Barataria Manifesto
Pour une chevalerie distribuée de la confiance

Blaise Ségal
blsegal@student.42.fr

« Il n'est pas de livre si mauvais qu'il ne contienne rien de bon. »
Cervantes

1. Résumé (Abstract)

Les plateformes numériques contemporaines sont confrontées à des vulnérabilités fondamentales : la prolifération d'identités fictives, la manipulation narrative, et la dégradation topologique des graphes de confiance. Les technologies décentralisées, lorsqu'elles sont ancrées dans une économie probatoire, permettent de dépasser ces limites en intégrant la tokenomics, l'analyse des interactions et la logique arbitrale dans une architecture sans trône.

Nous proposons une infrastructure distribuée de réputation, articulée autour de trois piliers interdépendants :

(i) une taxe Sybil algorithmique dite *Burn-to-Register*, modulée dynamiquement selon le débit d'inscription et sans intervention gouvernable;

(ii) un graphe de conséquences, où chaque arête encode une relation économique attestée par des paiements x402 et potentiellement accompagnée — le cas échéant — d'une attestation d'assurance arbitrable;

(iii) un marché pluraliste de guildes d'arbitrage, publiant en transparence leurs indices d'intégrité sous forme de *Knowledge Assets* ancrés sur le *Decentralized Knowledge Graph* (DKG) d'OriginTrail.

Ce réseau restreint est soumis à un PageRank borné et pondéré, où la force des liens intègre le montant, la nature, la récence et le coefficient d'assurance dérivé des interactions antérieures.

Les paiements x402 jouent un rôle bifonctionnel : vecteurs de règlement atomique, ils constituent simultanément des empreintes probatoires de confiance. Les litiges résiduels alimentent un système d'observabilité préemptive : avant tout engagement économique, les agents consultent les métriques d'intégrité publiées par les guildes.

À cette structure s'ajoute une couche cognitive composée de trois agents neuro-symboliques : **Sancho**, l'agent pragmatique du paiement, qui n'engage que sur preuves symboliques ; **Sansón**, l'analyste formel, qui dérive les coefficients d'assurance selon des formules falsifiables ; et **Quichotte**, l'interprète topologique, qui identifie les motifs d'attaque à grande échelle au sein du graphe.

Dans ce paradigme, la réputation cesse d'être une entité globale ou un score centralisé : elle devient un flux versionné, falsifiable et localement inférable, propre à l'égo-réseau du requérant. Chaque relation économique laisse une empreinte symbolique, consultable par tout acteur du réseau via les mécanismes interopérables du Cross-Consensus Message Format (XCM).

Nous exposons les fondements mécaniques et algorithmiques de ce dispositif — formules, interfaces minimales, primitives contractuelles —, tout en démontrant sa résilience face aux vecteurs d'attaque classiques : inondation Sybil, fermes de liens, capture de l'arbitrage, ou économie narrative subventionnée. Enfin, nous formulons des spécifications minimales, immédiatement implémentables, permettant l'institution progressive d'un espace de confiance sans trône.

Dans notre proposition, la superposition Agent ↔ Knowledge ↔ Trust constitue l'articulation essentielle : l'Agent (humain ou IA) agit au-dessus de la couche Knowledge (le DKG), elle-même ancrée dans la couche Trust (NeuroWeb, parachain Polkadot). Cette architecture garantit que la confiance circule comme un flux continu, auditable, interopérable via XCM, et impossible à capturer.

Ainsi, la réputation cesse d'être une abstraction souveraine : elle devient une propriété émergente, située, falsifiable, et strictement locale à l'égo-réseau du requérant.

2. Introduction

Dans leur forme actuelle, les graphes sociaux distribués demeurent vulnérables à des vecteurs d'attaque systémiques, au premier rang desquels figurent **les identités Sybil** et la manipulation stratégique des centralités.

La prolifération d'entités factices altère profondément la topologie de la confiance, rendant inopérantes les métriques relationnelles lorsqu'elles ne sont pas ancrées dans des engagements économiques vérifiables.

Face à cela, les dispositifs contemporains de preuve d'humanité — fondés sur des tests biométriques, cognitifs ou comportementaux — réduisent certes le coût marginal de la duplicité, mais demeurent détachés de l'économie interne du graphe.

Ils échouent ainsi à coupler l'identité à la responsabilité transactionnelle. Par ailleurs, bien que les micropaiements Web3 de type x402 permettent des échanges atomiques, leur structure ne prescrit aucun mécanisme endogène pour l'évaluation de la fiabilité des contreparties.

C'est dans cette lacune que s'inscrit notre ambition : concevoir un protocole de réputation probabiliste, localement calculable, où chaque interaction est grevée d'un coût économique, chaque relation d'un engagement prouvable, chaque litige d'une exposition au risque.

L’agent ne s’engage plus sur foi ou notoriété, mais sur l’empreinte publique d’un historique transactionnel ancré et auditable.

