LIVRE BLANC V1.0

Rédigé par

Serge Destin TAMPOLLA

4IR Advisor

serge.destin@tampolla.com

www.tampolla.com

Ankh AI

Le Protocole d'Intelligence Artificielle Collaborative Décentralisée

Un écosystème ouvert où le calcul, les talents et les données convergent pour créer l'IA de demain — détenue par tous, contrôlée par personne.

Rédigé par
Serge Destin TAMPOLLA
4IR Advisor
serge.destin@tampolla.com
www.tampolla.com

Résumé Exécutif

Ankh AI propose un protocole décentralisé qui coordonne, via une blockchain dédiée, les trois piliers fondamentaux de l'intelligence artificielle moderne : la puissance de calcul distribuée, l'expertise humaine collective, et les ressources data. Le protocole récompense équitablement chaque contributeur grâce au token $COLLAB et établit un système de vérification hybride (preuves ZK, vérification optimiste, échantillonnage aléatoire) garantissant l'intégrité des modèles produits. Ankh AI vise à produire des modèles de fondation compétitifs — de 7B à 100B+ paramètres — tout en démocratisant la propriété intellectuelle et la gouvernance de l'IA, créant ainsi une alternative résiliente, transparente et équitable aux solutions centralisées dominées par quelques acteurs. Le marché de l'IA décentralisée, estimé à 3,5 milliards de dollars en 2025, devrait atteindre 9,2 milliards d'ici 2034, porté par une demande croissante d'infrastructure IA accessible et transparente.

Table des Matières

1. Introduction

1.1. Le problème : Centralisation de l'IA

L'intelligence artificielle est à un point de bascule historique. Alors que les modèles de fondation atteignent des capacités sans précédent — rédaction de code, raisonnement mathématique, génération multimodale — leur développement et leur contrôle se concentrent entre les mains d'une poignée d'acteurs disposant de ressources computationnelles, humaines et data inégalées. Les chiffres parlent d'eux-mêmes : en 2025, les trois premiers acteurs du marché enterprise LLM API — Anthropic, OpenAI et Google — concentrent à eux seuls **88%** des dépenses d'entreprise en intelligence artificielle générative.[22] OpenAI, malgré une perte de parts de marché passant de 50% en 2023 à 27% fin 2025, maintient une valorisation de **260 milliards de dollars** et coordonne le projet Stargate de **500 milliards de dollars** en infrastructure IA.[22]

Cette concentration pose des risques systémiques qui dépassent le simple enjeu économique. La censure algorithmique s'exerce de manière opaque : les modèles centralisés filtrent, modifient ou refusent des réponses selon des critères que ni les utilisateurs ni les régulateurs ne contrôlent pleinement. L'extraction de valeur ne bénéficie qu'à un cercle restreint : les chercheurs, ingénieurs et annotateurs qui contribuent à l'amélioration des modèles ne perçoivent aucune participation proportionnelle à la valeur créée. Les barrières à l'entrée sont devenues insurmontables : un entraînement de six mois pour un modèle de nouvelle génération comme GPT-5 (nom de code Orion) coûte à lui seul plus de **500 millions de dollars** en ressources computationnelles, sans compter les talents et les données.[23] Le CEO d'Anthropic a confirmé que certains modèles en développement dépassent le milliard de dollars de coût d'entraînement.[33]

Les conséquences de cette centralisation s'étendent à la résilience technologique elle-même. Lorsqu'une poignée d'entités contrôle les infrastructures critiques de l'IA, une décision unilatérale — modification des conditions d'utilisation, suspension d'un service, faille de sécurité — peut affecter des centaines de millions d'utilisateurs instantanément. Pew Research a constaté que **34%** des adultes américains ont utilisé ChatGPT, dont une majorité de **58%** chez les moins de 30 ans.[22] Cette dépendance massive à des systèmes opaques constitue une vulnérabilité structurelle pour nos économies et nos démocraties.

1.2. La vision : Une IA véritablement publique et collaborative

Ankh AI naît d'une conviction fondamentale : l'intelligence artificielle, en tant que technologie transformative de notre époque, doit être traitée comme un bien public mondial. Cette vision ne s'oppose pas à l'excellence technique — au contraire, elle la conditionne. Les meilleurs modèles ne se construiront pas en silos corporatifs, mais en mobilisant l'intelligence collective de chercheurs, ingénieurs, fournisseurs de données et opérateurs de calcul répartis sur l'ensemble de la planète.

Le nom « Ankh » fait référence à l'ancien symbole égyptien de vie et de pouvoir créateur. Il incarne notre ambition : créer un système d'IA qui soit non seulement performant, mais aussi *vivant* dans le sens où il évolue, s'adapte et croît grâce à la contribution continue d'une communauté mondiale. Trois principes directeurs structurent cette vision.

Démocratiser l'accès à la création de valeur IA. Chaque acteur — du data center professionnel au particulier disposant d'une carte graphique gaming, du chercheur PhD à l'annotateur spécialisé, du fournisseur de corpus de données à l'auditeur de sécurité — doit pouvoir contribuer et être récompensé proportionnellement à l'utilité de sa contribution. Le protocole élimine les barrières d'entrée artificielles et crée un mérite technologique véritable.

Aligner les incitations entre tous les contributeurs. Le token $COLLAB et les smart contracts associés établissent un système de récompenses programmatiques où chaque contribution — calcul, expertise, données — reçoit une contrepartie tangible et immédiate. Les royalties perpétuelles garantissent que les contributeurs initiaux continuent de bénéficier de la valeur générée par leurs apports, même années après leur contribution.

Créer une alternative résiliente, transparente et équitable. En distribuant la production de l'IA sur un réseau mondial de contributeurs coordonnés par une blockchain, Ankh AI élimine les points de défaillance uniques. La gouvernance décentralisée via DAO garantit que les décisions concernant l'évolution des modèles se prennent de manière collective et transparente, sans intervention unilatérale d'une entité privée.

1.3. Pourquoi la blockchain est la solution

La technologie blockchain résout précisément les problèmes de coordination que pose la production collaborative d'IA à grande échelle. Quatre propriétés fondamentales la rendent indispensable à notre architecture.

Coordination sans confiance (trustless). Les contributeurs du réseau Ankh AI n'ont pas besoin de se connaître ni de se faire confiance. Les smart contracts exécutent automatiquement les engagements : distribution des tâches de calcul, vérification des résultats, paiement des récompenses. Cette propriété élimine le besoin d'intermédiaires centralisés et réduit drastiquement les frictions de coordination.

Traçabilité des contributions. Chaque contribution — heure de calcul fournie, ligne de code soumise, jeu de données déposé — est enregistrée de manière immuable sur la blockchain. Cette traçabilité constitue la base du système de réputation et de la distribution des royalties. Un chercheur dont une méthode d'optimisation est utilisée dans un modèle déployé peut prouver son apport et percevoir des redevances perpétuelles.

Distribution programmatique de la valeur. Les tokens programmables permettent de distribuer automatiquement la valeur créée par l'écosystème selon des règles transparentes et immuables. Lorsqu'un utilisateur paie pour une inférence, 70% du montant sont directement transférés au fournisseur de calcul, 15% aux contributeurs intellectuels via le RoyaltyEngine, 10% à la trésorerie DAO, et 5% sont brûlés — le tout exécuté atomiquement par un smart contract.

Gouvernance transparente. La DAO Ankh AI permet aux détenteurs de veCOLLAB de voter sur les décisions stratégiques : choix des architectures de modèles, allocation du budget de recherche, mise à jour du protocole, sélection des priorités de sécurité. Chaque vote est public et vérifiable, et les résultats s'exécutent automatiquement via des smart contracts, éliminant les risques de manipulation ou d'inaction.

Le marché de l'IA décentralisée, évalué à **3,5 milliards de dollars** en 2025, devrait croître à un taux annuel composé (CAGR) de **9,5%** pour atteindre **9,2 milliards de dollars** d'ici 2034.[49] Ce marché naissant, porté par la demande croissante d'infrastructure IA accessible et par la pression réglementaire pour une IA transparente, offre à Ankh AI un terrain de développement considérable.

2. Fondements Théoriques

2.1. Les trois piliers de l'IA moderne

La production d'intelligence artificielle de pointe repose sur trois piliers interdépendants : la puissance de calcul, les ressources humaines, et les données. Comprendre leur dynamique respective et leurs interactions est essentiel pour concevoir un protocole de coordination efficace.

Pilier 1 : La puissance de calcul

Le calcul constitue le moteur matériel de l'IA moderne. L'entraînement d'un modèle de fondation de 70 milliards de paramètres requiert des milliers de GPU haute performance fonctionnant en parallèle pendant des semaines, voire des mois. Le marché mondial des GPU pour data centers, estimé à **87,3 milliards de dollars** en 2024, devrait atteindre **228 milliards de dollars** d'ici 2030, avec un CAGR de **15,8%**.[55] La pénurie structurelle de GPU NVIDIA H100, avec des délais d'attente de 6 à 12 mois pour les acheteurs entreprises, a fait exploser les prix du cloud GPU, les fournisseurs décentralisés comme Akash proposant des tarifs **50 à 80%** inférieurs aux hyperscalers traditionnels.[34]

La décentralisation du calcul présente un avantage structurel majeur : elle permet d'agrégérer des ressources sous-utilisées à l'échelle planétaire. Les data centers professionnels, les stations de travail des créateurs, les GPU gaming des particuliers et même les ressources edge computing des appareils IoT représentent une capacité potentielle colossale. Akash Network a démontré la viabilité de ce modèle en atteignant **80%** de taux d'utilisation de son réseau GPU en 2025, avec plus de **34 300 locations** actives par trimestre.[34]

Pilier 2 : Les ressources humaines

Sans l'expertise humaine, le calcul reste inerte. Ce pilier englobe les chercheurs qui conçoivent les architectures de modèles, les ingénieurs ML qui les implémentent et les optimisent, les ingénieurs MLOps qui assurent le déploiement et la maintenance, les annotateurs qui fournissent les données d'alignement RLHF, et les experts domaine qui valident la qualité et la pertinence des sorties. Dans l'industrie, les coûts liés aux talents représentent typiquement **30 à 35%** du budget total de développement d'un modèle de fondation, contre **45 à 55%** pour le calcul et **10 à 20%** pour les données.

La décentralisation des ressources humaines permet de mobiliser une expertise mondiale bien plus large que celle accessible à une seule entreprise, même la plus riche. Les contributions peuvent provenir de chercheurs universitaires, d'ingénieurs indépendants, de communautés open-source, et d'experts sectoriels répartis dans des fuseaux horaires et des cultures diverses. Ce pluralisme intellectuel constitue un atout pour l'innovation et pour la robustesse des modèles produits.

Pilier 3 : Les données

Les données sont le carburant de l'apprentissage machine. La qualité, la diversité et le volume des corpus d'entraînement déterminent directement les capacités des modèles. Cependant, l'accès aux données de qualité est de plus en plus contraint : les grandes sources publiques (Common Crawl, Wikipedia, livres numérisés) ont été largement exploitées, et les données propriétaires demeurent verrouillées dans les silos des grandes entreprises technologiques. Les chercheurs d'OpenAI ont eux-mêmes reconnu que « l'internet public ne contient pas assez de données diversifiées et de haute qualité » pour entraîner les modèles de nouvelle génération.[23]

La décentralisation des données ouvre des possibilités radicalement nouvelles. Des corpus spécialisés — médicaux, juridiques, scientifiques, multilingues — peuvent être mis à disposition par leurs détenteurs tout en préservant la confidentialité via des techniques de compute-to-data, d'apprentissage fédéré et de chiffrement homomorphe. Le protocole Ankh AI établit un cadre où les données deviennent un actif productif, rémunéré équitablement, plutôt qu'un avantage compétitif verrouillé.

Interdépendances et ratios d'investissement

Les trois piliers ne fonctionnent pas de manière indépendante mais forment un système dynamique où chaque pilier renforce les deux autres. Des données de meilleure qualité réduisent les besoins en calcul (moins d'itérations nécessaires). Des chercheurs plus talentueux conçoivent des architectures plus efficaces qui extraient plus de valeur des données disponibles. Un calcul plus abondant permet d'explorer des approches plus ambitieuses et de valider plus rapidement les hypothèses.

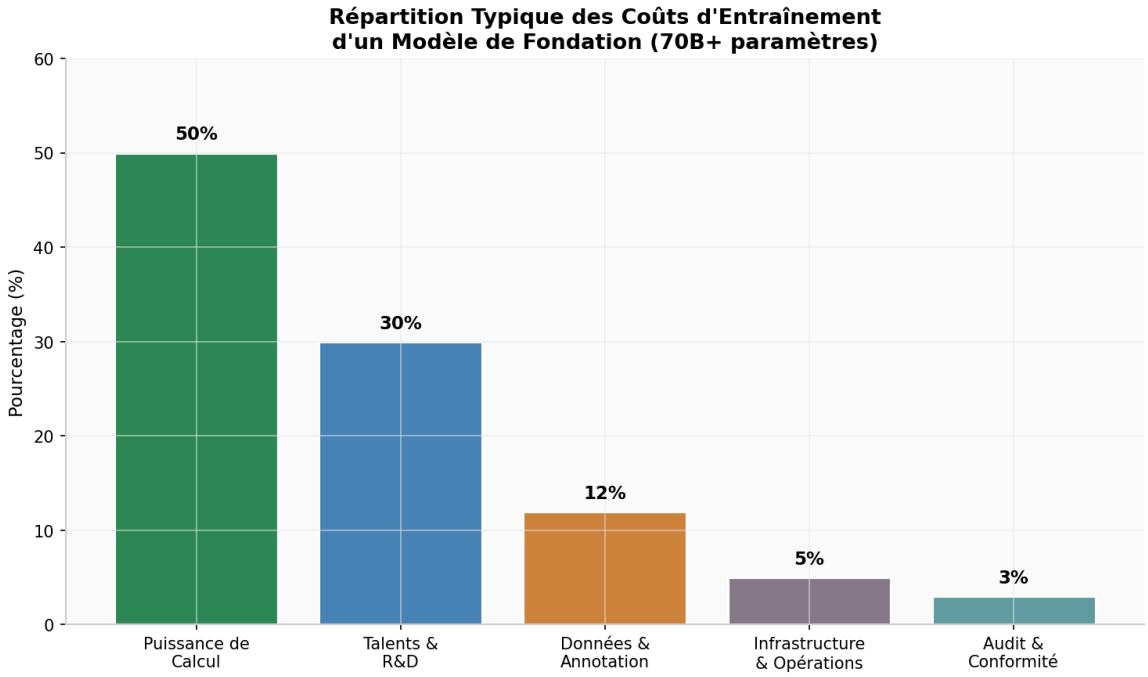

Figure 1 *Répartition typique des coûts d'entraînement d'un modèle de fondation de grande taille (70B+ paramètres)*

2.2. État de l'art des systèmes décentralisés existants

Plusieurs projets ont exploré la convergence blockchain-IA avant Ankh AI. Une analyse critique de leurs forces et faiblesses est nécessaire pour positionner notre contribution.

Bittensor (TAO)

Bittensor est le protocole décentralisé le plus établi dans le domaine, avec une capitalisation de **2,54 milliards de dollars** et un réseau de **128 à 129 sous-réseaux** actifs.[3] Son mécanisme de consensus par « Preuve d'Intelligence » (Proof of Intelligence) récompense les mineurs selon la qualité de leurs contributions IA, évaluées par des validateurs. Le token TAO, avec une offre maximale de 21 millions (calquée sur Bitcoin), a connu son premier halving en décembre 2025, réduisant l'émission journalière de 7 200 à 3 600 TAO.[2] Le système de gouvernance dynamique (dTAO) permet aux investisseurs d'allouer du capital vers des sous-réseaux spécifiques.

