AIM²P — DOCUMENT DIRECTEUR

African AI for Mathematics & Mathematical Physics Initiative — Master Plan 2026–2028

Statut : Version 1.1 · Phase actuelle : Phase 3 (baselines expérimentales) · Auteur / Lead : Initiative AIM²P

Synthèse exécutive

L'initiative **AIM²P** vise à faire passer l'intelligence artificielle du stade de générateur de texte plausible à celui d'**assistant de recherche rigoureux, certifié et auto-correcteur** pour les mathématiques pures et la physique théorique. En intégrant de façon synergique modèles de raisonnement (LLM), calcul symbolique (SymPy/SageMath) et vérification formelle de preuve (Lean 4), AIM²P s'attaque aux étapes invalides et aux affirmations de réussite non vérifiées dans le raisonnement scientifique.

1. VISION ET POSITIONNEMENT STRATÉGIQUE

Le projet ambitionne d'évoluer d'un prototype individuel vers une infrastructure internationale de recherche ouverte. Il s'attaque à la limitation fondamentale des LLM actuels : l'absence de garantie de validité d'un raisonnement complexe tant qu'il n'est pas vérifié par un outil formel.

Positionnement (niche scientifique) :

- **Au-delà des olympiades :** tandis que les géants industriels (Google DeepMind, Meta, OpenAI) se focalisent sur les mathématiques de concours, et que les premiers prouveurs Lean pour la physique (apparus depuis octobre 2025) restent centrés sur l'information quantique et la physique de premier cycle, AIM²P vise la vérification formelle en **physique mathématique théorique** (mécanique hamiltonienne, géométrie différentielle, holographie, complexité de Krylov).
- **Architecture ouverte à trois composants :** des systèmes spécialisés associant LLM, calcul formel et Lean existent déjà pour des domaines précis (LIPS, NSPI, MechGeo) ; AIM²P vise une boucle ouverte et généraliste LLM ↔ solveur symbolique ↔ vérificateur formel, avec rétroaction et correction d'erreurs, où la fiabilité du critique symbolique est mesurée plutôt que présumée et où l'apport de chaque composant est mesuré par ablation.
- **Pôle d'excellence open science :** création d'une infrastructure de recherche ouverte basée en Afrique, dédiée à l'IA pour les mathématiques et la physique formelle.

2. PROBLÈME SCIENTIFIQUE ET HYPOTHÈSE CENTRALE

2.1 Problématique scientifique

Les modèles d'IA générative modernes produisent des démonstrations en apparence convaincantes, mais souffrent de faiblesses structurelles : apparition d'étapes invalides, hallucinations symboliques, imprécision sur les conditions aux limites et absence de mécanisme de preuve certifiée autonome.

2.2 Hypothèse de travail

Une architecture hybride neuro-symbolique combinant modèles de raisonnement guidé, solveurs d'algèbre computationnelle et assistant de preuve formelle (Lean 4) produit, à budget de calcul égal, davantage de solutions correctes et intégralement vérifiées que le même LLM utilisé seul ou filtré par le même vérificateur. Les hypothèses mesurables H1 à H3 sont pré-enregistrées dans le document 01 et le Research Proposal v0.1.

Problème → Raisonnement (LLM) → Calcul symbolique (SymPy) → Preuve formelle (Lean 4) → Auto-correction → Résultat certifié