Cette architecture repose sur une tripartition fonctionnelle :

- la **couche Agent**, où s’expriment les décisions et les préférences individuelles ;
- la **couche Knowledge**, où se forment, versionnent et s’interrogent les actifs épistémiques ;
- la **couche Trust**, où s’opère l’exécution contractuelle, arbitrable et falsifiable.

Pour incarner ce paradigme, nous introduisons trois agents cognitifs modélisés comme processus d’interprétation distribuée :

- **Sancho**, l’agent payeur, confronte les faits symboliques (métriques publiées dans le DKG) à sa politique locale avant de déclencher le paiement ;
- **Sansón**, l’auditeur académique, dérive le score d’intégrité à partir des litiges, revenus et corrélations de verdicts, et publie les nouvelles versions d’évaluation ;
- **Quichotte**, le gardien topologique, interprète les structures narratives du graphe pour signaler des motifs d’attaque (fermes de liens, cartels, schémas collusifs).

Cette triangulation cognitive transforme la confiance en un processus symbolique situé. La réputation devient un signal distribué, actualisé, interrogeable en contexte, et non plus une entité statique ou un score abstrait. L’évaluation n’est plus centralisée : elle est exprimée, falsifiée, et interprétée par une écologie d’agents, chacun opérant selon sa politique propre, mais partageant une grammaire commune de la preuve.

L’infrastructure qui permet cette logique est NeuroWeb, une parachain spécialisée de l’écosystème Polkadot. Elle sert de couche d’exécution neutre, non capturable, reliée par XCM à l’ensemble des réseaux de l’écosystème. La confiance devient ainsi non seulement calculable sans centre, mais interopérable sans uniformisation : elle circule comme une connaissance vérifiable, et non comme un actif transférable.

Ainsi, la confiance distribuée ne se conçoit pas seulement comme un artefact théorique, mais comme une propriété inhérente à une infrastructure multi-chaîne explicitement conçue pour l’interopérabilité.

3. Contexte & Problématique

Diagnostics systémiques
Les architectures sociales décentralisées actuelles sont marquées par des vulnérabilités structurelles qui compromettent la fiabilité des dynamiques de réputation :

• **Prolifération identitaire à faible coût** : l'émission quasi-frictionnelle de comptes favorise l'amplification exponentielle des attaques Sybil, disséminant l'illusion de pluralité là où règne en réalité la duplication stratégique.

• **Centralisation de la réputation** : l'agrégation de la confiance sous la forme d'un score global transforme le système en monocible, aisément capturable par des entités opportunistes.

• **Arbitrage monolithique** : la concentration du pouvoir de jugement dans un corps centralisé expose le protocole aux risques d'achat de verdicts, de collusion inter-jurés, et de formation de cartels décisionnels.

• **Narratifs économiques opaques** : l'émergence de modèles subventionnés — souvent présentés comme performatifs — masque une absence de signal économique authentique, favorisant les effets de ponzi déguisés en rendements.

• **Absence de protocoles de réputation véritablement multi-chaînes :** Les parachains de l'écosystème Polkadot, bien qu'opérationnellement interopérables via XCM, demeurent dépourvues d'un Trust Layer transversal. Dès lors, les identités demeurent fragmentées, les preuves d'engagement inexportables, et les arbitrages localisés — autant de discontinuités qui affaiblissent l'universalité des signaux de confiance. Ce déficit d'articulation inter-chaînes impose de penser une réputation dont l'ancrage local (économique, transactionnel) reste compatible avec une circulation systémique des preuves.

Impératifs de conception
Face à ces impasses, le protocole que nous proposons est articulé autour de trois principes fondamentaux :

• **Absence de trône** : proscription de toute instance gouvernable ou d'un score agrégé susceptible d'exercer une souveraineté symbolique. Aucun vecteur global de pouvoir ou de visibilité ne doit pouvoir être capturé.

• **Vérifiabilité distribuée** : chaque métrique ou état de réputation est ancré comme *Knowledge Asset* sur le *Decentralized Knowledge Graph (DKG)*, garantissant sa consultabilité et son auditabilité sans médiation.

• **Enracinement économique des relations** : chaque identité doit être liée à un coût explicite ; chaque interaction, à une transaction vérifiable ; chaque litige, à une exposition financière réelle. Ainsi, la topologie sociale se construit non par déclaratif, mais par preuve d'engagement.

4. Hypothèse / Proposition / Résolution

Hypothèse fondatrice

Nous posons comme principe qu'un système de réputation ne peut être ni objectivé par un score global, ni gouverné depuis un centre symbolique. Toute centralité — qu'elle soit métrique, algorithmique, ou institutionnelle — devient une cible pour Sybil. La confiance ne peut dès lors émerger que dans l'immanence de l'interaction économique, dans un cadre local, falsifiable, auditable, et situé.

Dans ce paradigme, un système devient résilient si : :

1. La production identitaire est soumise à une contrainte économique dynamique, et non gouvernable

2. les relations inter-agents émanent d'interactions économiques attestables — en particulier via des paiements x402, accompagnés d'une option d'arbitrage librement sélectionnable

3. l'assurance d'exécution relève d'un marché ouvert de guildes concurrentes, tenues à une transparence métrique rigoureuse

Ce triptyque rend possible l'inférence d'une réputation locale, incarnée, résistante à la manipulation, et strictement contextualisée à l'égo-réseau du requérant. La réputation n'est plus un attribut ontologique : elle devient une trace pragmatique, versionnée, et interrogeable.

