvent entraînés sur des données massives et diverses, en vue d'apprendre des représentations **générales et réutilisables** ¹ ⁶. Idéalement, les entrées et sorties seraient **unifiées** dans un même espace de représentation : c'est le principe d'**unifiabilité** qui consiste à encoder les données hétérogènes dans une représentation commune pour modéliser leurs interactions ². Cette approche permet d'encourager un meilleur apprentissage collaboratif entre diverses tâches et modalités » ². On pense notamment aux réseaux de transformateurs universels qui prennent comme entrées des séquences textuelles représentant toute information (image convertie en tokens, signaux de capteurs, etc.), facilitant l'apprentissage multi-tâche et le transfert de connaissances entre domaines.
- **Modularité et flexibilité.** L'architecture doit être **modulaire**, c'est-à-dire composée de sous-systèmes spécialisés (encodeurs, décodeurs, modules de traitement par modalité, etc.) que l'on peut assembler ou remplacer de façon « plug-and-play ». La modularité offre « plus de flexibilité et d'évolutivité » : chaque module peut être conçu, entraîné et testé de façon indépendante, puis intégré au système global ⁷. Par exemple, on peut avoir des encodeurs spécifiques pour la vision, le langage, le signal, raccordés à un module central (par exemple un grand modèle neuronale multi-tâche) via des adaptateurs. Cette conception facilite la maintenance et la mise à jour (on peut mettre à jour un encodeur ou un module sans re-entraîner tout le système), et permet d'analyser indépendamment chaque composant pour comprendre leur contribution au comportement global ⁷. En pratique, on pourra combiner différentes configurations d'entraînement (certains modules gelés, d'autres entraînés) et utiliser des architectures de type **RAG** (Retrieval-Augmented Generation) pour intégrer des sources de connaissances externes de manière modulaire ⁷.

- **Adaptabilité et apprentissage continu.** L'IA doit s'ajuster en ligne aux nouveaux contextes sans nécessiter un réentraînement complet. Cela passe par des objectifs d'entraînement larges (éviter la surspécialisation) et des techniques comme le *prompt tuning*, le *fine-tuning* sur des exemples spécifiques, ou l'expansion du modèle. De fait, grossir un modèle (plus de paramètres) améliore implicitement son adaptabilité car il apprend à représenter des données plus variées ⁸. Au-delà de l'échelle, on cherche à renforcer l'**adaptabilité** du système : c'est « une caractéristique fondamentale » qui permet un meilleur transfert de connaissances entre tâches et modalités, et ainsi d'aborder de nouvelles tâches sans réentraînements lourds ⁹. Par exemple, des méthodes d'apprentissage par renforcement, de méta-apprentissage ou de travail en mode *continual learning* peuvent être intégrées pour que l'IA affine ses performances en situation réelle.
- **Boucle de perception-action.** Le template devrait inclure un cycle *percevoir - décider - agir* alimenté par les retours (feedback) environnementaux. Autrement dit, on imagine un **orchestrateur adaptatif** qui supervise la collecte de données multi-sensorielles (caméras, microphones, capteurs divers), la fusion de ces données via les modules d'analyse, puis la prise de décision par des modules de planification/prise de décision (par exemple, planificateur symbolique, agent RL, etc.). Les décisions peuvent alors déclencher des actions ou des requêtes (ex. nouveau prompt, recherche externe). Ce cycle est itératif : chaque action produisant un nouvel état, l'IA met constamment à jour son contexte, ajuste ses modules (entraînements incrémentaux) et réévalue les probabilités de ses choix. Ce schéma rappelle les architectures des systèmes autonomes (ex. robotique avec SLAM visuel, ou agents conversationnels qui affichent du feedback), mais à l'échelle d'une IA générale chaque composant reste interchangeable selon les besoins.
- **Contextes et situations multiples.** Le cadre théorique doit prévoir de gérer plusieurs niveaux de contexte : **immédiat** (contexte conversationnel ou sensoriel actuel), **historique** (données accumulées, connaissances acquises), et **stratégique** (objectifs à long terme, contraintes éthiques/réglementaires). Par exemple, l'IA pourrait maintenir un *graphe de connaissances* dynamiquement mis à jour pour stocker les entités, relations et expériences passées, servant de base à la prise de décision dans divers scénarios. Elle doit également considérer des **objectifs multiples** : par exemple, combiner performance technique, sécurité et alignement éthique (scénario typique en IA critique). Aucun système unique n'excellerait dans tous ces aspects (surtout en cas de conflits d'objectifs), d'où l'intérêt d'une architecture multi-agents ou multi-objectifs. Cette approche rappelle la notion de *fédération* où plusieurs sous-modèles (agents) spécialisés négocient ou arbitrent pour aboutir à une décision globale.
- **Incorporation d'IA générative et de connaissances.** Les modèles modernes à base de transformateurs (LLMs, diffusion, etc.) fournissent un pouvoir de généralisation très élevé ¹⁰ ⁶. Le template peut utiliser ces modèles de base pour la compréhension du langage, la génération de plans ou d'hypothèses, etc. On peut exploiter l'**apprentissage par instruction** (instruction tuning) et l'**apprentissage par contexte** (prompting) pour guider ces modèles sans réentraînement complet ¹¹ ⁶. Par ailleurs, la capacité à intégrer de nouvelles connaissances (via indexation de données, RAG, mémoires externes) est cruciale pour que l'IA puisse élargir son champ de connaissance après déploiement.
- **Contrôle et supervision humaine.** Enfin, pour la sécurité et la fiabilité, le cadre théorique peut intégrer des mécanismes de contrôle humain (*human-in-the-loop*). Par exemple, un module d'**orchestration** pourrait solliciter l'intervention d'un opérateur lorsque les décisions sont incertaines