Les limites de Bittensor sont instructives. La concentration du pouvoir de validation entre les mains des 64 validateurs du Root Subnet pose un problème de centralisation : les 12 premiers validateurs contrôlent **79%** du stake total du réseau.[2] Le protocole ne coordonne pas nativement les trois piliers (calcul, talents, données) dans un système unifié : son focus principal est sur les sorties des modèles plutôt que sur la chaîne de production complète. Enfin,

Gensyn

Gensyn se concentre spécifiquement sur la vérification du calcul d'apprentissage machine à grande échelle. Son protocole de « Preuve d'Utilité de Calcul » (Proof of Useful Work) permet de vérifier que les nœuds du réseau ont effectivement exécuté les tâches d'entraînement demandées, via un mécanisme de délégation référée et des preuves cryptographiques.[39] Gensyn a développé des innovations notables comme NoLoCo (communication à faible bande passante pour l'entraînement distribué), CheckFree (tolérance aux pannes sans points de contrôle), et RL Swarm (apprentissage par renforcement collaboratif).[46]

Cependant, Gensyn reste principalement un protocole de calcul. Il ne traite pas nativement la coordination des ressources humaines ni la gouvernance des modèles produits. Son système de récompense, bien que sophistiqué pour le calcul, n'intègre pas de mécanismes de royalties perpétuelles pour les contributeurs intellectuels.

Ocean Protocol

Ocean Protocol a été pionnier dans l'économie des données décentralisées. Son architecture Compute-to-Data (C2D) permet d'exécuter des algorithmes sur des données sans les déplacer, préservant ainsi la confidentialité.[5] Les Data NFTs et les Datatokens créent un marché liquide pour les actifs data. Ocean a été l'un des trois projets fondateurs de l'Artificial Superintelligence Alliance (ASI) aux côtés de Fetch.ai et SingularityNET.[15]

La force d'Ocean réside dans la couche data, mais le protocole ne couvre ni le calcul d'entraînement ni la coordination des talents. L'intégration dans l'alliance ASI, bien qu'ambitieuse, a connu des difficultés : Ocean Protocol s'est retiré de l'alliance en octobre 2025.[15] Cette fragmentation illustre les défis de coordination entre protocoles aux objectifs partiellement divergents.

Fetch.ai / ASI Alliance

Fetch.ai, maintenant intégré dans l'ASI Alliance, se focalise sur les agents autonomes et la coordination décentralisée. L'alliance rassemble Fetch.ai (agents autonomes), SingularityNET (marketplace IA et recherche AGI), et CUDOS (infrastructure GPU distribuée), avec un token FET (transition vers ASI) à **450 millions de dollars** de capitalisation.[15] L'ASI Chain en développement vise à créer une blockchain L1 dédiée à l'IA agentique.

L'ASI Alliance souffre cependant de sa propre complexité organisationnelle. La sortie d'Ocean Protocol a fragilisé la vision d'intégration verticale. Le focus sur les agents autonomes, bien que

prometteur, ne répond pas directement au besoin de coordination des trois piliers de la production de modèles de fondation. Le mécanisme de tokenomics FET/ASI, sans hard cap et avec 83% de supply en circulation, offre moins de prévisibilité que les modèles à offre limitée.[15]

Akash Network et Render Network

Akash Network (AKT) et Render Network (RENDER) représentent la couche infrastructure GPU décentralisée. Akash a atteint **3,36 millions de dollars** de volume mensuel de calcul en Q3 2025,[37] avec AkashML proposant une API compatible OpenAI. Render Network a traité plus de **68 millions** de frames rendues et a lancé Dispersed, un sous-réseau dédié aux charges IA.[35]

Ces protocoles excellent dans la fourniture de ressources computationnelles mais n'adressent pas les couches supérieures : entraînement coordonné de modèles, gouvernance des architectures, récompense des contributions intellectuelles, et marketplace d'inférence.

Matrice de Différenciation des Protocoles d'IA Décentralisée
(Score 1-5 par dimension)

	Coordination 3 Piliers	Vérification Hybride ZK	Réputation SBT	Royalties Programmables	veTokenomics	Marketplace Inférence
Ankh AI (Protocole)	**5**	**5**	**5**	**5**	**5**	**5**
Bittensor (TAO)	3	2	1	2	2	3
Gensyn (AI)	2	4	1	1	1	2
Ocean Protocol (OCEAN)	2	1	1	3	1	2
Fetch.ai/ASI (FET)	3	1	1	1	1	3
Akash (AKT)	2	1	1	1	1	3

Figure 2 *Matrice de différenciation des protocoles d'IA décentralisée par dimension fonctionnelle (score 1-5)*

2.3. Innovations clés du protocole Ankh AI

Ankh AI ne se contente pas d'assembler les briques existantes. Le protocole introduit cinq innovations fondamentales qui le différencient de l'ensemble des approches précédentes.

Vérification hybride du calcul (ZK + Optimistic + Sampling)

Au lieu de s'appuyer sur un seul mécanisme de vérification, Ankh AI déploie une architecture à trois couches complémentaires. Pour l'inférence, les preuves à connaissance nulle (ZK-ML via ezKL) permettent de prouver cryptographiquement qu'un modèle a bien été exécuté sur des données spécifiques sans révéler ni le modèle ni les données.[4] L'outil ezKL convertit automatiquement les modèles ONNX en circuits compatibles ZK-SNARK, avec des preuves générées en quelques minutes et une vérification en secondes.[8] Pour l'entraînement, la vérification optimiste réduit les coûts : les résultats sont acceptés par défaut et contestés dans un délai donné. Enfin, l'échantillonnage aléatoire basé sur des fonctions vérifiables de hasard (VRF) sélectionne des comités de vérification indépendants pour auditer un sous-ensemble représentatif des calculs.

Réputation on-chain avec Soulbound Tokens

Ankh AI implémente un système de réputation sophistiqué basé sur les Soulbound Tokens (SBT) — des tokens non transférables attachés à l'identité d'un contributeur.[20] Chaque SBT enregistre un aspect de la réputation : contributions acceptées au code, évaluations par les pairs, ancienneté, domaines de spécialisation, et historique de vérification. Contrairement aux systèmes de réputation centralisés (étoiles, scores), les SBT sont immuables, vérifiables et composables. Ils servent à pondérer les votes de gouvernance, déterminer l'accès aux tâches sensibles, et qualifier les récompenses.

Royalties programmables pour les contributeurs intellectuels

Lorsqu'un modèle déployé sur Ankh AI génère des revenus (via l'inférence payante), une fraction automatique est redistribuée aux contributeurs dont le travail a été utilisé. Un chercheur dont une architecture d'optimisation est intégrée perçoit **0,1%** des frais d'inférence générés par le modèle, pour toute la durée de son exploitation. Ces royalties sont exécutées automatiquement par le RoyaltyEngine, sans intervention humaine, et sont inscrites de manière transparente sur la blockchain.

Valorisation dynamique des données par oracles

La valeur d'un jeu de données n'est pas statique : elle dépend de son impact sur la performance du modèle, de sa rareté, et de la demande actuelle. Ankh AI déploie des oracles de valorisation qui évaluent dynamiquement chaque contribution data selon une formule à trois facteurs :

$$V(D) = \alpha \cdot \Delta P(D) + \beta \cdot R(D) + \gamma \cdot Dem(D) \tag{1}$$

où $\Delta P(D)$ mesure l'amélioration de la perplexité du modèle après intégration des données D, $R(D)$ évalue la rareté relative du corpus, et $Dem(D)$ quantifie la demande exprimée par les consommateurs de modèles spécialisés.

Coordination unifiée des trois piliers

L'innovation la plus fondamentale d'Ankh AI réside dans l'intégration des trois piliers — calcul, talents, données — dans un protocole unifié. Contrairement à Bittensor (focus sur les sorties de modèles), Gensyn (focus sur le calcul), ou Ocean Protocol (focus sur les données), Ankh AI coordonne l'ensemble de la chaîne de production. Cette unification crée des effets de réseau puissants : les fournisseurs de calcul attirent les développeurs de modèles, qui attirent les fournisseurs de données, qui améliorent les modèles, qui génèrent plus de demande d'inférence, qui rémunère tous les contributeurs.

3. Architecture du Protocole

3.1. Vue d'ensemble architecturale

L'architecture d'Ankh AI s'organise en cinq couches distinctes mais interopérables, chacune responsable d'un aspect spécifique de la coordination. Cette approche modulaire permet l'évolution indépendante de chaque couche tout en maintenant la cohérence du système global.

Couche 1 — Blockchain de coordination. C'est le socle du protocole. Une blockchain dédiée (appchain Cosmos SDK) assure la finalité des transactions, l'exécution des smart contracts fondamentaux, et la gouvernance on-chain. Elle utilise un mécanisme de consensus Proof of Stake avec délégation, offrant un débit de plusieurs milliers de transactions par seconde et une finalité inférieure à 3 secondes.

Couche 2 — Smart Contracts fondamentaux. Six smart contracts principaux définissent la logique métier du protocole : ComputeRegistry (enregistrement des fournisseurs de calcul), TaskScheduler (distribution des tâches), DataMarketplace (gestion des actifs data), RoyaltyEngine (distribution des revenus), ReputationSystem (gestion des SBT et scores), et GovernanceCore (votes et exécution des propositions).

Couche 3 — Couche de vérification (Verification Layer). Cette couche hétérogène combine les nœuds validateurs exécutant les preuves ZK pour l'inférence, les comités de vérification optimiste pour l'entraînement, et les mécanismes d'échantillonnage aléatoire pour l'audit continu. Elle constitue la garantie d'intégrité du protocole.

Couche 4 — Couche de stockage et de données. Le stockage décentralisé (IPFS, Filecoin, Arweave) persiste les modèles, datasets et artefacts d'entraînement. L'indexation (via un protocole type The Graph) permet la découverte et la requête efficace. Le chiffrement et le contrôle d'accès (TEE, chiffrement homomorphe partiel) protègent les données sensibles.

Couche 5 — Application Layer. Les interfaces utilisateurs (portails web, CLI, SDK), les API d'inférence, les marketplaces de modèles spécialisés, et les outils de gouvernance constituent la couche d'interaction avec les utilisateurs finaux, développeurs et contributeurs.

3.2. La blockchain de coordination

Le choix d'une blockchain dédiée plutôt que d'une simple extension d'Ethereum ou d'une solution L2 générique repose sur des considérations de performance, de souveraineté et de coût.

Une appchain construite avec le Cosmos SDK offre plusieurs avantages décisifs pour Ankh AI.

Souveraineté et paramétrage. Contrairement à une L2 Ethereum soumise aux congestions et aux fluctuations de frais du L1, une appchain dédiée permet de calibrer précisément les paramètres du protocole : taille des blocs, frais de transaction minimum, temps de bloc, et mécanismes de consensus. Cette souveraineté est essentielle pour un protocole dont l'activité principale (distribution de milliers de micro-tâches de calcul par seconde) génère un volume de transactions incompatible avec les coûts actuels d'Ethereum mainnet.

Mécanisme de consensus. Ankh AI utilise un consensus Proof of Stake (PoS) avec délégation inspiré de Cosmos/Tendermint. Les validateurs sécurisent le réseau en stakeant des tokens $COLLAB et sont récompensés par les frais de transaction et une fraction de l'inflation programmée. Le temps de bloc cible est de **2 secondes**, avec une finalité instantanée (instant finality) via Tendermint BFT. Ce débit permet de traiter les milliers de transactions nécessaires à la coordination d'un entraînement distribué : soumission de preuves, attribution de récompenses, mises à jour de réputation.

Interopérabilité. Grâce au protocole IBC (Inter-Blockchain Communication) natif dans l'écosystème Cosmos, Ankh AI peut communiquer nativement avec les chaînes compatibles (Cosmos Hub, Osmosis, Akash, etc.). Des ponts sécurisés vers Ethereum et autres L1 EVM permettent le transfert de tokens et l'interaction avec l'écosystème DeFi établi. Cette interopérabilité garantit que $COLLAB reste liquide et accessible quel que soit l'écosystème de prédilection des utilisateurs.

Tableau 1 Comparaison des options d'infrastructure blockchain

Critère	L1 Dédiée (Ankh)	L2 Ethereum	Parachain Polkadot
Débit (TPS)	3 000+	500-2 000	1 000+
Finalité	~2 secondes	~12 secondes	~12 secondes
Frais de tx	<$0.001	$0.01 — 1	<$0.01
Souveraineté	Complète	Dépendante du L1	Partagée
Interopérabilité	**IBC + Ponts**	**Natif Ethereum**	XCM
Déploiement	Appchain Cosmos SDK	OP Stack / ZK Stack	Substrate

3.3. Les smart contracts fondamentaux

Six smart contracts constituent le cœur logique du protocole Ankh AI. Chacun encapsule une responsabilité métier spécifique et interagit avec les autres via des interfaces bien définies.

ComputeRegistry

Le ComputeRegistry maintient le registre canonique de tous les fournisseurs de calcul du réseau. Chaque nœud s'enregistre en stakeant un montant minimum de $COLLAB comme garantie de bonne conduite. Le contrat enregistre les spécifications matérielles (type de GPU, mémoire VRAM, bande passante réseau), les métriques de performance historiques (FLOPS mesurés, uptime, latence moyenne), et le statut actif du nœud. Le slashing — confiscation partielle ou totale du stake — punit les comportements malveillants : résultats de calcul incorrects, downtime prolongé, ou tentative de Sybil attack.

```
// Pseudo-code Solidity : Enregistrement d'un nœud de calcul
function registerNode(
    bytes32 hardwareFingerprint,
    uint256 stakeAmount,
    NodeSpec calldata specs
) external {
    require(stakeAmount >= MIN_STAKE, "Stake insuffisant");
    require(specs.vramGB >= MIN_VRAM, "VRAM insuffisante");

    nodes[msg.sender] = Node({
        stake: stakeAmount,
        specs: specs,
        status: NodeStatus.ACTIVE,
        reputationScore: BASE_REPUTATION,
        registeredAt: block.timestamp
    });

    totalStaked += stakeAmount;
    emit NodeRegistered(msg.sender, hardwareFingerprint);
}
```

TaskScheduler

Le TaskScheduler orchestre la distribution des tâches d'entraînement et d'inférence. Il maintient une file d'attente priorisée des tâches en attente, sélectionne les nœuds appropriés en fonction des exigences de la tâche et des capacités des nœuds, suit l'exécution en temps réel, et déclenche les paiements à la validation des résultats. Pour l'entraînement distribué, le TaskScheduler

implémente un protocole de type DiLoCo (Distributed Low-Communication) adapté, découpant l'entraînement en îlots de calcul asynchrones avec synchronisation périodique des gradients.[17]

DataMarketplace

Le DataMarketplace gère le cycle de vie complet des actifs data : dépôt, indexation, découverte, achat, et utilisation. Chaque dataset est représenté par un Data NFT (standard ERC-721 étendu) qui enregistre la propriété, la provenance, et les termes de licence. Les Datatokens (ERC-20) permettent l'accès fractionné et la liquidité des datasets. Le contrat intègre le Compute-to-Data : les algorithmes d'entraînement sont envoyés vers l'emplacement des données plutôt que l'inverse, préservant la confidentialité.[5]

RoyaltyEngine

Le RoyaltyEngine automatise la distribution des revenus générés par les modèles déployés. Chaque modèle enregistre sa « recette de contribution » — la liste des contributions intellectuelles, datasets, et optimisations qui ont permis sa création. Lorsqu'un utilisateur paie pour une inférence, le RoyaltyEngine décompose le montant selon les pourcentages programmés et distribue les paiements à chaque contributeur en temps réel, via des flux de paiement continus de type Superfluid.[24]

ReputationSystem

Le ReputationSystem gère l'émission, la mise à jour, et la révocation des Soulbound Tokens de réputation. Chaque SBT est non transférable et associé à une adresse Ethereum spécifique. Le système calcule un score composite à partir des métriques on-chain : contributions acceptées, évaluations par les pairs, ancienneté, spécialisation technique. Ce score détermine le poids du vote en gouvernance, l'accès aux tâches à haute valeur, et les multiplicateurs de récompense.

GovernanceCore

Le GovernanceCore implémente les mécanismes de proposition, vote, et exécution des décisions de la DAO. Il supporte plusieurs types de propositions : améliorations du protocole (PAP — Propositions d'Amélioration du Protocole), allocation budgétaire, sélection des priorités de recherche, et décisions d'urgence. Le système intègre des délais de timelock pour les modifications critiques, un mécanisme de vote quadratique pour les décisions d'allocation, et des quorums adaptatifs selon le type de décision.

3.4. Couche de vérification (Verification Layer)

L'intégrité des modèles produits par Ankh AI repose sur une couche de vérification sophistiquée qui combine trois mécanismes complémentaires, chacun adapté à un type spécifique de calcul.