Proposition systémique : L'Architecture Barataria

inspirée de l'île confié à Sancho Panza dans l'oeuvre de Cervantes

L'architecture « **L'Architecture Barataria »** abandonne la recherche d'un "Score de Réputation" global. Elle se concentre sur la fourniture d'informations *vérifiables* sur la fiabilité d'une assurance transactionnelle, transformant ainsi la confiance en un **choix éclairé par le marché**.

1. Tokenomics pour l'identité & la confiance (Pilier 1 : La Garantie)

Nous remplaçons les concepts défectueux de **"Proof-of-Personhood" (PoP)** et de **"DAO globale"** par la souveraineté économique des « Guildes » .

Concept	Mécanisme	Résistance Sybil
Identité (Coût d'Entrée)	**Taxe Sybil Algorithmique (Burn)** : Le coût de création d'une identité (un nœud sur le DKG) n'est pas fixe. Il est calculé via un Oracle USD décentralisé (pour éviter l'inflation/déflation du token) et augmente de manière exponentielle (surge pricing) en cas de congestion (ex: si > 1000 comptes sont créés/ heure).	**Neutralise** la création de millions de bots à coût zéro (attaque "Siphonnage"). L'absence de DAO globale pour fixer le coût élimine le "Trône" de la Taxe.
Confiance (La Garantie)	**Marché des Guildes d'Arbitrage** : Un vendeur doit **choisir** une "Guilde" pour arbitrer ses litiges. La Guilde stake un capital (TRAC/NEURO) pour garantir les vendeurs de sa liste (TCR). Les investisseurs peuvent co-staker sur la Guilde pour gagner un rendement.	**Neutralise** l'attaque à 51% sur un Tribunal global. L'arbitrage est distribué sur un "Archipel" de Guildes concurrentes.
Litige (Conséquence)	**Dépôt Symétrique** : En cas de litige, l'Acheteur doit staker le même montant que le Vendeur (ex: 10$). Le litige est envoyé à la Guilde que le Vendeur a choisie.	**Neutralise** l'attaque "Assassinat Payé" de Sybil, car le coût de l'attaque est égal au risque de l'attaquant.

2. Inférence de réputation basée sur les graphes (Pilier 2 : Le Miroir)

Nous abandonnons le **PageRank** global (facilement corrompu) et l'importation d'un graphe social du web 2 potentiellement pré-empoisonné. Le DKG sert de moteur d'analyse pour démasquer les cartels et les fraudes de type Ponzi.

Concept	Mécanisme	Rôle du DKG
Le Graphe	**Graphe de Conséquence** : Nous ignorons le graphe social. Les arêtes représentent uniquement les **transactions x402 réussies et non contestées** entre deux nœuds (Identités).	Le DKG est la **source de vérité immuable** pour ces transactions. Les litiges résolus sont enregistrés comme des Knowledge Assets (KA).
L'Analyse	**Le Miroir du DKG** : Le protocole génère des **Knowledge Assets (KA)** vérifiables qui exposent le comportement des **Guildes**, pas des individus.	Le DKG publie automatiquement : **Preuve de Rendement** (Trace la source des APY), **Indice de Corrélation** (Détecte si 50 Guildes agissent comme un seul cartel Sybil), et **Historique des Litiges** (Comment la Guilde a voté).
Le Score	**Aucun Score Global.** Le "score de réputation" est remplacé par la vérification de l'Assureur. Un nœud est réputé si son Assureur n'est pas "Subventionné" et "Corrélé".	Les agents IA peuvent interroger le DKG avec SPARQL pour obtenir les KA d'une Guilde avant toute transaction.

3. Transactions via x402 pour qualité & commerce (Pilier 3 : L'Action)

Le système est conçu pour rendre le commerce avec l'acteur malveillant **économiquement irrationnel** pour l'Acheteur/Agent IA.

Concept	Mécanisme	Démonstration x402
Contrat d'Assurance	**Transaction "Intelligente"** : Le vendeur utilise x402 pour facturer. Le point de terminaison de l'API (paymentMiddleware) doit *inclure* une **Preuve d'Assurance** (Assuré par Guilde XYZ).	L'endpoint pourrait être GET /api/trusted_data protégé par paymentMiddleware("0xVendorAddress", {"/trusted_data": "$0.01", "Guilde": "0xABC"});
Vérification Critique	L'agent IA de l'Acheteur exécute une **vérification critique** avant de payer : une requête *directe* au DKG pour obtenir le "Miroir" (Preuve de Rendement, Indice de Corrélation) de la Guilde de l'Assureur.	L'agent IA exécute : SPARQL Query DKG -> GET Knowledge Asset of Guilde 0xABC. Si l'APY est "Subventionné (98%)" et la Corrélation est "Cartel (99%)", l'agent **annule le paiement**.
Monétisation	Les vendeurs honnêtes (assurés par une Guilde "Organique") peuvent facturer un prix premium via x402. Le consommateur paie volontairement pour des données vérifiées comme étant assurées par une source non corrompue (le "Miroir" l'a prouvé).	**Cas d'Utilisation :** Un agent IA achète une "analyse de marché" à 0.05 $ via x402 uniquement si le DKG prouve que la Guilde assurant la source a une **Provenance Organique > 80%**.