ou critiques. Cette supervision permet de combiner l'autonomie de l'IA avec le jugement humain en contexte complexe.

Template conceptuel proposé

En synthèse, un template théorique pour une IA universelle multi-contexte pourrait inclure les éléments suivants :

1. **Capteurs et perception** (multimodalité) : Réseau de capteurs variés (caméras, microphones, IoT, interfaces logicielles). Chaque type de donnée est traité par un **module d'encodage dédié** (vision, langage, son, signaux, etc.).
2. **Fusion et représentation unifiée** : Un composant central unifie les données encodées en une **représentation commune** (par exemple par un modèle transformeur universel) 2 6. Cela peut se faire en sérialisant chaque modalité sous forme de tokens compatibles.
3. **Modules spécialisés** : Divers sous-systèmes (ex. analyseur d'intention, détecteur d'objets, analyseur de sentiments, calcul de physique interne, etc.) qui consomment la représentation unifiée ou les sorties d'autres modules. Chaque module peut être activé en fonction du contexte/tâche.
4. **Orchestrator adaptatif** : Cœur du système qui gère le pipeline de données, active/désactive les modules selon les besoins, et combine leurs résultats. Il inclut la logique décisionnelle (planification symbolique, agent RL, réseaux de neurones métacognitifs, etc.) et peut apprendre à **réorienter les requêtes internes** (Meta-Controller) pour optimiser les performances globales.
5. **Cycle apprentissage/validation continue** : Boucle qui utilise le retour d'information (feedback, récompenses, nouvelles données) pour mettre à jour les modules (par fine-tuning ou reconfiguration adaptive). On y intègre des mécanismes d'apprentissage continu ou par renforcement, afin que le système s'améliore en opération 8 9.
6. **Mémoires et bases de connaissances dynamiques** : Structure mémoire (types base de connaissances, graphes de connaissances, mémoires à long terme) pour stocker les expériences, contextes et solutions rencontrés. Ces mémoires alimentent l'orchestrator et les modules avec un contexte élargi.
7. **Diversité d'algorithmes et de paradigmes** : Composants utilisant différentes approches (algorithmes génétiques, optimisations, logiques, apprentissage statistique profond, etc.) pour couvrir un éventail de scénarios. Le système peut choisir selon les situations l'algorithme le plus adéquat, conformément au *no free lunch* 4.
8. **Interface de supervision** : Mécanismes de gestion des risques (audit, logs de décisions) et options d'intervention humaine quand nécessaire.

Ce schéma n'est pas un produit fini, mais un **cadre flexible** : on peut y ajouter ou retirer des modules, reconfigurer l'architecture (par ex. passer d'une configuration unifiée à une configuration fortement modulaire) selon le cas d'usage. L'idée est de concevoir une plate-forme IA « multi-tout »: multimodale, modulable, multi-stratégies, interopérable. Avec un tel template, un développeur ou un chercheur peut choisir parmi toutes les options disponibles (différents modèles de langage, frameworks de vision, systèmes experts, etc.) et les « bricoler » ensemble dans une architecture générale. En se référant aux critères d'**unifiabilité, modularité et adaptabilité** soulignés dans la littérature récente 2 7 9, cette approche garantit que le système pourra évoluer vers des tâches et contextes de plus en plus variés tout en conservant une cohérence globale.

Sources: travaux récents sur les architectures IA adaptatives et multimodales 3 1 2 7 9 4 6.

1 2 7 8 9 10 11 Generalist Multimodal AI: A Review of Architectures, Challenges and Opportunities
<https://arxiv.org/html/2406.05496v1>

3 5 [2406.11272] Development of an Adaptive Multi-Domain Artificial Intelligence System Built using Machine Learning and Expert Systems Technologies
<https://arxiv.org/abs/2406.11272>

4 No Free Lunch Theorem for Machine Learning - MachineLearningMastery.com
<https://machinelearningmastery.com/no-free-lunch-theorem-for-machine-learning/>

6 What Are Foundation Models? | IBM
<https://www.ibm.com/think/topics/foundation-models>