Vérification ZK pour l'inférence

Pour l'inférence — l'exécution d'un modèle entraîné sur de nouvelles données — Ankh AI utilise les preuves à connaissance nulle via la bibliothèque ezKL. Le processus se déroule en quatre étapes : (1) le modèle est exporté au format ONNX, (2) ezKL le convertit automatiquement en un circuit compatible ZK-SNARK, (3) le circuit génère une preuve cryptographique que l'inférence a été exécutée correctement, (4) n'importe qui peut vérifier cette preuve en quelques secondes sans accéder au modèle ni aux données d'entrée.[8]

Cette approche résout un problème fondamental : comment un utilisateur peut-il être sûr que le fournisseur d'inférence exécute réellement le modèle promis (par exemple GPT-4 et non GPT-3) sans révéler les paramètres propriétaires ? Les preuves ZK offrent cette garantie cryptographique. Modulus Labs a démontré la faisabilité de cette approche à grande échelle en vérifiant des modèles jusqu'à **18 millions de paramètres** directement on-chain.[4]

Vérification optimiste pour l'entraînement

L'entraînement de modèles de fondation, bien plus intensif en calcul que l'inférence, ne peut pas être vérifé entièrement par des preuves ZK en raison des coûts prohibitifs. Ankh AI adopte une approche optimiste : les résultats d'entraînement (gradients agrégés, poids mis à jour) sont soumis et acceptés par défaut. Durant une période de contestation (challenge window, typiquement 24 heures), n'importe quel validateur peut soumettre une preuve de fraude (fraud proof) si le résultat est incorrect. Si la fraude est confirmée, le nœud malveillant est pénalisé (slashing) et le contestataire reçoit une récompense.

Comités de vérification aléatoire (VRF-based sampling)

Prenant le relais des mécanismes précédents, des comités de vérification sont sélectionnés aléatoirement parmi les nœuds validateurs via des fonctions vérifiables de hasard (VRF). Cette sélection garantie l'impartialité : aucun nœud ne peut prédire ou influencer sa sélection. Les comités auditent un échantillon statistiquement significatif des calculs, fournissant une couche de validation indépendante qui complète les preuves ZK et la vérification optimiste.

Slashing et incitations

Le mécanisme de slashing constitue le fondement économique de la sécurité. Tout fournisseur de calcul doit staker des tokens $COLLAB pour participer. En cas de comportement malveillant — preuve de calcul invalide, fraude confirmée, downtime prolongé sans justification — une partie du stake est confisquée et redistribuée aux validateurs honnêtes. Cette menace économique, combinée aux récompenses pour les validateurs vigilants, aligne les incitations de tous les acteurs vers l'honnêteté.

3.5. Couche de stockage et de données

La gestion des données dans un système décentralisé d'IA pose des défis uniques : volume massif (les modèles de fondation pèsent des centaines de gigaoctets), confidentialité (les données sensibles ne doivent pas être exposées), et traçabilité (la provenance des données doit être vérifiable). La couche de stockage d'Ankh AI adresse ces trois défis.

Stockage décentralisé

Les modèles entraînés, les datasets, et les artefacts d'entraînement sont persistés sur des réseaux de stockage décentralisés. IPFS (InterPlanetary File System) assure l'adressage par contenu et la distribution peer-to-peer. Filecoin ajoute une couche d'incentive économique garantissant la durabilité du stockage. Arweave offre une persistance permanente via son mécanisme de blockweave. Le choix du backend de stockage est abstrait par une couche d'interface unifiée, permettant aux utilisateurs de choisir le compromis coût/durabilité adapté à leurs besoins.

Indexation et découverte

Un système d'indexation décentralisé (inspiré de The Graph) permet de requêter efficacement les métadonnées des datasets et modèles : type de données, langue, domaine, qualité estimée, licence, historique d'utilisation. Cette indexation transforme un océan de données brutes en un catalogue navigable et découvrable.

Chiffrement et contrôle d'accès

Les données sensibles sont protégées par une combinaison de techniques. Les Environnements d'Exécution de Confiance (TEE), comme Intel SGX, AMD SEV-SNP, ou AWS Nitro Enclaves, créent des enclaves matérielles isolées où les données peuvent être traitées sans être exposées ni au

système d'exploitation hôte ni au fournisseur de cloud.[27] Le chiffrement homomorphe partiel permet certains calculs sur des données chiffrées. Ces techniques permettent l'utilisation de données médicales, financières, ou juridiques sans compromettre la confidentialité.

Provenance et traçabilité

Chaque dataset enregistré sur Ankh AI possède un historique de provenance immuable : origine, transformations appliquées, contributions d'annotation, et utilisations dans les modèles entraînés. Cette traçabilité est essentielle pour la conformité réglementaire (AI Act européen), l'audit des biais, et la transparence scientifique. Un consommateur de modèle peut retracer exactement quelles données ont contribué à son entraînement et sous quelles conditions de licence.

4. Mécanismes de Contribution et de Récompense

Le cœur économique d'Ankh AI réside dans sa capacité à convertir les contributions de trois types d'acteurs — fournisseurs de calcul, experts humains, et fournisseurs de données — en récompenses tangibles et équitablement distribuées. Cette section détaille les mécanismes spécifiques de chaque pilier, de l'enregistrement initial à la distribution finale des récompenses.

4.1. Pilier 1 : La puissance de calcul

4.1.1. Types de fournisseurs

Le réseau de calcul Ankh AI accueille quatre catégories de fournisseurs, chacune apportant des ressources complémentaires qui, agrégées, forment une infrastructure mondiale résiliente et hétérogène.

Data centers professionnels. Ces fournisseurs déploient des clusters de GPU haute performance (NVIDIA H100, H200, et prochainement Blackwell B200) dans des infrastructures optimisées. Ils offrent la plus grande densité de calcul et les garanties de disponibilité les plus élevées. Leur participation est essentielle pour les phases d'entraînement intensive des modèles de fondation. Akash Network a démontré la viabilité de ce modèle en intégrant plus de **65 data centers** professionnels à son réseau en 2025.[34]

Fournisseurs cloud décentralisés. Des réseaux comme Akash et Render Network agissent comme intermédiaires infrastructurels, agrégeant des ressources de multiples fournisseurs sous une interface unifiée. Ces partenaires stratégiques permettent à Ankh AI d'accéder rapidement à une capacité de calcul élasticité sans investissement direct en infrastructure physique.

Particuliers (GPU gaming, stations de travail). Les cartes graphiques des gamers et créateurs individuels — NVIDIA RTX 4090, RTX 5090, AMD Radeon — représentent une capacité sous-utilisée colossale. Un RTX 4090 clusterisé peut réduire les coûts d'inférence de **75%** par rapport à un H100 avec une perte de performance minimale pour le traitement par lots.[37] Ces contributeurs individuels, bien que moins puissants unitairement, forment une masse critique qui améliore la décentralisation et la résilience du réseau.

Edge computing (IoT, mobile). Les appareils périphériques — smartphones, objets connectés, micro-ordinateurs — peuvent contribuer aux tâches de calcul léger : pré-traitement de données,

validation de petits batches, et inférence de modèles quantifiés. Cette catégorie, encore émergente, devient pertinente avec l'amélioration des puces NPUs (Neural Processing Units) intégrées aux appareils mobiles modernes.

4.1.2. Preuve de Calcul Utile (Proof of Useful Work)

Ankh AI n'utilise pas une simple preuve de travail (Proof of Work) comme Bitcoin, où le calcul est « gaspillé » à résoudre des énigmes cryptographiques sans utilité pratique. À la place, le protocole implémente une **Preuve de Calcul Utile** : le travail effectué par les nœuds sert directement à l'entraînement ou à l'inférence de modèles d'IA. Chaque cycle GPU consommé produit une valeur réelle pour l'écosystème.

La métrique fondamentale est le **FLOPS vérifié** (Floating Point Operations Per Second) : le nombre d'opérations en virgule flottante effectivement exécutées et vérifiées. Cette mesure est complétée par des indicateurs de qualité de service : l'**uptime** (disponibilité du nœud, cible >99%), la **latence** (temps de réponse aux requêtes), et la **fidélité** (conformité des résultats aux attentes). Le processus de vérification, décrit en section 3.4, garantit que les FLOPS déclarés correspondent à un travail réel et utile.

La gestion des pannes est intégrée au protocole. Si un nœud devient indisponible pendant une tâche d'entraînement, le TaskScheduler redistribue automatiquement le workload aux nœuds de secours sans interruption du processus global. Le nœud défaillant subit une pénalité proportionnelle à l'impact de sa panne — une légère déduction pour un downtime bref, un slashing sévère pour une déconnexion prolongée ou répétée.

4.1.3. Récompenses par phase

Les récompenses des fournisseurs de calcul varient selon la phase du cycle de vie du modèle, reflétant les différentes valeurs économiques créées à chaque étape.

Phase 1 — Entraînement. Les fournisseurs participant à l'entraînement initial d'un modèle de fondation sont récompensés via un mécanisme d'enchère inversée. Le protocole publie un appel d'offres pour un volume de calcul donné, et les fournisseurs compétitifs proposent leurs tarifs. Le stake obligatoire garantit l'engagement : pour participer à un appel d'offres, un fournisseur doit verrouiller un montant de $COLLAB proportionnel à la valeur du contrat. Les récompenses sont versées en $COLLAB neufs (issus de l'inflation programmée), créant une distribution initiale large et équitable.

Phase 2 — Inférence. Les fournisseurs d'inférence perçoivent **70%** des frais payés par les utilisateurs du service. Ces frais peuvent être réglés en stablecoins (USDC, USDT) pour la stabilité, ou en $COLLAB avec un bonus de **5%** pour encourager l'utilisation du token natif. Ce

modèle génère des revenus récurrents et prévisibles pour les opérateurs de nœuds, proportionnels à la demande réelle des services IA.

Phase 3 — Fine-tuning. Les fournisseurs participant aux sessions de fine-tuning reçoivent une part des frais du modèle spécialisé résultant, plus un bonus de qualité si le modèle atteint des seuils de performance prédéfinis sur des benchmarks indépendants. Cette structure incite à l'excellence technique plutôt qu'à la simple fourniture de cycles de calcul.

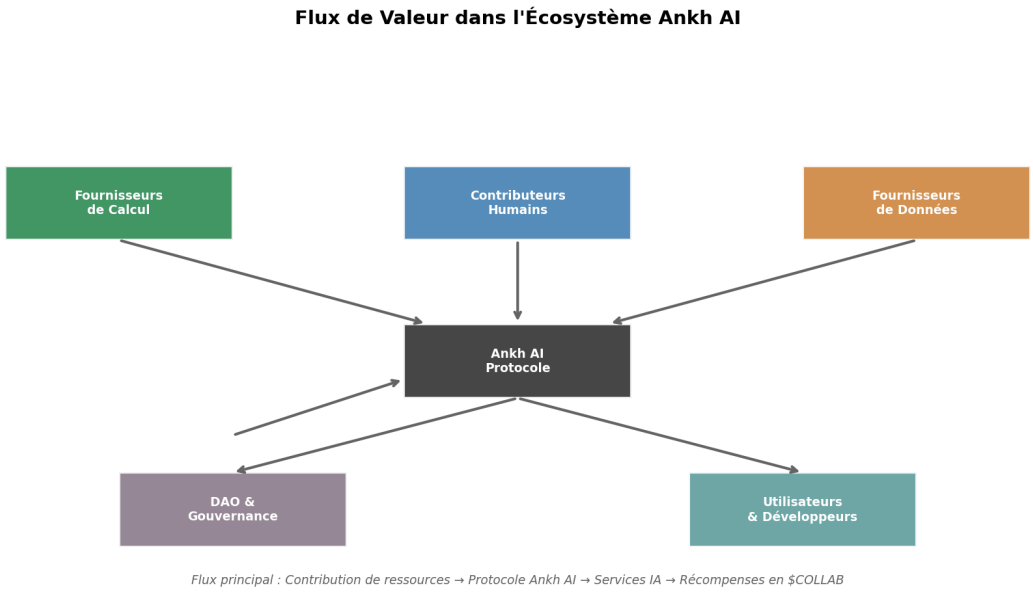

Figure 3 *Flux de valeur dans l'écosystème Ankh AI — de la contribution des ressources aux récompenses en $COLLAB*

4.2. Pilier 2 : Les ressources humaines

4.2.1. Catégories de contributeurs

La production d'intelligence artificielle de qualité exige une diversité d'expertises qu'aucune organisation ne possède dans son intégralité. Ankh AI reconnaît cinq catégories de contributeurs humains, chacune avec des mécanismes de contribution et de récompense adaptés.

Chercheurs. Ils conçoivent les architectures de modèles, les méthodes d'optimisation, et les techniques d'alignement. Leur contribution prend la forme de propositions d'amélioration du protocole, de publications de recherche appliquée, et d'expérimentations sur les architectures candidates. Un chercheur proposant une méthode d'optimisation des gradients qui réduit de

Ingénieurs ML. Ils implémentent les architectures conçues par les chercheurs, optimisent les performances des modèles, et assurent la compatibilité avec l'infrastructure distribuée. Leurs contributions sont mesurables en lignes de code, en amélioration de métriques de performance, et en résolution de problèmes techniques.

Ingénieurs MLOps/Infra. Ils gèrent le déploiement, la maintenance, et la scalabilité des systèmes d'entraînement et d'inférence. Leur expertise garantit que les modèles fonctionnent de manière fiable dans un environnement distribué hétérogène, avec des contraintes de bande passante, de latence, et de fiabilité variable.

Annotateurs et experts domaine. Ils fournissent les données d'alignement RLHF (Reinforcement Learning from Human Feedback) et les connaissances spécialisées nécessaires au fine-tuning par domaine. Un médecin annotant des cas cliniques pour un modèle médical, un juriste validant des analyses de contrats, ou un éducateur créant des données de tutorat personnalisé apportent une valeur difficilement substituable.[47]

Auditeurs. Ils évaluent la sécurité, les biais, et la conformité éthique des modèles produits. Leur rôle devient critique à mesure que les modèles sont déployés dans des domaines à haut risque (santé, finance, justice). Ils effectuent des audits de « red-teaming » pour identifier les vulnérabilités et les comportements potentiellement nuisibles.

4.2.2. Preuve de Compétence (Proof of Expertise)

Contrairement aux systèmes anonymes où n'importe qui peut prétendre à n'importe quelle expertise, Ankh AI établit une **Preuve de Compétence** vérifiable et immuable via les Soulbound Tokens (SBT).[20]

Chaque contributeur possède un portfolio on-chain de SBT qui enregistre : les contributions acceptées (code fusionné, propositions adoptées, datasets validés), les évaluations par les pairs (notoriété dans la communauté technique), l'ancienneté et la régularité de participation, et les domaines de spécialisation certifiés. Ces SBT sont non transférables : ils sont « liés à l'âme » de leur détenteur et ne peuvent être vendus ni cédés.

Le score de réputation composite, calculé à partir de ces SBT, détermine trois privilèges clés. Le **poids du vote** en gouvernance : un contributeur avec un score élevé a plus d'influence sur les décisions du protocole. L'**accès aux tâches** : certaines missions sensibles (audit de sécurité, validation de modèles critiques) sont réservées aux contributeurs ayant démontré leur expertise. Le **multiplicateur de récompense** : les contributeurs de haute réputation reçoivent un bonus sur leurs récompenses, reconnaissant la qualité et la fiabilité de leurs apports historiques.

4.2.3. Système de bounties et de royalties

Ankh AI déploie un écosystème de récompenses à quatre niveaux pour mobiliser et retenir les talents.

Bounties pour tâches spécifiques. La DAO publie des primes pour des missions définies : implémentation d'une feature, correction d'un bug critique, optimisation d'un algorithme. Les contributeurs soumettent leur travail, les évaluateurs de la communauté votent sur la meilleure soumission, et le gagnant reçoit la prime en $COLLAB. Ce mécanisme de marché compétitif garantit que les tâches sont réalisées par les meilleurs talents disponibles au meilleur prix.