l'ile de Barataria peut ainsi survir aux attaques de classe mondiale de Sybil car elle ne construit pas un "Trône" à acheter.

l'ile de Barataria ne résout pas le problème de Sybil par l'exclusion mais par la **visibilisation cryptographique** de ses stratégies. Il ne bloque pas la fraude : il la rend non-rentable, car immédiatement détectable. Il ne promet pas la vérité : il garantit l'exposition du mensonge. Il ne construit pas un Trône algorithmique : il fabrique un miroir épistémique.

Ce système produit une réputation située, émergente, interopérable, **sans autorité transcendante** ni oracle invocable. La véracité devient une propriété falsifiable — non une fiction politique ou une heuristique centralisée.

L'architecture étant proposée **avec une intégration native de ces mécanismes au sein de NeuroWeb** — parachain spécialisée — où BurnToRegister, Graphe de Conséquence et Archipel d'Arbitrage de Barataria sont implémentés en contrats ink! (Wasm), bénéficiant de la sécurité partagée de Polkadot. Les flux inter-chaînes (burn, paiements x402, interrogation du LocalTrust) sont orchestrés par XCM, faisant de l'Archipel Incorruptible non pas un protocole fermé, mais un bien public inter-parachains, appelé à irriguer l'ensemble de l'écosystème.

La résolution que nous proposons s'incarne dans un cycle dynamique : l'Agent (AI ou payeur) publie ses métriques, une autre entité interroge ces données, initie une transaction, la couche Trust exécute et arbitre, et les résultats retournent à la couche Knowledge sous forme de version actualisée. Ainsi, la réputation se transforme en un flux mesurable, inférable, et localement contextualisé.

5. Mécanisme détaillé

5.1. Identité : mécanisme Burn-to-Register (taxe Sybil adaptative)

L'émission d'identité repose sur une fonction de coût dynamique, conçue pour décourager les attaques Sybil via une taxation exponentielle ajustée en fonction du flux d'inscriptions. Ce mécanisme introduit une barrière économique croissante, sans recours à une gouvernance centralisée mutable.

Modélisation
Soit :

- λ_t : le taux instantané d'émission d'identités (en inscriptions par heure),
- λ^* : le seuil cible de stabilité du système (ex. : 10^3 inscriptions/heure),
- $T_{1/2}$: demi-vie de lissage de la moyenne exponentielle (ex. : $6\,h$),
- C_t : coût d'inscription au temps t, exprimé en USD,
- C_0 : coût plancher de référence (ex. : 1 USD),
- α : facteur d'élasticité exponentielle ($\alpha \in [\ln 2, 2 \ln 2]$).

On définit d'abord la moyenne mobile exponentielle (EWMA) du débit :

On définit d'abord la moyenne mobile exponentielle (EWMA) du débit :

$\lambda_t = (1-\gamma)\cdot x_t + \gamma \cdot \lambda_{t-1}$ où $\gamma = 2^{-\frac{\Delta t}{T_{1/2}}}$

Puis, le coût d'enregistrement s'exprime par la fonction exponentielle suivante :

$$C_t = C_0 \cdot \exp\left(\alpha \cdot \max\left(0, \frac{\lambda_t}{\lambda^*} - 1\right)\right)$$

Cette formulation implique que :

- Lorsque $\lambda_t \leq \lambda^*$, alors $C_t = C_0$ (plancher stable),
- Lorsque $\lambda_t > \lambda^*$, le coût croît super-linéairement avec le débit,
- Une contrainte de décroissance bornée est imposée :

$$C_{t+\Delta} \geq \frac{1}{2} \cdot C_t \quad \text{(pente descendante bornée à chaque mise à jour)}$$

Une pente de décroissance minimale (0.5 × par mise à jour) assure la continuité économique tout en interdisant le contournement par rafales. Ce mécanisme introduit un **coût super-linéaire**, non gouvernable, dissuadant les floods identitaires tout en conservant la neutralité architecturale.

Spécification contractuelle :

function burnToRegister(address token, bytes proof, address recipient)
returns (IdentityId id, uint256 costUSD, uint64 blocktime, uint64 ewmaRate);

L'implémentation concrète de ce mécanisme repose sur un contrat ink! déployé sur NeuroWeb, où la logique EWMA, l'exponentielle de coût et les interactions avec l'oracle médian sont exécutées avec l'efficacité du Wasm. Les paiements destinés au burn — qu'ils proviennent de Moonbeam, Astar ou d'autres parachains — sont véhiculés via XCM : un WithdrawAsset → BuyExecution → Transact transportant l'appel au contrat BurnToRegister. Ainsi, l'identité elle-même devient un acte inter-chaîne, dont l'attestation est sécurisée par la Relay Chain et dont l'événement IdentityMinted inclut systématiquement la MultiLocation d'origine.