3. OBJECTIFS SCIENTIFIQUES ET PRINCIPES DIRECTEURS

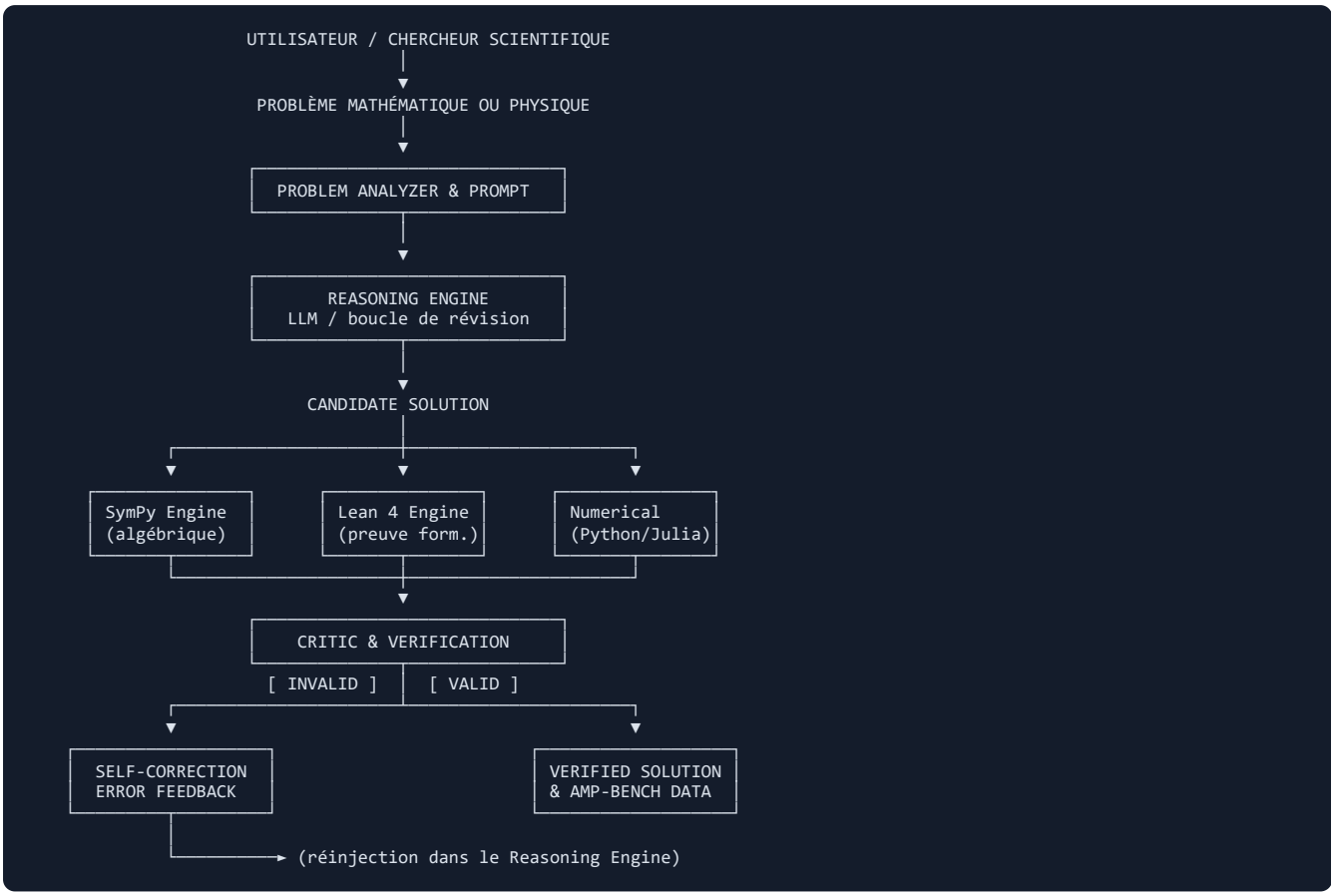
Code	Objectif	Description et cible scientifique
O1	Mathematical Reasoning	Générer un raisonnement étape par étape avec justification explicite des inférences.
O2	Formal Verification	Coupler Lean 4 pour certifier les propositions et lemmes formalisés (le noyau vérifie l'énoncé formel ; sa fidélité au problème est auditée séparément).
O3	Benchmark mathématique	Établir une évaluation multi-niveaux (N1 à N5), du calcul élémentaire à la recherche.
O4	Mathematical Physics	Modéliser les systèmes dynamiques, la mécanique quantique et la relativité.
O5	Scientific Discovery	Assister la formulation de conjectures, d'invariants et de nouvelles stratégies.
O6	African Capacity	Construire un réseau d'excellence open source accessible aux chercheurs africains.