Royalties perpétuelles. Lorsqu'une contribution intellectuelle est intégrée dans un modèle déployé, son auteur perçoit des redevances continues. Si un chercheur développe une technique d'optimisation des transformers qui est utilisée dans le modèle Ankh-GPT-7B, il reçoit **0,1%** de tous les frais d'inférence générés par ce modèle, pour toute la durée de son exploitation. Ces royalties sont programmées dans le RoyaltyEngine et s'exécutent automatiquement sans intervention humaine.

Grants de recherche exploratoire. La DAO alloue un budget annuel (issu de la trésorerie) à des projets de recherche à haut risque et haut potentiel. Contrairement aux bounties, les grants ne visent pas un livrable spécifique mais soutiennent l'exploration de nouvelles directions. Les chercheurs financés par des grants rendent compte périodiquement de leurs progrès à la communauté.

Streaming payments. Pour les contributions continues — maintenance d'une infrastructure, support technique, mise à jour régulière d'un modèle — Ankh AI utilise des flux de paiement continus via des protocoles de money streaming (Superfluid ou équivalent).[24] Un ingénieur MLOps responsable de la maintenance d'un cluster de calcul reçoit un paiement continu par seconde, plutôt qu'un salaire mensuel ou ponctuel. Ce modèle aligne parfaitement les incitations : le paiement s'interrompt automatiquement si la contribution cesse.

4.3. Pilier 3 : Les données

4.3.1. Types de données acceptées

La qualité et la diversité des données déterminent la performance des modèles. Ankh AI accepte et valorise cinq catégories de données, chacune répondant à des besoins spécifiques du pipeline d'entraînement.

Données brutes. Corpora issus du crawl web, archives numériques, corpus textuels divers. Ces données constituent le substrat initial de l'entraînement. Bien que volumineuses, elles nécessitent un important travail de nettoyage et de filtrage pour en extraire la valeur.

Données curées. Données nettoyées, dédupliquées, filtrées selon des critères de qualité. Le processus de curation, coûteux en temps et en expertise, ajoute une valeur considérable car il transforme des données brutes en ressources directement utilisables pour l'entraînement. Une dataset curée de haute qualité peut avoir plus d'impact sur la performance d'un modèle qu'un corpus brut dix fois plus volumineux.

Données d'alignement (RLHF). Paires de préférences, comparaisons de réponses, et évaluations humaines utilisées pour l'alignement des modèles. Ces données sont particulièrement précieuses car elles conditionnent directement le comportement, la sécurité, et l'utilité perçue des modèles finaux. La qualité de l'alignement distingue un modèle « utile et honnête » d'un modèle « capable mais dangereux ».[47]

Données spécialisées. Corpora domaine-spécifiques : dossiers médicaux anonymisés, corpus juridiques, données financières, code source, littérature scientifique. Ces données sont rares, coûteuses à acquérir, et déterminantes pour la performance des modèles spécialisés. Leur valeur économique est proportionnelle à leur rareté et à leur impact sur la qualité du modèle fine-tuné.

Données synthétiques. Données générées par des modèles existants pour augmenter les corpus d'entraînement ou créer des scénarios de test. Bien que moins précieuses que les données réelles, elles offrent une solution économique pour l'augmentation de données et la couverture de cas limites.

4.3.2. Preuve de Contribution de Données (Proof of Data)

La valorisation des contributions data est l'un des problèmes les plus délicats du protocole. Comment évaluer objectivement la valeur d'un jeu de données sans connaître à l'avance son impact sur un modèle qui n'existe pas encore ? Ankh AI résout ce défi via un système d'oracle de valorisation à trois facteurs.

$$V(D) = \alpha \cdot \Delta P(D) + \beta \cdot R(D) + \gamma \cdot Dem(D) \tag{2}$$

où $\alpha + \beta + \gamma = 1$ et :

- $\Delta P(D)$ représente l'impact sur la perplexité du modèle — la réduction relative de la métrique de perplexité après intégration du dataset D dans l'entraînement. Une baisse de perplexité signifie que le modèle prédit mieux les données de test.

- $R(D)$ mesure la rareté relative du corpus — l'inverse de la fréquence d'apparition de contenus similaires dans les datasets existants. Un dataset contenant des données médicales spécialisées en swahili aura une rareté élevée.
- $Dem(D)$ quantifie la demande exprimée — le nombre et la valeur des requêtes pour des modèles entraînés sur ce type de données.

La vérification de l'unicité utilise une combinaison de hachage cryptographique et de similarité cosinus vectorielle. Chaque document est haché et son embedding vectoriel est comparé à ceux des datasets existants pour détecter les doublons et les plagiats. La vérification de la qualité emploie un « modèle proxy » — un petit modèle entraîné rapidement — pour estimer l'impact potentiel du dataset sur un modèle de taille réelle.

4.3.3. Marketplace de données

Le DataMarketplace d'Ankh AI constitue une place de marché décentralisée où les fournisseurs de données rencontrent les consommateurs (entraîneurs de modèles, développeurs de fine-tuning).

Dépôt et découverte. Les fournisseurs déposent leurs datasets via une interface standardisée. Les métadonnées (type, langue, domaine, taille, qualité estimée, licence) sont indexées et rendues consultables. Les acheteurs peuvent parcourir, filtrer, et évaluer les datasets avant achat.

Compute-to-Data (C2D). Plutôt que de télécharger les données sensibles, les acheteurs envoient leurs algorithmes d'entraînement vers l'emplacement où résident les données.[5] L'exécution se déroule dans un environnement sécurisé (TEE ou container sandbox), et seuls les résultats (gradients, métriques) quittent le périmètre de sécurité. Cette approche préserve la confidentialité des données tout en permettant leur exploitation.

Modèles de prix. Deux modèles coexistent. L'enchère permet aux fournisseurs de fixer un prix minimum et aux acheteurs d'enchérir. Le prix fixe offre une transaction directe à un tarif déterminé. Les licences programmables encodent dans des smart contracts les conditions d'utilisation : durée, nombre d'entraînements autorisés, droits de redistribution, et obligations d'attribution.

Confidentialité avancée. Pour les données hautement sensibles, Ankh AI supporte l'apprentissage fédéré (les modèles sont entraînés localement et seuls les gradients agrégés sont partagés) et la confidentialité différentielle (du bruit statistique est ajouté aux résultats pour protéger l'identité des individus dans les données).[27] Ces techniques, combinées aux TEE, offrent un éventail complet de protections adaptées à chaque niveau de sensibilité.

5. Les Trois Phases du Cycle de Vie du Modèle

La production d'un modèle d'intelligence artificielle sur Ankh AI suit un cycle de vie structuré en trois phases distinctes, chacune caractérisée par ses objectifs techniques, ses mécanismes de coordination, et ses modalités de récompense. Ce cycle s'inscrit dans une logique d'amélioration continue : chaque modèle produit alimente la demande pour le calcul, les données, et l'expertise, qui à leur tour alimentent le modèle suivant.

5.1. Phase 1 : Entraînement initial du modèle de fondation

5.1.1. Objectifs et spécifications

La phase d'entraînement initial a pour objectif de produire un modèle de fondation compétitif qui servira de base à toutes les applications et spécialisations ultérieures. Le premier modèle cible — nom de code **Ankh-GPT-7B** — vise une taille de 7 milliards de paramètres, offrant un compromis optimal entre capacité et efficacité pour un déploiement sur infrastructure décentralisée.

Le budget estimé pour cette première itération est de **5 à 10 millions de dollars**, répartis sur une période de **3 à 6 mois**. Ce budget couvre : l'allocation de ressources computationnelles (enchères inverses auprès des fournisseurs de calcul), les récompenses pour les contributeurs de données et d'expertise, les audits de sécurité, et les réserves opérationnelles. À titre de comparaison, l'entraînement de GPT-4 a coûté plus de 100 millions de dollars,[33] mais l'efficacité des architectures modernes et l'agrégation de ressources décentralisées permettent à Ankh AI d'atteindre des résultats compétitifs à une fraction de ce coût. DeepSeek R1 a démontré en 2025 qu'une architecture efficace pouvait égaler les performances de GPT-4 à un coût de calcul dix fois inférieur.[40]

5.1.2. Coordination de l'entraînement distribué

L'entraînement distribué à grande échelle sur un réseau hétérogène de contributeurs pose des défis techniques considérables. Ankh AI adapte le protocole DiLoCo (Distributed Low-Communication) développé par Google DeepMind pour orchestrer l'entraînement par îlots.[17]

Le principe consiste à diviser l'entraînement global en « îlots » de calcul indépendants — typiquement des data centers ou des clusters régionaux — qui entraînent localement des copies du modèle avec leurs propres données. Périodiquement, les îlots synchronisent leurs gradients

via un mécanisme de gossip léger plutôt que par un all-reduce coûteux en bande passante. Cette approche, qualifiée de « Découpled DiLoCo » par DeepMind, isole les pannes locales : si un îlot devient indisponible, les autres continuent leur progression sans interruption.[17]

La gestion de l'hétérogénéité matérielle est assurée par le ComputeRegistry qui classe les nœuds selon leurs capacités (VRAM, FLOPS, bande passante) et assigne des workloads adaptés. Les nœuds les plus puissants traitent les batches les plus volumineux, tandis que les ressources plus modestes contribuent via des tâches de fine-tuning local ou de validation. La tolérance aux pannes est native : le TaskScheduler redistribue automatiquement les tâches des nœuds défaillants et reprend l'entraînement depuis le dernier point de synchronisation valide.

5.1.3. Distribution des récompenses

Les récompenses de la phase d'entraînement sont prélevées sur un pool d'émission initial dédié. La répartition suit les ratios sectoriels établis :

Tableau 2 Répartition des récompenses — Phase 1 (Entraînement initial)

Catégorie	Ratio	Description
Puissance de calcul	45-50%	Fournisseurs de GPU, data centers, particuliers
Talents & R&D	20-25%	Chercheurs, ingénieurs ML, MLOps
Données	15-20%	Fournisseurs de datasets, annotateurs
Trésorerie & Audit	5-10%	Réserves DAO, audits de sécurité
Réserve communautaire	5%	Bounties, grants, programme d'incitation

Une période de vesting de **12 à 24 mois** s'applique aux récompenses des contributeurs majeurs, alignant leurs intérêts avec la réussite à long terme du protocole. Les récompenses sont libérées progressivement, avec un cliff initial de 3 mois suivi d'un déblocage linéaire.

5.1.4. Financement initial

Le financement de la première phase d'entraînement provient de trois sources complémentaires. La vente de tokens aux early backers (investisseurs stratégiques, fonds de venture capital spécialisés en IA et Web3) fournit le capital initial. Les grants de fondations Web3 (Ethereum Foundation, Cosmos Hub, et programmes de grants décentralisés) apportent un financement non dilutif. Les partenariats stratégiques avec des data centers, des fournisseurs de données, et

des institutions académiques offrent des ressources en nature (crédits de calcul, datasets, expertise) qui réduisent les besoins en liquidités.

5.2. Phase 2 : Déploiement et inférence

5.2.1. Architecture d'inférence décentralisée

Une fois le modèle de fondation entraîné et validé, il est déployé sur le réseau d'inférence décentralisé. Cette architecture doit répondre à trois exigences apparemment contradictoires : la performance (latence faible), la scalabilité (capacité à absorber des pics de demande), et la décentralisation (pas de point de défaillance unique).

Le **routage des requêtes** est assuré par un mécanisme de load balancing on-chain. Lorsqu'un utilisateur soumet une requête d'inférence, le TaskScheduler sélectionne dynamiquement le nœud optimal en fonction de la latence estimée, de la charge actuelle, et du niveau de service demandé. Les requêtes peuvent être routées vers plusieurs nœuds en parallèle pour les tâches critiques, avec un mécanisme de consensus sur le résultat.

Trois **niveaux de service** sont disponibles. Le niveau Standard offre une latence modérée (~2-5 secondes par requête) à coût réduit, adapté aux applications grand public. Le niveau Faible Latence garantit des réponses en moins d'une seconde pour les applications temps réel (chatbots, assistants vocaux). Le niveau Confidentiel utilise des TEE pour exécuter l'inférence dans un environnement isolé, garantissant que ni le modèle ni les données d'entrée ne sont exposés au fournisseur de calcul.[27]

Le **scaling automatique** ajuste le nombre de nœuds actifs en fonction de la demande mesurée. En période de faible activité, seuls les nœuds les plus efficaces restent actifs. Lors des pics de demande, le protocole réactive automatiquement des nœuds en veille et distribue la charge. Ce mécanisme garantit une utilisation optimale des ressources tout en maintenant la qualité de service.

5.2.2. Modèle économique de l'inférence

La tarification de l'inférence s'adapte à trois modèles de consommation. Le **pay-per-use** facture chaque requête au prix unitaire, adapté aux utilisations sporadiques. L'**abonnement** offre un volume mensuel de requêtes à tarif préférentiel, adapté aux applications régulières. La **licence** permet le déploiement privé d'un modèle sur l'infrastructure du client, avec des redevances basées sur l'utilisation.

La répartition des frais d'inférence est programmée dans le RoyaltyEngine et s'exécute automatiquement :

Tableau 3 Répartition des frais d'inférence

Destination	Pourcentage	Description
Fournisseur de calcul	70%	Rémunération directe du nœud exécutant l'inférence
Royalties contributeurs	15%	Redistribution aux auteurs des contributions intellectuelles
Trésorerie DAO	10%	Financement du développement, R&D, marketing
Brûlage (burn)	5%	Destruction définitive de tokens, pression déflationniste

Le mécanisme de **burn** (brûlage de 5% des frais) crée une pression déflationniste structurelle sur l'offre de $COLLAB. Plus le réseau est utilisé, plus de tokens sont brûlés, réduisant mécaniquement l'offre circulante et soutenant la valeur du token — à l'image du mécanisme EIP-1559 d'Ethereum.

5.2.3. Cas d'usage consommateurs

L'API d'inférence Ankh AI, compatible avec les standards OpenAI, permet aux développeurs de remplacer simplement leur endpoint existant par le nœud Ankh AI le plus proche. Cette compatibilité minimise les frictions de migration et élargit immédiatement l'adresseable market aux millions d'applications déjà intégrées avec les APIs IA centralisées.

Les **applications grand public** — chatbots, générateurs de contenu, assistants créatifs — bénéficient d'une tarification compétitive et d'une résistance à la censure. Les **agents IA autonomes** équipés de portefeuilles crypto peuvent interagir directement avec le protocole pour s'autoalimenter en ressources de calcul, créant un écosystème d'agents économiquement autonomes. Les **solutions entreprises** peuvent déployer des instances dédiées avec des garanties de confidentialité (via TEE) et des SLA personnalisés, à des tarifs significativement inférieurs aux solutions centralisées équivalentes.

5.3. Phase 3 : Fine-tuning et spécialisation

5.3.1. Processus de spécialisation

La phase de fine-tuning transforme le modèle de fondation généraliste en modèles spécialisés optimisés pour des domaines verticaux spécifiques. Ce processus, ouvert et démocratisé, suit un pipeline structuré en cinq étapes.

Proposition de projet. Tout membre de la communauté peut proposer un projet de fine-tuning en spécifiant le domaine cible, les datasets requis, l'architecture de fine-tuning envisagée, et le budget estimé. La proposition est soumise à la DAO pour approbation.

Financement. Le projet peut être auto-financé (le proposant fournit les ressources), financé par la DAO (via un vote d'allocation de trésorerie), ou financé par crowdfunding (collecte de contributions de la communauté contre une part des futurs revenus).

Contribution des données spécialisées. Les fournisseurs de données domaine-spécifiques déposent leurs datasets sur le DataMarketplace. Le système de valorisation dynamique (section 4.3.2) détermine leur compensation.

Exécution du fine-tuning distribué. Le fine-tuning s'exécute sur le réseau de calcul décentralisé, avec les mêmes mécanismes de vérification que l'entraînement initial. La charge computationnelle est inférieure (quelques GPU-jours contre des milliers pour l'entraînement initial), ce qui abaisse significativement la barrière à l'entrée.

Validation et déploiement. Le modèle spécialisé est évalué sur des benchmarks indépendants du domaine. S'il atteint les seuils de performance définis, il est validé par le comité de vérification et déployé sur le réseau d'inférence.

5.3.2. Tokenisation des modèles spécialisés

Chaque modèle spécialisé validé est représenté par un **NFT fractionné** (ERC-1155) qui encode la propriété collective du modèle. Les fractions sont distribuées aux contributeurs du projet de fine-tuning : fournisseurs de données, opérateurs de calcul, et chercheurs ayant conçu l'approche de spécialisation.