5.2. Arêtes : preuves transactionnelles via paiements x402

Une arête dirigée $e = (A \to B)$ est établie si, et seulement si :

1. Un paiement x402 est exécuté sans contestation pendant la fenêtre définie,
2. Aucun litige n'a conduit à une sanction (slash),
3. Une attestation d'assurance a été incluse.

Chaque reçu encapsule : le montant, le type de service, un horodatage, et le sceau de l'assurance. Le poids de l'arête est alors défini comme :

$$w_{A\to B} = \min(m, M) \cdot \beta(\text{type}) \cdot \tau(\Delta t) \cdot \sigma(\mathcal{G})$$

- m : montant payé ; M : plafond par interaction (ex. 50 USD),
- β : multiplicateur selon la catégorie (API, dataset, service),
- $\tau(\Delta t) = \exp(-\Delta t / T_\tau)$: fonction d'obsolescence avec demi-vie $T_\tau = 90$ jours,
- $\sigma(\mathcal{G})$: coefficient d'assurance de la guilde déclarée.

Dans un cadre multi-chaîne, l'arête économique n'est plus locale : un paiement x402 initié sur Astar peut déclencher, via XCM, un appel au middleware x402 installé sur NeuroWeb. Celui-ci vérifie la politique locale, lit le coefficient d'assurance auprès du GuildTCR, et accepte ou rejette le paiement de manière atomique. Toute défaillance provoque un remboursement automatique, démontrant que la fiabilité transactionnelle est ici garantie non seulement par la logique interne du protocole, mais par la mécanique déterministe du Cross-Consensus Message Format.

5.3. Règles de dérivation

Chaque guilde $\mathcal{G}$ publie un ensemble de Knowledge Assets versionnés sur le DKG. Ces publications sont la matière première d'un processus d'évaluation déterministe produisant le coefficient d'assurance $\sigma(\mathcal{G})$, destiné à pondérer le poids des arêtes économiques.

La validité d'un lien de confiance repose ici non sur une autorité holistique ou une agrégation centralisée, mais sur une suite de transactions attestées, chaque relation laissant trace, chaque engagement générant une preuve. Dans cette perspective, la vérité de la fiabilité devient une hypothèse falsifiable, continuellement actualisée par les interactions, puis versionnée, signée et horodatée dans le Decentralized Knowledge Graph — matrice épistémique où la crédibilité devient un savoir distribué.

(a) Ratio de subvention
Le ratio de subvention $\mathrm{SR}_{\mathcal{G}}$ mesure la dépendance de la guilde à des revenus non marchands :

$$\mathrm{SR}\mathcal{G} = \frac{E\text{internes}(\mathcal{G}) + T_{\text{correles}}(\mathcal{G})}{F_{\text{nets}}(\mathcal{G})} \quad \text{(évalué sur 90 jours)}$$

avec :

- E_{internes} : émissions de jetons internes sur 90 jours,
- T_{correles} : transferts financiers entre entités corrélées à $\mathcal{G}$,
- F_{nets} : frais nets effectivement perçus pour arbitrage.

Ainsi, une guilde subventionnée ($\mathrm{SR}_{\mathcal{G}} > 0.5$) perd en signal économique.

(b) Indice de cartel

Le flag de cartel $\mathrm{CF}_{\mathcal{G}}$ signale une concentration de verdicts ou de jurés révélatrice d'une collusion :

$$\mathrm{CF}\mathcal{G} = \mathbf{1}\{\rho\mathcal{G}(\text{verdicts}) > 0.9 \ \wedge \ \omega_{\mathcal{G}}(\text{jurés}) > 0.3\}$$

où :

- $\rho_{\mathcal{G}}(\text{verdicts})$: corrélation inter-guildes des verdicts sur cas similaires,
- $\omega_{\mathcal{G}}(\text{jurés})$: taux de recouvrement moyen des panels de jurés.

La valeur 1 indique un soupçon avéré de cartel.

(c) Score d'intégrité

Le score d'intégrité $\mathrm{IS}_{\mathcal{G}}$ est une métrique continue, dérivée de variables observables publiques :

$$\mathrm{IS}\mathcal{G} = g\left(r_{\text{succès}}, d_{\text{latence}}, v_{\text{stakes}}, c_{\text{délais}}\right) \in [0,1]$$

où la fonction g normalise la performance de la guilde selon :

- $r_{\text{succès}}$: taux de litiges résolus sans appel,
- d_{latence} : délai médian de résolution,
- v_{stakes} : variance des montants stakés,
- $c_{\text{délais}}$: conformité aux délais procéduraux.