Principes fondamentaux

- **Rigueur :** aucune affirmation sans preuve ou expérimentation.
- **Reproductibilité :** scripts, pipelines et graines aléatoires figés.
- **Open science :** datasets et code 100 % publics.
- **Comparaison stricte :** benchmarks systématiques contre les baselines de l'état de l'art.
- **Vérification automatique :** validation par le compilateur de preuve.
- **Neutralité technologique :** indépendance vis-à-vis d'un fournisseur unique de LLM.

4. ARCHITECTURE TECHNIQUE ET MOTEUR DE RAISONNEMENT

L'architecture cible est organisée en couches modulaires étanches :



5. CAS D'ÉTUDE PHARE : COMPLEXITÉ DE KRYLOV ET HOLOGRAPHIE

Pour faire la preuve de concept de la capacité d'AIM²P en physique mathématique, le projet intègre les travaux sur la **complexité de Krylov** (*Krylov complexity*) comme cas d'étude initial.

- **Dérivation formelle :** génération de la base d'opérateurs de Krylov via l'algorithme de Lanczos symbolique.
- **Vérification d'invariants :** auto-vérification de la croissance d'opérateurs (*operator growth*) et des conditions de régularité des coefficients de Lanczos b_n.
- **Transition classique-quantique :** comparaison automatisée des trajectoires dans l'espace de Krylov avec les modèles holographiques de gravité duale, limitée aux cas où les deux côtés admettent une forme fermée (SYK à double échelle et gravité JT, régime SL(2,$\mathbb{R}$)).

6. ROADMAP COMPLÈTE : 19 PHASES (P0 À P18)

Phase	Intitulé	Objectif majeur et livrable clé
P0	Cadrage scientifique	Revue de littérature, identification du gap, publication de `Research_Gap.md`.
P1	Cartographie du domaine	Matrice comparative des LLM, Lean 4, SymPy, Rocq (ex-Coq) et Isabelle.

Phase	Intitulé	Objectif majeur et livrable clé
P2	Spécification du benchmark	Création du dataset initial N1 à N5 (mathématiques et physique).
P3	Baselines expérimentales	Évaluation des modèles actuels (OpenAI, Anthropic, DeepSeek, Llemma…), versions figées au lancement des tests.
P4	Moteur neuro-symbolique	Développement du prototype Reasoning Engine V0.1 + Critic SymPy.
P5	Intégration de Lean 4	Mise en place de la boucle de génération et de vérification de tactiques Lean.
P6	Boucle d'auto-correction	Implémentation de la boucle agentique ou de la recherche arborescente guidée par les erreurs Lean.
P7	Physique mathématique	Module de mécanique hamiltonienne + cas d'étude complexité de Krylov.
P8	Benchmark AMP-Bench	Publication du benchmark de référence en physique mathématique.
P9	Publications scientifiques	Rédaction et soumission des preprints (NeurIPS, ICLR, revues de physique).
P10	Écosystème open source	Structuration de l'organisation GitHub AIM2P (dépôts complets et documentés).
P11	Constitution de l'équipe	Recrutement des rôles clés : Math Researcher, AI Engineer, Physicist, Dev.
P12	Advisory Board	Installation du conseil scientifique consultatif (SAB).
P13	Stratégie de financement	Campagne de subventions et de compute grants (5 k$ à 100 k$ et plus).
P14	Collaborations internationales	Partenariats visés : AIMS, Oxford, Perimeter Institute, ENS/INRIA.
P15	Conférences et outreach	Présentations aux workshops AI for Math et Mathematical Physics.
P16	African Research Network	Lancement de groupes de recherche, hackathons et séminaires.
P17–P18	Discovery & Scaling	Formulation autonome de conjectures et grand programme international.