Ces fractions confèrent un droit proportionnel sur les revenus générés par le modèle. Lorsqu'un utilisateur paie pour une inférence utilisant le modèle spécialisé, les frais sont automatiquement répartis entre les détenteurs de fractions selon leurs parts. Un marché secondaire permet l'échange de ces fractions, créant une liquidité pour les investissements dans les modèles spécialisés.

5.3.3. Exemples de verticales

Les verticales de fine-tuning couvrent les domaines où les modèles spécialisés apportent une valeur significativement supérieure aux modèles généralistes.

Médical. Diagnostic assisté, analyse de dossiers patients, recherche biomédicale. Les données médicales, hautement sensibles et réglementées, bénéficient particulièrement de l'approche Compute-to-Data et des TEE d'Ankh AI qui préservent la confidentialité tout en permettant l'entraînement.

Juridique. Analyse de contrats, recherche jurisprudentielle, assistance à la rédaction d'actes juridiques. Les corpus juridiques multilingues et multijuridictionnels sont rares et précieux, ce qui donne une valeur élevée aux contributions data dans cette verticale.

Finance. Analyse de marchés, détection de fraude, génération de rapports financiers. La latence faible et la confidentialité des données de trading font de l'inférence confidentielle (via TEE) un atout majeur pour cette verticale.

Éducation. Tutorat personnalisé adaptatif, génération d'exercices, évaluation automatisée. Les datasets éducatifs multilingues et multiculturalisés permettent de créer des modèles accessibles aux populations non anglophones, réduisant les inégalités d'accès à l'IA.

Code. Assistant de développement, revue de code, génération de tests. Avec Anthropic capturant **54%** du marché du coding IA en 2025,[22] cette verticale représente une opportunité considérable pour un modèle open-source spécialisé.

6. Gouvernance

La gouvernance d'Ankh AI incarne le principe fondamental de décentralisation : les décisions concernant le protocole sont prises collectivement par ses contributeurs et utilisateurs, non par une entité centralisée. Cette section détaille l'architecture organisationnelle, les mécanismes de vote, et les processus de décision qui permettent à la DAO Ankh AI de fonctionner efficacement à grande échelle.

6.1. Structure de la DAO

La DAO Ankh AI adopte une structure fédérée qui équilibre participation large et expertise technique. Cette architecture évite les écueils des DAO purement directes (susceptibles aux attaques Sybil et à l'apathie des votants) et des structures trop centralisées (qui trahissent l'esprit décentralisé).

Assemblée générale. Tous les détenteurs de veCOLLAB (tokens de gouvernance verrouillés) constituent l'Assemblée générale. Elle vote sur les décisions stratégiques majeures : changements de paramètres fondamentaux du protocole, allocation de grandes enveloppes budgétaires, et élection des conseils spécialisés. Le pouvoir de vote est pondéré par la quantité de veCOLLAB détenue et la durée du verrouillage.

Conseils spécialisés. Trois conseils permanents apportent une expertise ciblée : le **Conseil Technique** (architectes blockchain, chercheurs IA, ingénieurs sécurité) évalue les propositions d'amélioration technique ; le **Conseil Économique** (économistes, analystes tokenomics, financiers) conseille sur les questions de tokenomics et de trésorerie ; et le **Conseil d'Éthique** (philosophes, juristes, représentants de la société civile) veille aux implications éthiques et sociétales des décisions.

Sous-DAOs. Quatre sous-organisations autonomes gèrent des domaines fonctionnels spécifiques avec une autonomie opérationnelle significative : la **Data DAO** supervise le marketplace de données et les politiques de qualité ; la **Compute DAO** gère les relations avec les fournisseurs de calcul et les politiques de pricing ; la **Models DAO** coordonne les initiatives d'entraînement et les standards de validation ; et la **Ecosystem DAO** gère les partenariats, les grants, et le marketing.

Mécanisme de délégation. Les détenteurs de veCOLLAB peuvent déléguer leur pouvoir de vote à des représentants de confiance. Cette délégation résout le problème de l'apathie des votants : un contributeur qui ne souhaite pas suivre toutes les propositions peut confier son vote à un expert

dont il partage les valeurs, tout en conservant la possibilité de reprendre son vote à tout moment.

6.2. Mécanisme de vote

Le système de vote d'Ankh AI s'inspire du modèle veTokenomics de Curve Finance, ajusté aux spécificités d'un protocole d'IA.[12]

Verrouillage et veCOLLAB. Pour obtenir du pouvoir de vote, un détenteur de $COLLAB doit verrouiller ses tokens pour une durée déterminée — de 1 semaine minimum à 4 ans maximum. Le verrouillage génère des veCOLLAB (vote-escrowed COLLAB) dont la quantité est proportionnelle à la fois au montant verrouillé et à la durée du verrouillage. La formule exacte est :

$$veCOLLAB = COLLAB_{locked} \times \frac{t_{remaining}}{t_{max}} \tag{3}$$

où $t_{remaining}$ est le temps restant avant déverrouillage et t_{max} = 4 ans. Cette formule garantit qu'un détenteur verrouillant 100 $COLLAB pour 4 ans obtient autant de pouvoir de vote que quelqu'un verrouillant 400 COLLAB pour 1 an. Le pouvoir de vote décroît linéairement à mesure que la date de déverrouillage approche, incitant à des re-verrouillages réguliers pour maintenir une influence maximale.

Quorums adaptatifs. Le quorum requis varie selon l'importance de la décision. Les propositions techniques mineures (mise à jour d'un paramètre) requièrent 10% de participation et une majorité simple. Les propositions majeures (changement d'architecture, allocation de trésorerie >5%) exigent 25% de participation et une majorité qualifiée de 60%. Les décisions critiques (modification des règles de gouvernance) nécessitent 40% de participation et une majorité des deux tiers.

Vote quadratique pour l'allocation. Pour les décisions d'allocation de ressources (budget de recherche, récompenses de grants), Ankh AI utilise le vote quadratique où le coût d'un vote est proportionnel au carré du nombre de voix allouées. Cette mécanique atténue l'influence disproportionnée des « baleines » (détenteurs de grands montants) et donne plus de poids aux préférences des petits contributeurs.

Délais de timelock. Toutes les décisions, sauf celles marquées comme urgences de sécurité, sont soumises à un délai de timelock de 48 heures entre l'approbation du vote et l'exécution effective.

Ce délai permet aux contributeurs de réagir en cas de décision controversée et constitue une protection contre les attaques de gouvernance.

6.3. Processus de décision

Le cycle de vie d'une proposition d'amélioration du protocole (PAP — Proposition d'Amélioration du Protocole) suit un processus rigoureux conçu pour garantir la qualité des décisions tout en maintenant l'agilité.

Cycle de vie d'une proposition

Phase 1 — Discussion. Tout membre de la communauté peut soumettre une idée sur le forum de gouvernance. Cette phase de discussion libre dure minimum 7 jours et permet d'affiner la proposition, d'identifier les objections, et de construire un consensus préliminaire.

Phase 2 — Formalisation. La proposition est formalisée selon un template standard (description, motivation, spécifications techniques, analyse d'impact, budget si applicable) et soumise on-chain via le GovernanceCore. Un dépôt de 100 $COLLAB est requis pour éviter le spam ; ce dépôt est remboursé si la proposition atteint le quorum.

Phase 3 — Vote. La période de vote dure 5 jours. Les détenteurs de veCOLLAB expriment leur vote (Pour, Contre, Abstention). Leur pouvoir de vote est pondéré par leur solde veCOLLAB au moment du snapshot.

Phase 4 — Exécution. Si la proposition est approuvée (quorum atteint, majorité requise obtenue), elle entre dans la file d'attente d'exécution. Après le délai de timelock, le smart contract GovernanceCore exécute automatiquement la décision programmée.

Gestion des urgences

Certaines situations exigent une réaction immédiate sans passer par le cycle complet de gouvernance : découverte d'une vulnérabilité de sécurité critique, détection d'un modèle dangereux (biases systémiques, comportements nuisibles), ou menace sur la stabilité économique du protocole.

Dans ces cas, le **Conseil de Sécurité** — un comité de 7 membres élus pour leur expertise technique — peut déclencher une procédure d'urgence. La décision du Conseil s'exécute immédiatement mais doit être ratifiée par un vote de l'Assemblée générale dans les 7 jours suivants. Si la ratification échoue, la décision est annulée et les membres du Conseil ayant voté pour subissent une pénalité de réputation.

Résolution des conflits

En cas de désaccord profond au sein de la communauté, un mécanisme de **fork social** est prévu. Si une faction significative (minimum 20% des veCOLLAB) rejette une décision majeure, elle peut initier un processus de fork qui crée une nouvelle instance du protocole avec les paramètres souhaités. Cette possibilité, coûteuse en ressources, sert de dernier recours et incite les factions à négocier des compromis.

6.4. Trésorerie

La trésorerie de la DAO constitue le principal moteur financier du développement du protocole. Sa gestion transparente et décentralisée est essentielle pour maintenir la confiance de la communauté et des investisseurs.

Sources de revenus. La trésorerie est alimentée par trois flux. Les **frais de protocole** (10% de chaque transaction d'inférence) constituent la source principale et la plus prévisible. L'**inflation programmée** (émission de nouveaux $COLLAB selon le calendrier d'émission) alimente la croissance. Les **grants et partenariats** (subventions de fondations, investissements stratégiques) apportent des ressources complémentaires.

Allocation budgétaire. La répartition des dépenses est votée annuellement par la DAO. Le budget type se répartit ainsi : **40%** R&D (développement du protocole, recherche sur les architectures de modèles, amélioration de la vérification) ; **25%** marketing et adoption (campagnes de sensibilisation, programmes d'ambassadeurs, documentation) ; **20%** audits et sécurité (audits de smart contracts, bug bounty, tests de pénétration) ; **10%** réserves opérationnelles ; et **5%** événements communautaires et hackathons.

Sécurisation des fonds. La trésorerie est sécurisée via un système multi-signatures (multi-sig) requérant l'approbation de 5 signatures sur 9 pour toute dépense supérieure à un seuil défini. Les dépenses exceptionnelles sont soumises à un timelock de 7 jours. Les rapports de transparence, détaillant chaque entrée et sortie de fonds, sont publiés on-chain en temps réel et synthétisés dans un rapport mensuel.

7. Tokenomics

La tokenomics d'Ankh AI constitue le système nerveux économique du protocole. Conçue pour aligner les intérêts de tous les acteurs sur le long terme, elle combine des mécanismes inflationnistes (pour récompenser les contributions), déflationnistes (pour soutenir la valeur du token), et de gouvernance (pour démocratiser le contrôle).

7.1. Le token $COLLAB

Le token natif $COLLAB est un token d'utilité et de gouvernance qui remplit quatre fonctions fondamentales au sein de l'écosystème.

Utilité.
COLLAB est le moyen de paiement privilégié pour les services du protocole. Les utilisateurs peuvent p
COLLAB, bénéficiant d'un bonus de 5% par rapport aux paiements en stablecoin. Les fournisseurs de calcul perçoivent leurs récompenses en $COLLAB, créant une demande structurelle liée à l'activité réelle du réseau.

Gouvernance. Le pouvoir de vote dans la DAO est proportionnel à la quantité de veCOLLAB détenue. Seuls les tokens verrouillés confèrent du pouvoir de vote, ce qui aligne les gouvernants avec la stabilité à long terme du protocole.

Récompense. Les contributions au réseau — calcul, données, expertise — sont rémunérées en $COLLAB nouvellement émis (dans la limite du calendrier d'émission) ou prélevés sur les frais de protocole. Cette double source de récompenses garantit la pérennité du système de rémunération.

Staking. Les détenteurs de $COLLAB peuvent staker leurs tokens pour sécuriser le réseau (en devenant validateurs ou en déléguant à des validateurs) et percevoir des récompenses de staking proportionnelles à leur participation.

Techniquement, $COLLAB est un token standard ERC-20 (pour la compatibilité EVM) avec des extensions spécifiques : mécanisme de burn programmable, fonctions de mint contrôlées par gouvernance, et hooks pour le système de streaming payments.

7.2. Distribution initiale

La supply totale de $COLLAB est fixée à **1 milliard de tokens** (1,000,000,000). Cette offre maximale est immuable — aucun mécanisme ne permet d'augmenter ce plafond, garantissant la rareté du token à long terme. La distribution initiale au Token Generation Event (TGE) se répartit comme suit :

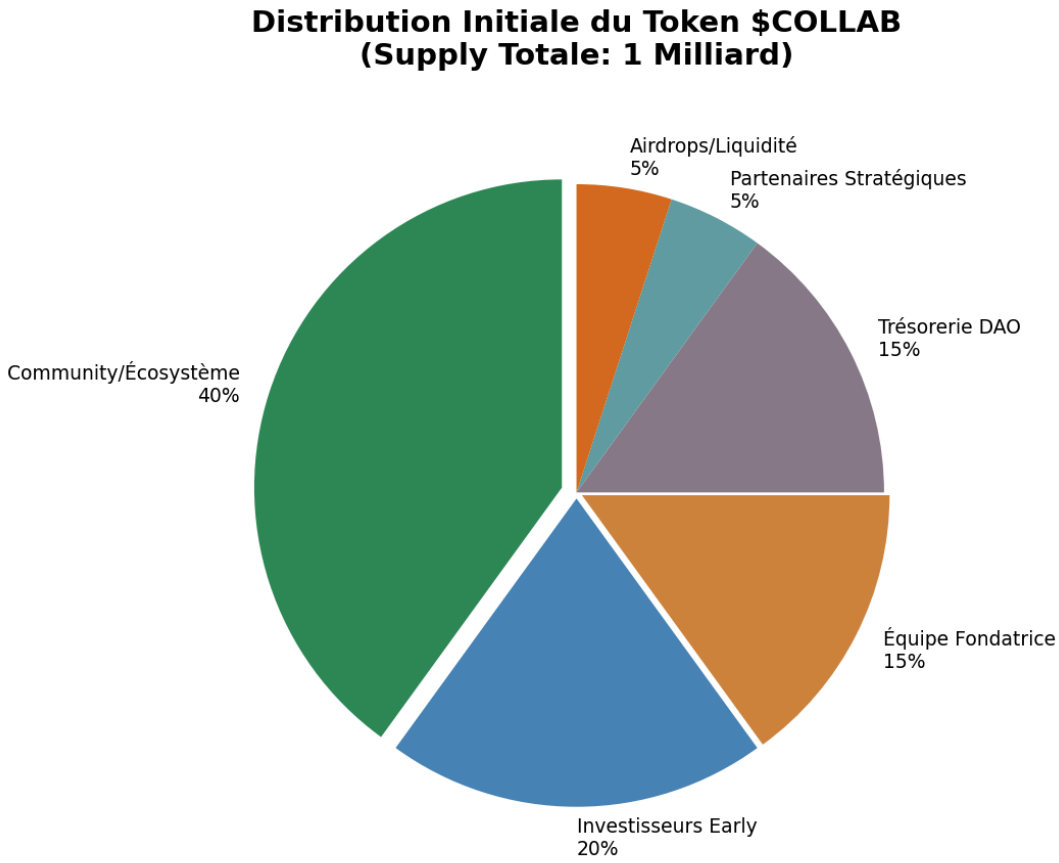

Figure 4 *Distribution initiale du token $COLLAB (supply totale: 1 milliard)*

Le vesting (déblocage progressif) est un mécanisme de sécurité économique qui empêche les dumping massifs et aligne les intérêts des détenteurs initiaux avec la santé à long terme du protocole. L'équipe fondatrice, avec le vesting le plus long (5 ans), démontre son engagement à long terme. Les 400 millions de $COLLAB alloués à l'écosystème alimentent les récompenses des contributeurs, les programmes d'incitation, et les grants communautaires sur les premières années.

7.3. Mécanismes de valeur

La valeur du token $COLLAB repose sur un équilibre subtil entre forces inflationnistes et déflationnistes, conçu pour soutenir une croissance durable.