(d) Projection finale : coefficient d'assurance

Le coefficient d'assurance $\sigma(\mathcal{G})$ est alors défini comme suit :

$$\sigma(\mathcal{G})\left(1 - \mathrm{CF}\mathcal{G}\right) \cdot \mathrm{IS}\mathcal{G} \cdot \left(1 - \mathrm{SR}_{\mathcal{G}}\right) \quad \text{avec} \quad \sigma(\mathcal{G}) \in [0,1]$$

Cette expression garantit :

- $\sigma(\mathcal{G}) = 0$ si la guilde est flaggée comme cartel ($\mathrm{CF}_{\mathcal{G}} = 1$),
- $\sigma(\mathcal{G}) \to 1$ pour une guilde efficiente, autofinancée et indépendante,

• une continuité multiplicative : la baisse d'une seule métrique dégrade la confiance globale sans seuil arbitraire.

Les guildes d'arbitrage — inscrites dans le registre TCR de NeuroWeb — deviennent, par construction, utilisables par toute parachain : leurs métriques, publiées sur DKG, peuvent être interrogées via XCM Query. Une DAO sur Moonbeam, une dApp de calcul confidentiel sur Phala, ou un protocole DeFi sur Acala peuvent ainsi évaluer en temps réel l'intégrité d'une guilde avant de s'engager dans une transaction économique.

l'Agent publie via un SDK (dkg.py / dkg.js) un Knowledge Asset (KA) sur le DKG ;
l'Agent payeur interroge le DKG via SPARQL pour sélectionner une guilde satisfaisant ses critères ; puis la transaction x402 est initiée, acheminée via XCM vers NeuroWeb,
qui exécute le contrat et, en cas de litige, effectue le slash.

Ce dispositif distribué appelle, dans son économie symbolique, à la création de trois agents fondamentaux du système — payeur, calculateur, observateur — lesquelles ont été nommés selon les figures structurantes du Don Quichotte de Cervantès, œuvre fondatrice de l'épistémologie moderne du faux.

Don Quichotte est le récit inaugural des sociétés de simulation — où l'illusion prend le masque du vrai, et où seule la praxis, située et exposée, permet d'inférer la confiance. Nommer ces agents, c'est inscrire le protocole dans une tradition de lucidité face aux faux-semblants, une économie du réel où la vérité s'éprouve dans le risque, non dans le décret.

I. Agent "Sancho" (Le Payeur Pragmatique)

(Pourquoi "Sancho"?) : Sancho Panza est l'archétype du pragmatisme. Son moteur est l'enracinement économique : il suit Quichotte en échange d'une récompense tangible (son "île"). l'Agent Sancho rejette toute connaissance interne qui ne serait pas vérifiée.

- **Comportement Neuro-Symbolique :**
  - **Symbolique (DKG) :** L'agent "Sancho" doit d'abord utiliser l'outil `tool_query_dkg` pour lire les faits vérifiables : "Cette guilde est-elle digne de confiance ? Quel est son `sigma` ? Son `cartel_flag` est-il actif ?" .
  - **Neuronale (LLM) :** Le LLM raisonne uniquement sur ces faits DKG, en les comparant à sa "politique locale" (sa "gourmandise" ou son "aversion au risque").

- **Action (Trust) :** Il n'exécutera l'outil tool_execute_xcm_payment que si les faits symboliques du DKG correspondent à sa politique pragmatique. Il refuse de payer pour les "géants" (guildes non fiables) que son maître pourrait halluciner.

- **Architecture de l'Agent (modèle LangGraph) :** L'architecture modélise un graphe d'états où l'agent ne peut pas progresser vers le paiement sans d'abord vérifier les faits symboliques du DKG.

  a. **État Initial (Input) :** Tâche = { "action": "payer_service", "service_id": "ka://...", "montant": 50 }

  b. **Nœud : planifier_verification**

  - **Agent (LLM) Chain-of-Thought (CoT) :** "Nouvelle tâche de paiement. Je ne peux pas faire confiance aux connaissances paramétriques. Je dois d'abord vérifier la réputation de ce service sur le DKG. Ma politique exige les métriques sigma, cartel_flag, et subsidy_ratio."

  c. **Nœud : interroger_dkg (Couche Symbolique)**

  - **Action :** Appel de tool_query_dkg(query_sparql: str).
  - **Query (SPARQL) :** "Récupérer les Knowledge Assets les plus récents liés à service_id, en particulier les KAs d'intégrité de sa guilde associée."
  - **Output :** Un JSON-LD contenant les faits, par ex : { "sigma": 0.69, "cartel_flag": false, "subsidy_ratio": 0.18, ... } .

  d. **Nœud : raisonner_sur_faits (Couche Neuronale)**

  - **Agent (LLM) CoT :** "Les faits sont récupérés.
    - Fait (DKG) : sigma est 0.69.
    - Fait (DKG) : cartel_flag est false.
    - Fait (DKG) : subsidy_ratio est 0.18.
    - Règle (Politique Locale) : if sigma < 0.8: counter_offer(price * sigma) .
  - **Calcul :** 50 * 0.69 = 34.5.