7. PLAN D'EXÉCUTION DES 90 PREMIERS JOURS

- **Jours 1 à 15 (cadre scientifique) :** rédaction des documents fondateurs : `Research_Question.md`, `State_of_the_Art.md`, `Research_Gap.md`, `Research_Proposal_v0.1.md`.
- **Jours 16 à 30 (spécifications) :** sélection des jeux de données de baselines, protocole de test de robustesse et définition des métriques.
- **Jours 31 à 60 (baselines et tests) :** exécution des benchmarks sur des modèles de pointe fermés et ouverts (OpenAI, Anthropic, DeepSeek, Llemma…), versions figées au lancement des tests. Rapport d'erreurs et d'hallucinations.
- **Jours 61 à 90 (prototype V0.1) :** premier moteur de raisonnement Python couplé à SymPy et vérification syntaxique Lean 4. Publication du dépôt GitHub.

8. PISTES DE FINANCEMENT ET RESSOURCES

État au 17 novembre 2026, après l'audit de clôture de P0. L'affiliation institutionnelle est le verrou principal : la plupart des programmes de crédits d'API (Anthropic, OpenAI, Google Cloud, AWS, NVIDIA) exigent un rattachement académique ou associatif.

- **Niveau 1 (0 à 5 k$) :** ressources gratuites accessibles sans affiliation : Hugging Face ZeroGPU (2 Spaces gratuits, 5 minutes de GPU par jour) et demande de *community GPU grant* depuis un Space (documentation) ; Google TPU Research Cloud, TPU gratuits sur candidature, conditions actuelles à vérifier, utile seulement pour l'entraînement de modèles ouverts (TRC). Avec une affiliation : OpenAI Researcher Access (jusqu'à 1 000 $ de crédits) et crédits de recherche Google Cloud (1 000 à 5 000 $, pour doctorants, postdocs et enseignants).
- **Niveau 2 (5 k$ à 25 k$) :** principale source de financement en numéraire réellement ouverte à un fondateur sans affiliation : les *seed grants* de l'AI for Math Fund (Renaissance Philanthropy × XTX Markets), jusqu'à 100 000 $, examinés au fil de l'eau, ouverts aux individus du monde entier, avec des catégories « jeux de données et benchmarks » et « outils et infrastructure » (page officielle). Une demande de 15 000 à 50 000 $ paraît proportionnée à un pilote (suggestion, pas une consigne du fonds) ; la date limite de réception est à confirmer auprès du fonds. Avec une affiliation : Anthropic AI for Science (jusqu'à 20 000 $ de crédits d'API sur 6 mois). Partenaire d'affiliation visé en priorité : AIMS (master AI for Science).
- **Niveau 3 (25 k$ à 100 k$ et plus) :** tour principal de l'AI for Math Fund (100 000 $ à 1 M$; tour 2026 clos le 30 mars 2026, tour 2027 non annoncé) ; John Templeton Foundation (MPS), orientée institutions, dont les thèmes du tour 2026 ne couvraient pas l'IA pour les mathématiques (consignes 2027 attendues) ; Schmidt Sciences, qui finance par appels ciblés ; recrutement de postdocs et infrastructure GPU dédiée.
- **Pistes retirées ou fermées :** Modal for Academics (signalé en pause par une source secondaire), Microsoft AFMR (programme terminé), programmes de la Simons Foundation pour l'Afrique (anciens ou clos), tour « Claude Science » d'Anthropic (clos le 15 juillet 2026), LINGUA Africa (clos le 15 juin 2026), subventions académiques NVIDIA (réservées aux enseignants titulaires), crédits Together AI (sur invitation).

9. RÈGLE ABSOLUE DE CONDUITE SCIENTIFIQUE

> **Règle d'or :** ne jamais écrire « Notre IA est intelligente » ou « Notre IA résout les mathématiques ».
>
> Toujours écrire : « Notre système obtient X % de réussite sur le benchmark AMP-Bench contre Y % pour la baseline, avec Z % de solutions intégralement vérifiées par Lean 4. »

HISTORIQUE DES VERSIONS

- **v1.0 :** version initiale du document directeur.
- **v1.1 (17 novembre 2026) :** mises à jour ciblées après l'audit bibliographique de clôture de P0 : positionnement de la section 1 aligné sur les verrous révisés (des systèmes spécialisés LLM + calcul formel + Lean existent déjà) ; phase P6 ouverte aux boucles agentiques ; section 8 (financement) mise à jour au 17 novembre 2026.