Émission inflationniste décroissante. Au-delà de la distribution initiale, de nouveaux $COLLAB sont émis pour récompenser les contributions au réseau (calcul, données, expertise). Le taux d'émission suit une courbe décroissante inspirée de Bitcoin mais adaptée aux besoins d'un protocole d'IA : taux annuel de **8%** la première année, diminuant de **20%** chaque année jusqu'à atteindre un taux plancher de **1%**. Cette courbe garantit une rémunération attractive pour les premiers contributeurs tout en assurant la durabilité à long terme.

Burn déflationniste. 5% de chaque frais d'inférence est définitivement brûlé (détruit). Ce mécanisme crée une corrélation directe entre l'utilisation du réseau et la réduction de l'offre : plus Ankh AI est utilisé, plus la supply de $COLLAB diminue. Si le volume d'inférence atteint des niveaux significatifs, le burn peut théoriquement dépasser l'émission, rendant le token nettement déflationniste.

Staking et verrouillage. Les tokens stakés ou verrouillés en veCOLLAB sont retirés de l'offre circulante. Avec des incitations attractives au staking (rendements de 8-15% APR selon la participation), une fraction significative de la supply sera naturellement verrouillée, réduisant la pression vendeuse et soutenant le prix.

Frais de protocole. 10% de chaque transaction d'inférence est prélevé au profit de la trésorerie DAO. Ces frais, dénommés en $COLLAB, créentunedemandeconstantepourletoken : laDAOdoitacquérirdesCOLLAB pour financer ses opérations, ou les contributeurs doivent en détenir pour accéder aux services.

7.4. Modèle veTokenomics

Le système veTokenomics (vote-escrow tokenomics) est l'un des piliers de l'alignement des incitations à long terme. Inspiré du modèle pionnier de Curve Finance,[12] il a été adapté aux spécificités d'un protocole d'IA collaborative.

Verrouillage. Un détenteur de $COLLAB peut verrouiller ses tokens pour une durée allant de 1 semaine à 4 ans via le smart contract VotingEscrow. En échange, il reçoit des veCOLLAB qui représentent son pouvoir de vote et ses droits aux récompenses boostées. Les veCOLLAB sont non transférables — ils sont liés à l'adresse qui a effectué le verrouillage.

Pouvoir de vote proportionnel. Le pouvoir de vote est calculé selon la formule linéaire : veCOLLAB = $COLLAB_locked × (t_remaining / t_max). Un verrouillage maximal (4 ans) donne un ratio 1:1 ; un verrouillage de 2 ans donne un ratio 1:0.5. Ce mécanisme récompense l'engagement à long terme : les acteurs les plus engagés ont le plus d'influence sur les décisions.

Boost de récompenses. Les stakers long terme reçoivent un boost multiplicateur sur leurs récompenses de contribution. Un contributeur verrouillant ses récompenses pour 4 ans peut percevoir jusqu'à **2,5 fois** plus de récompenses qu'un contributeur non verrouillé. Ce boost s'applique aux récompenses de calcul, de données, et d'expertise.

Decay linéaire. Le pouvoir de vote décroît linéairement à mesure que la date de déverrouillage approche. Un contributeur avec 1000 veCOLLAB verrouillés pour 4 ans verra son pouvoir passer à 750 veCOLLAB après 1 an, 500 après 2 ans, etc. Ce decay incite à des re-verrouillages réguliers pour maintenir une influence maximale, créant un effet de « flywheel » de verrouillage continu.

7.5. Flux de valeur dans l'écosystème

L'écosystème Ankh AI génère et redistribue de la valeur selon un cycle vertueux où chaque acteur contribue et bénéficie de manière interdépendante.

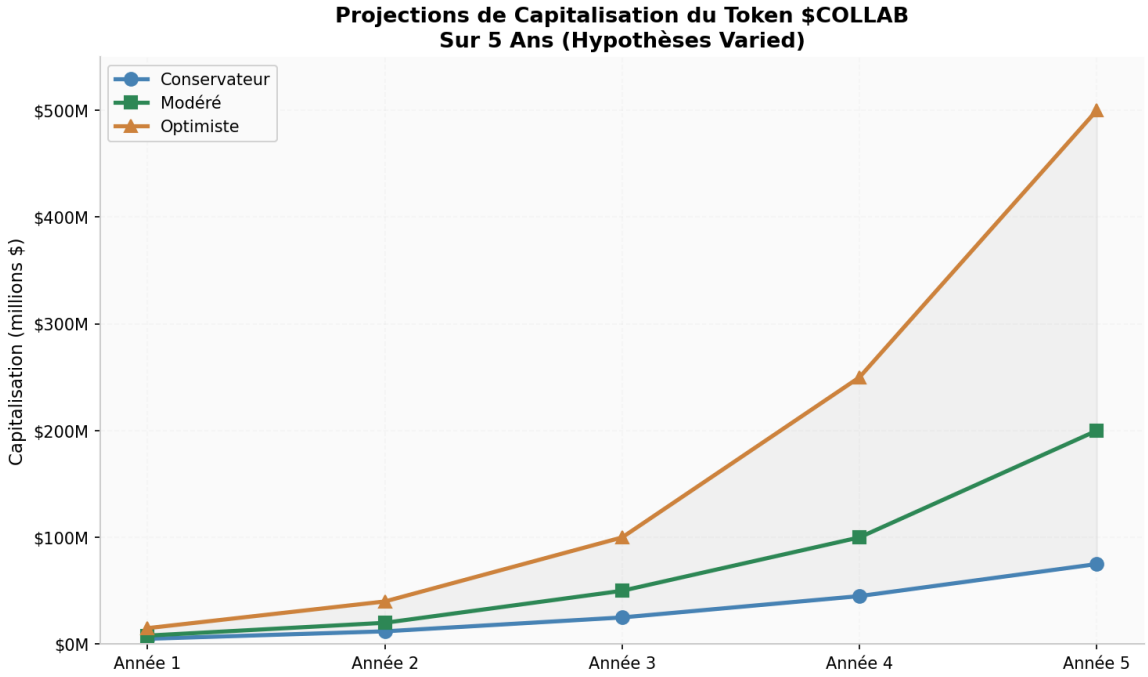

Figure 5 *Projections de capitalisation du token $COLLAB sur 5 ans — scénarios conservateur, modéré et optimiste*

Sources de revenus et distribution

Les revenus du protocole proviennent principalement de trois sources. Les **frais d'inférence** représentent le flux le plus important : chaque requête API génère des revenus distribués selon le schéma 70/15/10/5 (fournisseur/royalties/DAO/burn). Les **frais de marketplace** sont prélevés sur chaque transaction de données (typiquement 1-2% du montant). Les **frais de fine-tuning** sont facturés pour l'utilisation de l'infrastructure de spécialisation de modèles.

Projections financières sur 5 ans

Trois scénarios illustrent les trajectoires possibles du protocole. Le scénario **conservateur** suppose une adoption modérée, avec un premier modèle 7B compétitif mais une pénétration limitée du marché enterprise. Le scénario **modéré** suppose une croissance régulière, l'entraînement réussi d'un modèle 70B, et des partenariats stratégiques significatifs. Le scénario **optimiste** suppose une adoption rapide, un modèle 100B+ compétitif avec les solutions centralisées, et une reconnaissance réglementaire favorable.

Tableau 5 Projections de capitalisation (millions $) par scénario

Scénario	Année 1	Année 2	Année 3	Année 4	Année 5
Conservateur	5	12	25	45	75
Modéré	8	20	50	100	200
Optimiste	15	40	100	250	500

Ces projections, bien que spéculatives, s'appuient sur des hypothèses explicites : croissance du marché global de l'IA décentralisée à 9,5% CAGR,[49] expansion du marché GPU cloud à 35% CAGR,[50] et capacité d'Ankh AI à capturer 2-5% du marché de l'inférence décentralisée d'ici l'année 5. Les hypothèses macroéconomiques (régulation, concurrence, adoption blockchain) constituent les principales variables d'incertitude.

8. Sécurité, Confidentialité et Éthique

La production d'intelligence artificielle à grande échelle sur une infrastructure décentralisée soulève des défis de sécurité, de confidentialité, et d'éthique qui dépassent ceux des systèmes centralisés traditionnels. Cette section détaille les mécanismes de protection déployés par Ankh AI pour garantir l'intégrité du protocole, la confidentialité des données, l'alignement des modèles, et l'inclusion démocratique.

8.1. Sécurité du protocole

La sécurité d'Ankh AI repose sur une approche de défense en profondeur qui combine des mesures préventives, détectives, et correctives à chaque couche de l'architecture.

Vecteurs d'attaque et contre-mesures

Attaques Sybil. Un acteur malveillant pourrait créer de multiples identités fictives pour manipuler les votes de gouvernance ou capturer une part disproportionnée des récompenses. La contre-mesure combine le staking obligatoire (coût élevé par identité), le système de réputation SBT (qui ne peut pas être transféré ni fabriqué instantanément), et l'analyse on-chain des patterns de comportement pour détecter les réseaux de nœuds coordonnés.

Attaques sur l'entraînement (poisoning). Un fournisseur de données pourrait injecter des données malveillantes pour corrompre le modèle (backdoors, biais induits). Le système de vérification hybride (section 3.4) détecte les anomalies via l'échantillonnage aléatoire, et le slashing économique punit les fournisseurs de données frauduleux. La diversité des sources de données (aucun fournisseur ne contrôle plus de 10% du corpus d'entraînement) limite l'impact d'une source compromise.

Attaques sur l'inférence (évasion). Un utilisateur pourrait tenter d'extraire des informations sur le modèle ou de le faire générer du contenu nuisible via du prompt engineering avancé. Les nœuds d'inférence exécutent des filtres de sécurité temps réel, et les modèles déployés intègrent des mécanismes d'alignement RLHF qui résistent aux tentatives d'évasion.

Attaques économiques. Une entité disposant de capital massif pourrait tenter de manipuler le prix du token ou de prendre le contrôle de la gouvernance. Le plafond de supply (1 milliard de tokens), les périodes de vesting longues, et le mécanisme veTokenomics (qui récompense le verrouillage à long terme) rendent de telles attaques prohibitively coûteuses.

Audits et bug bounty

Tous les smart contracts fondamentaux font l'objet d'audits de sécurité par des cabinets indépendants de réputation établie (ex: Trail of Bits, OpenZeppelin, CertiK) avant tout déploiement mainnet. Les audits portent sur la logique métier, la résistance aux attaques connues (reentrancy, overflow, front-running), et l'optimisation du gas.

Un programme de bug bounty permanent récompense les chercheurs en sécurité qui identifient des vulnérabilités. Les primes s'échelonnent de **1 000 $COLLAB** pour des bugs mineurs à **500 000 $COLLAB** pour des vulnérabilités critiques pouvant compromettre les fonds. Ce programme externalise la détection des failles à la communauté mondiale de white-hat hackers.

Circuit de mise à jour d'urgence

Un mécanisme de mise à jour d'urgence, contrôlé par le Conseil de Sécurité multi-sig, permet de patcher rapidement une vulnérabilité critique sans attendre le cycle complet de gouvernance. Ce pouvoir est encadré : toute utilisation doit être justifiée publiquement dans les 24 heures, et la DAO peut annuler la mise à jour par un vote dans les 7 jours suivants.

8.2. Confidentialité des données

Dans un écosystème où des données sensibles (médicales, financières, juridiques) circulent entre des contributeurs qui ne se connaissent pas, la confidentialité n'est pas une option mais une exigence fondamentale.

Environnements d'exécution de confiance (TEE)

Les TEE constituent la brique principale de la confidentialité. Un TEE (Trusted Execution Environment) crée une enclave matérielle isolée au sein du processeur, où les données peuvent être traitées en clair sans être exposées au système d'exploitation hôte, à l'hyperviseur, ou au fournisseur de cloud.[27] Même si le serveur physique est compromis, les données dans le TEE restent protégées. Intel SGX, AMD SEV-SNP, et AWS Nitro Enclaves représentent les implémentations les plus matures, avec NVIDIA Confidential Computing étendant la protection aux GPU.[29]

Apprentissage fédéré

Pour les données trop sensibles pour quitter leur emplacement physique (données de patients hospitaliers, registres financiers régulés), Ankh AI supporte l'apprentissage fédéré. Le modèle est entraîné localement sur chaque site, et seuls les gradients agrégés (dénuisés de toute information individuelle identifiable) sont partagés avec le réseau central. Cette approche, combinée au chiffrement des gradients, garantit que les données brutes ne quittent jamais leur périmètre de sécurité.

Preuves ZK pour la vérification sans révélation

Les preuves à connaissance nulle permettent de vérifier qu'un calcul a été effectué correctement sans révéler ni les données d'entrée ni les paramètres du modèle.[4] Cette propriété est essentielle pour les scénarios où le propriétaire du modèle souhaite prouver qu'il utilise bien l'architecture promise (par exemple, un modèle médical certifié) sans divulguer ses paramètres propriétaires.

Conformité réglementaire

Ankh AI intègre les exigences des principaux cadres réglementaires. Le **RGPD européen** est respecté via le droit à l'effacement (les données personnelles peuvent être retirées sur demande), le droit à la portabilité, et la minimisation des données. L'**AI Act européen**, entré en vigueur le 1er août 2024 avec une application progressive jusqu'en août 2026,[54] impose des obligations de transparence, de traçabilité, et de gestion des risques que le système de provenance on-chain d'Ankh AI satisfait naturellement. Les modèles déployés sur Ankh AI sont classifiés selon le système de risque de l'AI Act, et les modèles à haut risque sont soumis à des procédures de conformité renforcées.

8.3. Alignment et safety

L'alignement — la capacité d'un système d'IA à poursuivre les objectifs prévus par ses concepteurs et utilisateurs — constitue l'un des défis les plus profonds de l'IA moderne. Dans un contexte décentralisé, ce défi devient encore plus complexe car il n'existe pas d'entité centralisée pour « imposer » l'alignment.

RLHF distribué

Ankh AI implémente un mécanisme d'apprentissage par renforcement à partir du feedback humain (RLHF) distribué. Les annotateurs du réseau, répartis géographiquement et culturellement, fournissent des évaluations de préférence sur les sorties des modèles. Cette diversité d'annotateurs réduit le risque de biais culturels ou idéologiques qui affectent souvent les modèles centralisés (entraînés principalement par des annotateurs d'Amérique du Nord). Le mécanisme de consensus par réputation pondère les évaluations selon l'expertise démontrée de chaque annotateur.[47]

Red-teaming communautaire

Avant tout déploiement majeur, chaque modèle est soumis à une période de « red-teaming » ouvert où la communauté est invitée à identifier des failles, des biais, et des comportements potentiellement nuisibles. Les participants au red-teaming sont récompensés pour chaque vulnérabilité validée, créant une incitation économique à la découverte proactive des problèmes. Ce processus distribué, inspiré des pratiques de sécurité informatique, a démontré son efficacité dans l'identification de comportements indésirables que les tests internes n'avaient pas détectés.

Circuit d'arrêt d'urgence décentralisé

En cas de détection d'un comportement dangereux d'un modèle déployé (génération systématique de contenu nuisible, manipulation des utilisateurs, exploitation de vulnérabilités de sécurité), un mécanisme d'arrêt d'urgence permet de suspendre temporairement le modèle. Contrairement à un système centralisé où un exécutif peut prendre cette décision unilatéralement, le circuit d'arrêt d'Ankh AI nécessite l'accord d'un super-quorum (75% des veCOLLAB votant) ou du Conseil de Sécurité en cas d'urgence avérée. Cette double sécurité évite à la fois l'inaction paralysée et les suspensions abusives.

Transparence des biais

Tout modèle déployé sur Ankh AI publie un « passeport de biais » documentant : les datasets d'entraînement utilisés (avec leur provenance), les caractéristiques démographiques des annotateurs RLHF, les résultats des évaluations de biais sur benchmarks standardisés, et les limitations connues du modèle. Cette transparence, inscrite de manière immuable on-chain, permet aux utilisateurs de prendre des décisions informées et aux chercheurs d'auditer les modèles.

8.4. Éthique et inclusion

La démocratisation de l'IA ne peut se limiter à l'ouverture technique ; elle doit également s'accompagner d'une démocratisation de l'accès et des bénéfices.

Diversité des contributeurs. Ankh AI met en place des programmes actifs pour réduire les barrières à la contribution : documentation multilingue, interfaces simplifiées pour les non-techniciens, et fonds de soutien aux contributeurs des régions sous-représentées. L'objectif est que la production de l'IA reflète la diversité de ses utilisateurs finaux, évitant les biais de représentation qui affectent les modèles centralisés.