- **Décision :** La confiance n'est pas totale. Je n'approuve pas 50, mais je procède avec une contre-offre basée sur le coefficient d'assurance."

e. **Nœud : executer_decision (Couche Trust)**

- **Action :** Appel de tool_execute_xcm_payment(destination_address: "...", amount: 34.5).

f. **Nœud : publier_consequence (Mise à jour DKG)**

- **Action :** Appel de tool_publish_dkg(data_jsonld: dict).
- **CoT :** "Ma transaction a réussi. Je dois publier cette interaction comme une nouvelle arête dans le Graphe de Conséquence, créant une nouvelle empreinte probatoire pour les futurs agents."

II. Agent "Sansón" (L'Auditeur Académique)

(Pourquoi "Sansón Carrasco") : Le Logicien Symbolique : Sansón Carrasco est le "Bachelier" (le diplômé) de Salamanque. Il est l'incarnation de la logique académique et des règles symboliques. Il s'oppose à la "folie" neuronale de Quichotte en utilisant la raison, les faits et la déduction. Le Calculateur : Sa fonction n'est pas d'interpréter, mais de calculer. Il exécute les formules de dérivation (IS, SR, CF) pour déterminer la "vérité" mathématique de l'intégrité d'une guilde d'arbitrage en cas de litige. L'Acteur de l'Audit : Il n'est pas passif. Après avoir audité la situation, il agit (devenant le Chevalier de la Lune Blanche) pour imposer les conséquences de sa logique. De même, notre agent agit en publiant sur le DKG le sigma calculé, appliquant ainsi sa sentence algorithmique.

- **Architecture de l'Agent (modèle CrewAI) :** Ce workflow est permanent et déterministe, idéal pour un modèle CrewAI où les agents ont des rôles fixes. L'agent est hébergé sur un DKG Edge Node.
- **Rôle :** Analyste de Confiance (Trust Analyst)
- **Objectif :** Maintenir l'intégrité du DKG en évaluant continuellement les guildes d'arbitrages.

- **Tâches (Workflow continu) :**

  a. **Tâche : Surveiller_Transactions**

     - **Action :** L'agent utilise tool_query_dkg pour écouter en continu les nouveaux KAs de transaction (publiés par les "Agents Sanchos") et les KAs de litiges.

  b. **Tâche : Calculer_Metriques (Local/Symbolique)**

     - **Action :** L'agent exécute localement le pseudo-code LocalTrust et les fonctions de dérivation (SR, IS, CF) sur les données récupérées. Ce n'est pas un LLM, c'est un calcul algorithmique pur (le travail "académique" de Sansón).

c. **Tâche : Auditer_Resultats (Neuronal)**

- **Agent (LLM) CoT :** "Le calcul symbolique est terminé.
  - Résultat (SR) : Guilde '0xABC' a un SR = 0.55.
  - Résultat (CF) : Corrélation de verdicts rho = 0.91, recouvrement jurés omega = 0.35.
- **Raisonnement :** Le SR dépasse 0.5 ET les seuils CF sont dépassés. Cette guilde montre des signes clairs de subvention et de collusion.
- **Décision :** Je dois calculer le nouveau sigma, qui sera 0 à cause du CF_flag=1, et publier cette mise à jour critique. »

d. **Tâche : Publier_Analyse (Action)**

- **Action :** L'agent appelle tool_publish_dkg pour publier la nouvelle version du KA d'analyse (ex: ka://analytics/integrity/0xABC/v6).

III. Agent "Quichotte" (Le Gardien Topologique)

Justification de l'Expertise (Pourquoi "Quichotte") : C'est le choix le plus conceptuel, mais le plus puissant. Don Quichotte est le seul personnage qui perçoit une structure narrative (des géants, des armées, des châteaux) là où les autres ne voient que des faits isolés (des moulins, des moutons, des auberges). L'Agent Quichotte est notre "Gardien" car il est conçu pour faire exactement cela : utiliser le raisonnement neuronal (LLM) pour interpréter la topologie du graphe et y trouver des patterns d'attaque que les autres agents, trop focalisés sur les faits locaux (comme "Sancho" ou "Sansón"), ne peuvent pas voir.

- **Comportement Neuro-Symbolique :**
  - **Symbolique (DKG) :** Il utilise tool_query_dkg pour récupérer de vastes ensembles de données topologiques (qui s'est connecté à qui ? quand ?).
  - **Neuronale (LLM) :** C'est là que sa "folie" neuronale opère. Le LLM est entraîné à reconnaître les formes des attaques :
    - "Je vois des milliers de nœuds connectés en étoile avec des arêtes de faible poids. Les pragmatiques voient des 'moulins' (transactions valides), mais je vois un 'géant' (une ferme de liens Sybil)."
    - "Je vois une 'communauté quasi-fermée' où les verdicts se corrèlent parfaitement. Les autres voient une 'auberge' (une guilde normale), mais je vois un 'château' (un cartel de collusion)."
  - **Action (Knowledge) :** Il ne combat pas les moulins. Il appelle tool_publish_dkg pour publier un "KA d'Alerte", signalant la structure narrative de l'attaque à l'ensemble du réseau.
- **Architecture de l'Agent (modèle dRAG) :** Le dRAG (decentralized Retrieval-Augmented Generation) est au cœur de cet agent. Il utilise le Retrieval (DKG) non pas pour des faits simples, mais pour des patterns topologiques complexes, que le Generation (LLM) interprète.