Représentation multilingue et multiculturelle. Les modèles entraînés sur Ankh AI intègrent des corpus dans plus de 50 langues, avec une attention particulière aux langues peu dotées en ressources numériques (langues africaines, langues autochtones, dialectes régionaux). Cette inclusion linguistique réduit les inégalités d'accès à l'IA entre les locuteurs de l'anglais et ceux d'autres langues.

Accessibilité économique. L'inférence sur Ankh AI est tarifée de manière compétitive (50-80% moins chère que les solutions centralisées équivalentes[37]), rendant l'IA de qualité accessible aux petites entreprises, aux chercheurs individuels, et aux organisations à but non lucratif. Un programme de crédits gratuits pour les usages éducatifs et humanitaires complète cette politique d'accessibilité.

Prévention des usages malveillants. Bien qu'Ankh AI soit résolument ouvert, des garde-fous empêchent les usages manifestement nuisibles : génération systématique de désinformation à grande échelle, création de malware, ou ingénierie sociale automatisée. Ces restrictions, décidées démocratiquement par la DAO et inscrites dans les conditions d'utilisation des nœuds d'inférence, sont implémentées via des filtres techniques et des mécanismes de vérification humaine pour les cas limites.

9. Feuille de Route (Roadmap)

Le développement d'Ankh AI s'articule autour de cinq phases successives, chacune marquée par des jalons techniques, communautaires, et économiques mesurables. Cette feuille de route est indicative et sera affinée par la gouvernance DAO à mesure de l'évolution du protocole et du marché.

9.1. Phase 0 : Fondation (Mois 0-6)

La phase de fondation pose les bases organisationnelles et techniques du protocole. Les livrables incluent : la publication de ce livre blanc et l'ouverture d'un canal de feedback public ; la constitution de l'équipe core (10-15 personnes couvrant blockchain, IA, produit, et opérations) ; le développement du testnet privé avec les smart contracts fondamentaux (ComputeRegistry, TaskScheduler, ReputationSystem) ; et la levée de fonds seed auprès d'investisseurs stratégiques en IA et Web3. Cette phase vise à démontrer la faisabilité technique et à constituer la communauté initiale de contributeurs.

9.2. Phase 1 : Testnet et Communauté (Mois 6-12)

La phase de testnet public valide l'architecture technique en conditions réelles. Les jalons incluent : le lancement du testnet public avec invitation des premiers fournisseurs de calcul et contributeurs de données ; l'entraînement d'un premier modèle de démonstration (1B paramètres) pour valider le pipeline d'entraînement distribué ; le programme d'incitation pour les early contributors (récompenses en tokens de test convertibles en mainnet) ; et les premiers audits de sécurité des smart contracts par des cabinets tiers. Cette phase vise à identifier et corriger les bugs, à calibrer les paramètres économiques, et à construire une communauté engagée de plusieurs milliers de contributeurs.

9.3. Phase 2 : Mainnet et Premier Modèle (Mois 12-18)

La phase mainnet marque le passage en production du protocole. Les événements clés sont : le lancement du mainnet avec le Token Generation Event (TGE) ; l'entraînement du premier modèle de fondation Ankh-GPT-7B (7 milliards de paramètres) sur le réseau décentralisé ; le lancement de l'API d'inférence publique avec compatibilité OpenAI ; et les premiers partenariats avec des applications consommatrices. Cette phase vise à prouver qu'un modèle compétitif peut être produit de manière décentralisée et à générer les premiers revenus du protocole.

9.4. Phase 3 : Mise à l'échelle (Mois 18-30)

La phase de mise à l'échelle amplifie significativement les capacités du réseau. Les objectifs incluent : l'entraînement d'un modèle de fondation 70B+ paramètres (Ankh-GPT-70B) compétitif avec les modèles open-source de référence ; le lancement complet du DataMarketplace avec des datasets spécialisés en médecine, code, et droit ; les premiers modèles spécialisés fine-tunés par la communauté ; et les intégrations partenaires avec des projets blockchain complémentaires (Akash pour le calcul, Ocean Protocol pour les données, des wallets pour les agents autonomes). Cette phase vise à démontrer la scalabilité du protocole et à capturer une part significative du marché de l'inférence décentralisée.

9.5. Phase 4 : Maturité (Mois 30+)

La phase de maturité consolide Ankh AI comme une infrastructure IA publique majeure. Les jalons de long terme sont : l'entraînement d'un modèle 100B+ paramètres compétitif avec les solutions centralisées commerciales ; la transition vers une gouvernance complètement décentralisée avec retrait progressif de l'équipe fondatrice ; l'abstraction de la blockchain pour les utilisateurs finaux (expérience comparable à une application Web2) ; et l'auto-suffisance économique de l'écosystème (revenus d'inférence couvrant l'ensemble des coûts opérationnels et de développement). Cette phase ultime incarne la vision d'une IA véritablement publique, détenue et gouvernée par sa communauté mondiale de contributeurs et d'utilisateurs.

Tableau 6 Feuille de route synthétique

Phase	Période	Jalon clé	Objectif
Phase 0	Mois 0-6	Testnet privé, levée seed	Faisabilité technique
Phase 1	Mois 6-12	Testnet public, modèle 1B	Validation communautaire
Phase 2	Mois 12-18	Mainnet, TGE, modèle 7B	Production & premiers revenus
Phase 3	Mois 18-30	Modèle 70B, marketplace data	Scalabilité & spécialisation
Phase 4	Mois 30+	Modèle 100B+, DAO autonome	Maturité & autosuffisance

10. Équipe et Partenaires

10.1. Auteur du Livre Blanc

Serge Destin TAMPOLLA est le rédacteur principal de ce livre blanc et conseiller stratégique (4IR Advisor) pour le projet Ankh AI. Expert en transformation digitale et technologies émergentes de la Quatrième Révolution Industrielle (4IR), il apporte une vision transversale combinant expertise technique en blockchain et intelligence artificielle, compréhension approfondie des mécanismes économiques décentralisés, et expérience en conseil stratégique pour des projets technologiques de pointe.

En tant que 4IR Advisor, Serge Destin TAMPOLLA accompagne la conception architecturale du protocole Ankh AI, la structuration de sa tokenomics, et l'alignement de sa gouvernance avec les meilleures pratiques de l'écosystème Web3. Sa contribution s'étend de la rédaction des spécifications techniques à la définition des mécanismes d'incitation économique, en passant par l'analyse des risques réglementaires et la stratégie d'adoption.

Contact : serge.destin@tampolla.com
Web : www.tampolla.com

10.2. Équipe fondatrice

La constitution de l'équipe fondatrice est en cours. Le profil type recherché pour chaque rôle clé est le suivant.

CEO / Chef de Produit. Profil hybride technologie-business avec une expérience avérée dans le lancement de produits IA ou blockchain à grande échelle. Capacité à articuler la vision technique en proposition de valeur commerciale, à lever des fonds, et à construire une communauté. Expérience souhaitée dans les deux écosystèmes (IA et Web3) ou dans l'un avec une forte appétence pour l'autre.

CTO / Architecte Blockchain. Expert en architecture de protocoles décentralisés, avec une expérience pratique du développement de smart contracts (Solidity, Rust), de la conception de mécanismes de consensus, et de la sécurité on-chain. Connaissance du Cosmos SDK et des systèmes de preuve à connaissance nulle fortement appréciée.

Chief AI Scientist. Chercheur ou ingénieur de renom en apprentissage profond, avec des contributions significatives aux architectures de modèles de langage ou aux systèmes

d'apprentissage distribué. Expérience de l'entraînement de modèles de grande taille sur infrastructure GPU distribuée. Capacité à diriger une équipe de recherche et à publier dans des conférences de premier plan.

COO / Ecosystem. Profile opérationnel avec une expérience de la construction de communautés open-source ou de réseaux de contributeurs. Compétences en gestion de communauté, relations publiques, partenariats stratégiques, et opérations internationales.

General Counsel. Juriste spécialisé dans la réglementation des crypto-actifs et de l'IA, avec une expérience des enjeux de propriété intellectuelle dans les systèmes open-source. Capacité à naviguer dans les cadres réglementaires multiples (EU AI Act, réglementations étatiques américaines, cadres asiatiques) et à structurer des entités DAO compliant.

10.3. Conseillers

Le conseil consultatif d'Ankh AI réunit des experts de quatre domaines complémentaires : **IA/ML** (chercheurs de premier plan en apprentissage profond et alignment) ; **Blockchain/Crypto** (architectes de protocoles établis, économistes tokenomics) ; **Juridique/Réglementaire** (avocats spécialisés en droit des crypto-actifs et de l'IA) ; et **Économie/Tokenomics** (chercheurs en mécanismes d'incitation, concepteurs de systèmes veTokenomics).

10.4. Partenaires stratégiques

Le succès d'Ankh AI dépend de l'établissement de partenariats stratégiques dans quatre catégories. Les **fournisseurs de calcul** (data centers professionnels, réseaux comme Akash et Render) fournissent la capacité GPU nécessaire. Les **fournisseurs de données** (institutions académiques, entreprises détentrices de corpus spécialisés, plateformes d'annotation) alimentent le pipeline d'entraînement. Les **institutions académiques** apportent l'expertise de recherche, les benchmarks indépendants, et la crédibilité scientifique. Les **projets blockchain complémentaires** (infrastructures de stockage, protocoles de confidentialité, solutions de scaling) enrichissent l'écosystème technique.

11. Analyse des Risques

Tout projet ambitieux fait face à des risques. La transparence sur ces risques et la description des mesures d'atténuation sont essentielles pour la confiance des investisseurs et des contributeurs.

11.1. Risques techniques

Scalabilité de l'entraînement distribué hétérogène. L'entraînement de modèles de fondation sur un réseau de GPU hétérogènes (différents modèles, différentes architectures, différentes latences réseau) est techniquement plus complexe que sur un cluster homogène. La mitigation repose sur le protocole DiLoCo adapté, qui tolère l'hétérogénéité via des îlots de calcul asynchrones, et sur des algorithmes de synchronisation robustes qui s'adaptent dynamiquement aux caractéristiques de chaque nœud.

Vérifiabilité à grande échelle. Vérifier cryptographiquement chaque opération de calcul dans un entraînement distribué de milliards d'opérations reste coûteux. La mitigation combine l'échantillonnage statistique (vérifier un sous-ensemble représentatif plutôt que l'intégralité), la vérification optimiste (accepter par défaut, prouver la fraude), et l'amélioration continue des preuves ZK qui deviennent plus rapides et moins coûteuses.

Latence de la coordination on-chain. La blockchain, aussi rapide soit-elle, introduit une latence par rapport à la coordination centralisée. La mitigation utilise une architecture hybride : la coordination temps réel (distribution des tâches, collecte des résultats) s'effectue off-chain via des canaux peer-to-peer, tandis que la blockchain enregistre les états finaux, les paiements, et les preuves de vérification.

11.2. Risques économiques

Volatilité du token. La valeur de $COLLAB, comme tout crypto-actif, est sujette à une volatilité potentiellement élevée. Cette volatilité peut décourager les fournisseurs de calcul qui préfèrent des revenus stables. La mitigation offre la possibilité de réceptionner les récompenses en stablecoins (USDC) avec une décote de 5%, et développe des mécanismes de couverture (hedging) via des protocoles DeFi pour les contributeurs qui le souhaitent.

Attaques économiques (Sybil, collusion). Un acteur disposant de ressources financières importantes pourrait tenter de corrompre le réseau. La mitigation combine le coût élevé du

staking, le système de réputation SBT qui ne peut pas être acheté, et la diversité géographique des nœuds qui rend la coordination d'une attaque massive logistiquement complexe.

Durabilité du modèle de récompenses. Si les revenus d'inférence ne croissent pas suffisamment vite pour couvrir les récompenses, le protocole pourrait devenir insoutenable économiquement. La mitigation réside dans la courbe d'émission décroissante (qui réduit mécaniquement les coûts de récompense au fil du temps) et dans le mécanisme de burn qui crée une boucle de rétroaction positive entre l'utilisation et la valeur du token.

11.3. Risques réglementaires

Classification du token. Les régulateurs pourraient classifier $COLLABcommeunsecuritytokenplutôtqu'unutilitytoken, soumettantleprojetàdesobligationsd'e$ COLLAB est un token d'utilité dont la valeur dérive de l'accès aux services du protocole, non d'un investissement passif. La gouvernance décentralisée via DAO renforce ce caractère utilitaire.

Régulation de l'IA. L'AI Act européen,[54] les exécutive orders américains, et les cadres réglementaires asiatiques émergents imposent des obligations de plus en plus strictes sur les systèmes d'IA. La mitigation s'appuie sur la conformité native d'Ankh AI avec les exigences de traçabilité et de transparence de l'AI Act, et sur la flexibilité de la gouvernance DAO qui permet d'adapter rapidement le protocole aux évolutions réglementaires.

Propriété intellectuelle des données. L'utilisation de datasets pour l'entraînement soulève des questions de droits d'auteur et de propriété intellectuelle. La mitigation repose sur un système de licences programmables qui enregistre explicitement les droits d'utilisation de chaque dataset, et sur des mécanismes de filtrage qui écartent les contenus manifestement protégés sans autorisation.

11.4. Risques d'adoption

Concurrence des géants centralisés. OpenAI, Google, Anthropic, et Meta disposent de ressources financières et techniques colossales. La mitigation repose sur la différenciation : Ankh AI offre une résistance à la censure, une transparence totale, des tarifs compétitifs (50-80% moins chers), et une gouvernance communautaire que les acteurs centralisés ne peuvent pas répliquer sans abandonner leur modèle économique.

Complexité utilisateur. L'interaction avec des blockchains, des wallets, et des tokens reste intimidante pour de nombreux utilisateurs. La mitigation passe par l'abstraction progressive de la complexité blockchain : paiements en carte bancaire possibles dès le lancement, interfaces

utilisateur familières, et documentation exhaustive. À long terme, l'objectif est que l'expérience utilisateur soit indiscernable de celle d'un service Web2 classique.

Problème de l'œuf et de la poule. Les fournisseurs de calcul rejoignent un réseau qui a besoin de modèles attractifs ; les modèles attractifs nécessitent des fournisseurs de calcul. La mitigation utilise un financement initial substantiel pour amorcer les deux côtés du marché simultanément, et des incitations particulièrement attractives pour les early contributors qui prennent le risque de rejoindre un réseau naissant.

12. Conclusion

Ankh AI représente une proposition de valeur fondamentalement nouvelle dans l'écosystème de l'intelligence artificielle : un protocole décentralisé qui coordonne, récompense, et gouverne la production collaborative de modèles d'IA via une blockchain dédiée. Cette vision, ambitieuse mais techniquement réalisable, s'appuie sur des innovations éprouvées — preuves ZK, veTokenomics, entraînement distribué — et sur des mécanismes originaux — vérification hybride, réputation SBT, royalties programmables.

La proposition de valeur d'Ankh AI s'adresse simultanément à trois publics distincts. Pour les **investisseurs et venture capitalists**, le protocole offre une exposition au marché de l'IA décentralisée (3,5 milliards de dollars en 2025, croissance projetée à 9,5% CAGR[49]) avec une tokenomics conçue pour l'alignement à long terme et des mécanismes de valeur déflationnistes. Pour les **contributeurs potentiels** — chercheurs, ingénieurs, fournisseurs de données, opérateurs de nœuds — Ankh AI offre une plateforme où chaque contribution est traçable, vérifiable, et rémunérée équitablement, avec des royalties perpétuelles qui transforment une contribution ponctuelle en un actif productif de long terme. Pour la **communité technique**, le protocole fournit une architecture complète, des spécifications ouvertes, et une gouvernance décentralisée qui place les décisions techniques entre les mains de ceux qui construisent le système.

Les défis sont réels mais surmontables. La scalabilité de l'entraînement distribué hétérogène, la vérifiabilité à grande échelle, et la concurrence avec des acteurs disposant de centaines de milliards de dollars de ressources représentent des obstacles significatifs. Cependant, les succès récents de projets comme Bittensor (129 sous-réseaux actifs[3]), Akash (80% d'utilisation GPU[34]), et Gensyn (protocoles de vérification ML[39]) démontrent que les blocs de construction techniques existent et fonctionnent. Ce qu'Ankh AI apporte de nouveau, c'est l'intégration de ces briques en un protocole unifié qui coordonne les trois piliers — calcul, talents, données — dans un cycle vertueux d'amélioration continue.

La vision à long terme dépasse le cadre technique et économique. Ankh AI vise à établir un précédent : que l'intelligence artificielle, cette technologie qui transforme nos sociétés, puisse être produite de manière démocratique, transparente, et équitable. Que les chercheurs du monde entier, quel que soit leur employeur, puissent contribuer aux modèles les plus avancés. Que les données de chaque communauté linguistique et culturelle puissent enrichir les systèmes d'IA au lieu d'être ignorées. Que les bénéfices de l'IA soient partagés par ceux qui la créent, pas seulement par ceux qui la contrôlent.

L'intelligence artificielle est trop importante pour être la propriété exclusive de quelques-uns. Ankh AI invite tous ceux qui partagent cette conviction — développeurs, chercheurs, investisseurs, citoyens — à rejoindre la construction de cette infrastructure publique de l'IA. Le protocole est ouvert, la gouvernance est décentralisée, et la contribution est récompensée. L'avenir de l'IA se construit ensemble.

Annexes

Annexe A. Glossaire Technique

Terme	Définition
Appchain	Blockchain dédiée à une application spécifique, construite sur un framework comme Cosmos SDK ou Substrate
DAO	Decentralized Autonomous Organization — organisation gérée par des règles codées dans des smart contracts et des votes de ses membres
DiLoCo	Distributed Low-Communication — protocole d'entraînement distribué par îlots asynchrones développé par DeepMind
FLOPS	Floating Point Operations Per Second — mesure de performance de calcul
IPFS	InterPlanetary File System — protocole de stockage et de partage de fichiers décentralisé
LLM	Large Language Model — modèle de langage de grande taille (GPT, LLaMA, etc.)
RLHF	Reinforcement Learning from Human Feedback — apprentissage par renforcement à partir du feedback humain
SBT	Soulbound Token — token non transférable lié à une identité spécifique
Slashing	Confiscation partielle ou totale d'un stake en cas de comportement malveillant
Smart Contract	Programme auto-exécutable sur une blockchain dont les règles sont immuables
TEE	Trusted Execution Environment — enclave matérielle sécurisée au sein d'un processeur
veToken	Vote-Escrowed Token — token verrouillé conférant des droits de gouvernance proportionnels à la durée de verrouillage
VRF	Verifiable Random Function — fonction de hasard vérifiable cryptographiquement
ZK-Proof	Zero-Knowledge Proof — preuve cryptographique vérifiant une assertion sans révéler les données sous-jacentes
zkML	Zero-Knowledge Machine Learning — application des preuves ZK aux modèles d'apprentissage machine

Annexe B. Spécifications Techniques Détaillées

B.1 Paramètres de la blockchain

Paramètre	Valeur
Framework	Cosmos SDK v0.50+
Consensus	Tendermint BFT (Proof of Stake)
Temps de bloc cible	~2 secondes
Débit maximal	3 000+ TPS
Finalité	Instantanée (single-block finality)
Frais de transaction moyens	< $0.001
Langage smart contracts	CosmWasm (Rust) + EVM compatible
Nombre initial de validateurs	100 (extensible)
Staking minimum validateur	10 000 $COLLAB

B.2 Paramètres tokenomics

Paramètre	Valeur
Supply totale	1 000 000 000 $COLLAB
Supply initiale au TGE	~200 000 000 $COLLAB (20%)
Taux d'émission année 1	8% de la supply restante
Décroissance annuelle	20% du taux précédent
Taux plancher	1%
Burn sur frais d'inférence	5%
Frais protocole (DAO)	10%
Durée max verrouillage veCOLLAB	4 ans
Boost max récompenses	2.5x

Annexe C. Modèles Mathématiques

C.1 Valorisation dynamique des données

La valeur d'un dataset D est déterminée par :

$$V(D) = \alpha \cdot \Delta P(D) + \beta \cdot R(D) + \gamma \cdot Dem(D) \tag{C.1}$$

avec $\alpha + \beta + \gamma = 1$ et :

- $\Delta P(D) = \frac{P_{base} - P_{with\_D}}{P_{base}}$: amélioration relative de la perplexité
- $R(D) = 1 - \cos(E_D, E_{pool})$: rareté basée sur la dissimilarité cosinus avec les datasets existants
- $Dem(D) = \frac{\sum_i w_i \cdot q_i}{\sum_i q_i}$: demande pondérée par les requêtes des modèles spécialisés

C.2 Score de réputation composite

$$Rep(U) = w_1 \cdot C(U) + w_2 \cdot E(U) + w_3 \cdot A(U) + w_4 \cdot S(U) \tag{C.2}$$

où $C(U)$ est le nombre de contributions acceptées, $E(U)$ la moyenne des évaluations par les pairs, $A(U)$ l'ancienneté normalisée, et $S(U)$ le score de spécialisation. Les poids w_i sont déterminés par gouvernance.

Annexe D. Comparaison Détaillée avec les Concurrents

Tableau D.1 Comparaison fonctionnelle détaillée

Dimension	Ankh AI	Bittensor	Gensyn	Ocean Protocol
Focus principal	3 piliers unifiés	Marketplace IA	Vérification calcul ML	Économie des données
Consensus	PoS + vérification hybride	Preuve d'Intelligence	Délégation référée	N/A (pas de consensus propre)
Vérification calcul	**ZK + Optimistic + Sampling**	Évaluation par validateurs	**Vérification probabiliste**	N/A
Réputation	SBT on-chain	Score implicite	Basique	N/A
Royalties	Programmables, perpétuelles	Non	Non	Datatokens
Gouvernance	DAO fédérée + veTokenomics	dTAO (subnet-level)	Conseil élu	DAO classique
Supply max	1 milliard	21 millions	Non spécifié	1,41 milliard
Marché cible	IA de fondation + spécialisée	Sous-réseaux IA divers	Entraînement ML	Données + compute-to-data

Annexe E. Références Académiques et Techniques

CoinMarketCap. (2026). Bittensor (TAO) — Market Data and Analysis. *coinmarketcap.com*.

CoinStats. (2026). Bittensor (TAO) — Investment Analysis May 2026. *coinstats.app*.

CoinEx Academy. (2026). Bittensor (TAO) Price Prediction 2026-2030. *coinex.com*.

Kudelski Security. (2025). ZKML: Verifiable Machine Learning using Zero-Knowledge Proof. *kudelskisecurity.com*.

Blockchain App Factory. (2025). How Ocean Protocol Marketed Decentralized Data for AI. *blockchainappfactory.com*.

ScienceDirect. (2025). Zero-knowledge machine learning models for blockchain peer-to-peer energy trading. *sciencedirect.com*.

Altcoin Buzz. (2025). Bittensor ($TAO) Deep Dive. *altcoinbuzz.io*.

EZKL Documentation. (2026). The EZKL System. *docs.ezkl.xyz*.

arXiv. (2025). A Survey of Zero-Knowledge Proof Based Verifiable Machine Learning. *arxiv.org*.

NDAX. (2026). What is Bittensor (TAO)? *ndax.io*.

Messari. (2025). Vote Escrow Explained. *messari.io*.

CoinGecko. (2023). What are veTokens and Understanding veTokenomics. *coingecko.com*.

CoinMarketCap Academy. (2023). What Is Vote Escrow? *coinmarketcap.com*.

Tally. (2026). Curve voting escrow. *docs.tally.xyz*.

CoinStats. (2026). Artificial Superintelligence Alliance (FET) — Fundamental Analysis. *coinstats.app*.

Cube Exchange. (2026). What is VeTokenomics? *cube.exchange*.

Google DeepMind. (2026). Decoupled DiLoCo: Resilient, Distributed AI Training at Scale. *deepmind.google*.

Crypto.com. (2025). What Is the Artificial Superintelligence Alliance? *crypto.com*.

Curve Finance. (2025). Why veCRV Unlocks Sustainable Tokenomics. *news.curve.finance*.

ACM Digital Library. (2025). Soulbound Token Applications: A Case Study in the Health Sector. *dl.acm.org*.

Medium. (2025). Fetch.ai & The ASI Alliance: Decentralized AI Powerhouse. *0xgreythorn.medium.com*.

SQ Magazine. (2026). OpenAI Statistics 2026. *sqmagazine.co.uk*.

Life Architect AI. (2026). GPT-5 (2025). *lifearchitect.ai*.

CryptoSlate. (2026). Superfluid — Technology. *cryptoslate.com*.

Shawn Kanungo. (2026). Anthropic vs OpenAI vs Google. *shawnkanungo.com*.

Gitcoin. (2026). Superfluid. *gitcoin.co*.

AI21 Labs. (2025). What are Trusted Execution Environments (TEE)? *ai21.com*.

Medium. (2025). The Real Cost of AI. *medium.com*.

Chainlink. (2026). What Is a Trusted Execution Environment (TEE)? *chain.link*.

Medium. (2025). Private AI at Scale: Deploying LLMs with TEEs. *medium.com*.

Alchemy. (2026). Superfluid. *alchemy.com*.

Duality Tech. (2026). TEEs & Confidential Computing. *dualitytech.com*.

Fanatical Futurist. (2025). OpenAI GPT-5 is costing $500 Million per training run. *fanaticalfuturist.com*.

Messari. (2026). State of Akash Q4 2025. *messari.io*.

Disruption Banking. (2026). Can RENDER Ride the AI Wave in 2026? *disruptionbanking.com*.

DePIN Hub. (2026). Gensyn — About. *depinhub.io*.

BlockEden. (2026). Decentralized Compute for AI: Akash, Render, and the GPU Shortage Solution. *blockeden.xyz*.

CoinPedia. (2026). Render (RNDR) Price Prediction. *coinpedia.org*.

Whales Market. (2025). What is Gensyn ($AI)? *whales.market*.

Akash Network. (2026). Akash: 2025 Year in Review. *akash.network*.

SCB 10X. (2024). Democratizing AI with Gensyn's Decentralized Compute. *scb10x.com*.

Render Network. (2026). Monthly Report — December 2025. *rendernetwork.medium.com*.

TradingView / Cointelegraph. (2025). How to use Render Network. *tradingview.com*.

Akash Network. (2025). Scaling the Supercloud. *akash.network*.

The Defiant. (2025). Akash Network Founder Unveils Plan to Scale GPU Supply. *thedefiant.io*.

Gensyn Documentation. (2026). Products & Research. *docs.gensyn.ai*.

ACM Digital Library. (2025). RLHF Deciphered: A Critical Analysis. *dl.acm.org*.

Altcoin Buzz. (2024). The Latest Guide About Akash Network. *altcoinbuzz.io*.

Intel Market Research. (2026). Decentralized AI Market Outlook 2026-2034. *intelmarketresearch.com*.

Business Research Insights. (2026). GPU Cloud Computing Market 2035. *businessresearchinsights.com*.

Nevermined. (2026). 42 Decentralized AI Payments Volume Statistics. *nevermined.ai*.

Scribd / MarketsandMarkets. (2026). Data Center GPU Market Growth to 2030. *scribd.com*.

Precedence Research. (2026). GPU as a Service Market Size. *precedenceresearch.com*.

EU AI Act. (2026). Official Updates and Compliance Information. *artificial-intelligence-act.com*.

Technavio. (2026). Data Center GPU Market Growth Analysis. *technavio.com*.

Research and Markets. (2025). Data Center GPU Strategic Research Report. *businesswire.com*.

Annexe F. Templates de Smart Contracts (Pseudo-code)

F.1 ComputeRegistry — Enregistrement et staking

```
// SPDX-License-Identifier: MIT
pragma solidity ^0.8.19;

contract ComputeRegistry {
    struct Node {
        uint256 stake;
        NodeSpec specs;
        NodeStatus status;
        uint256 reputationScore;
        uint256 registeredAt;
        uint256 totalFLOPS;
        uint256 successfulTasks;
    }

    mapping(address => Node) public nodes;
    mapping(bytes32 => address) public fingerprintToNode;

    uint256 public constant MIN_STAKE = 1000e18; // 1000 $COLLAB
    uint256 public constant MIN_VRAM = 16; // 16 GB
    uint256 public totalStaked;

    event NodeRegistered(address indexed node, bytes32 fingerprint);
    event NodeSlashed(address indexed node, uint256 amount, string reason);

    function registerNode(
        bytes32 hardwareFingerprint,
        uint256 stakeAmount,
        NodeSpec calldata specs
    ) external {
        require(stakeAmount >= MIN_STAKE, "Stake insuffisant");
        require(specs.vramGB >= MIN_VRAM, "VRAM insuffisante");
        require(nodes[msg.sender].registeredAt == 0, "Noeud deja enregistre");

        // Transfert du stake
        collateralToken.transferFrom(msg.sender, address(this), stakeAmount);

        nodes[msg.sender] = Node({
            stake: stakeAmount,
            specs: specs,
            status: NodeStatus.ACTIVE,
            reputationScore: BASE_REPUTATION,
            registeredAt: block.timestamp,
            totalFLOPS: 0,
            successfulTasks: 0
        });

        fingerprintToNode[hardwareFingerprint] = msg.sender;
        totalStaked += stakeAmount;

        emit NodeRegistered(msg.sender, hardwareFingerprint);
    }

    function slashNode(address nodeAddress, uint256 amount, string calldata reason)
```

F.2 RoyaltyEngine — Distribution automatique

```
// SPDX-License-Identifier: MIT
pragma solidity ^0.8.19;

contract RoyaltyEngine {
    struct ModelRecipe {
        address[] contributors;
        uint256[] shares; // en basis points (10000 = 100%)
        bytes32[] contributionHashes;
    }

    mapping(bytes32 => ModelRecipe) public modelRecipes;
    mapping(address => uint256) public pendingRoyalties;

    uint256 public constant ROYALTY_BASIS = 10000;
    uint256 public constant INFERENCE_ROYALTY_PCT = 1500; // 15%

    event RoyaltyDistributed(bytes32 indexed modelId, uint256 amount);
    event RoyaltyClaimed(address indexed contributor, uint256 amount);

    function registerModelRecipe(
        bytes32 modelId,
        address[] calldata contributors,
        uint256[] calldata shares,
        bytes32[] calldata contributionHashes
    ) external onlyGovernance {
        require(contributors.length == shares.length, "Longueurs incompatibles");
                require(contributors.length  ==  contributionHashes.length,  "Longueurs
incompatibles");

        uint256 totalShares = 0;
        for (uint i = 0; i < shares.length; i++) {
            totalShares += shares[i];
        }
        require(totalShares == ROYALTY_BASIS, "Total des parts != 100%");

        modelRecipes[modelId] = ModelRecipe({
            contributors: contributors,
            shares: shares,
            contributionHashes: contributionHashes
        });
    }

    function distributeInferenceFee(bytes32 modelId, uint256 totalFee)
        external onlyInferenceContract {
        ModelRecipe storage recipe = modelRecipes[modelId];
        uint256 royaltyPool = (totalFee * INFERENCE_ROYALTY_PCT) / ROYALTY_BASIS;

        for (uint i = 0; i < recipe.contributors.length; i++) {
            uint256 contributorShare = (royaltyPool * recipe.shares[i]) / ROYALTY_BASIS;
            pendingRoyalties[recipe.contributors[i]] += contributorShare;
        }

        emit RoyaltyDistributed(modelId, royaltyPool);
```

Annexe G. Scénarios de Simulation Tokenomics

G.1 Hypothèses des scénarios

Paramètre	Conservateur	Modéré	Optimiste
Volume inférence année 1 (M requêtes)	10	50	200
Croissance volume (annuel)	2x	3x	4x
Prix moyen par requête ($)	0.001	0.001	0.001
% supply staked/verrouillée	40%	50%	60%
Ratio burn/émission (année 3)	0.3	0.6	1.0

G.2 Résultats simulés

Métrique	Année 1	Année 3	Année 5
Scénario Modéré			
Revenus annuels (M$)	0.05	1.35	12.15
Tokens brûlés (M)	0.25	8.1	36.45
Supply circulante (M)	350	480	520
Capitalisation implicite (M$)	8	50	200