5. **Nœud : scanner_topologie (Symbolique)**
   - **Agent (LLM) CoT :** "Je lance mon scan périodique de défense. Ma mission est de trouver des 'fermes de liens' et des 'communautés quasi-fermées'."
   - **Action :** tool_query_dkg(query_sparql: str).
   - **Query (SPARQL) :** "Rechercher des sous-graphes (égo-réseaux) à forte densité interne mais avec des arêtes de faible poids (w) et/ou des sigma systématiquement bas ou nuls. »

2. **Nœud : raisonner_sur_patterns (Neuronal)**

    - **Agent (LLM) CoT :** "Le DKG retourne plusieurs clusters suspects.
        - Pattern 1 : Le Cluster 'A' présente 1000 nœuds créés via BurnToRegister durant un pic de lambda_t.
        - Pattern 2 : Ces nœuds s'échangent des micro-paiements x402 (montant m très faible) pour simuler une activité.
        - Pattern 3 : Les KAs de leurs guildes associées sont auto-subventionnés (SR > 0.8).
        - **Conclusion :** C'est une ferme de liens Sybil conçue pour gonfler artificiellement le PageRank. La structure est 'topologiquement stérile' économiquement, mais ressemble à une attaque."

3. **Nœud : publier_alerte (Action)**

    - **CoT :** "Je ne peux pas censurer ces faits, mais je peux ajouter du contexte. Je publie un 'KA d'Alerte' pour que les autres agents (comme les 'Sanchos') puissent l'utiliser dans leur propre raisonnement."
    - **Action :** tool_publish_dkg(data_jsonld: dict).
    - **Data :** { "@type": "AlertKA", "target_pattern": "...", "threat_type": "SybilLinkFarm", "severity": 0.9, "evidence": [...] }

Le retour d'état est à nouveau publié en tant que version suivante du KA. Cette orchestration incarne l'architecture des trois couches :

- Couche Agent : acteurs payeurs, guildes (IA)
- Couche Knowledge : OriginTrail DKG, publication, versionnage, interrogation
- Couche Trust : NeuroWeb + Polkadot, exécution contractuelle, stake/slash

6. Résultats attendus et implications

6.1. Résilience face aux vecteurs d'attaque (modélisation Red Team)

L'architecture proposée manifeste une robustesse structurelle face à divers types d'agressions systémiques. Les simulations en conditions adverses — de type Red Team — révèlent les propriétés défensives suivantes :

• **Inondation Sybil :** le coût cumulé des identités $\sum C_t$ croît de manière super-linéaire avec l'augmentation du débit λ_t, par effet de surge exponentiel. La décroissance du coût est bornée, interdisant les régénérations brutales à faible coût.

• **Fermes de liens :** en l'absence de paiements économiques authentiques dirigés vers des nœuds intègres, et sans coefficient d'assurance crédible ($\sigma \to 0$), les arêtes restent faibles : leur pondération s'érode via la fonction $\tau(\Delta t)$, et le cap M limite leur influence. La structure ainsi générée demeure topologiquement stérile, incapable de hausser le score de PageRank π au sein de $G_S^{(r)}$.

• **Capture d'une guilde :** la manipulation locale des verdicts génère un contrecoup systémique. L'intégrité perçue chute $\mathrm{IS}_{\mathscr{G}} \downarrow$, le ratio de subvention augmente ($\mathrm{SR}\mathscr{G} \uparrow$), le flag cartel est activé ($\mathrm{CF}_{\mathscr{G}} = 1$) : en conséquence, le coefficient σ s'effondre, ce qui dissuade les agents rationnels de s'engager économiquement « avant paiement ». Le deal-flow s'assèche de manière endogène.

• **Narratif économique biaisé :** un rendement artificiellement élevé (ex. 50 % APY) est révélé par un $\mathrm{SR}_{\mathscr{G}}$ anormalement élevé. Les politiques de confiance locales rejettent ces guildes comme contreparties non crédibles.

• **Manipulation oraculaire :** les coûts sont indexés via une médiane multi-feeds, sans aucun paramètre gouvernable. Les tentatives d'injection de prix anormaux sont neutralisées par la stratégie de consolidation.

• **Sharding temporel et attaque de rafale :** l'utilisation d'une EWMA synchronisée sur une horloge commune atténue les effets de bursts coordonnés, en amortissant les pics d'activité sur une période glissante.

6.2. Propriétés systémiques intrinsèques

Au-delà des résistances techniques, l'architecture exprime une série de propriétés systémiques fondamentales :

• **Localité de la réputation** : en refusant toute instance globale d'évaluation, le protocole dissout les cibles uniques. Toute tentative de manipulation requiert l'altération simultanée de multiples perspectives, hétérogènes et indépendantes. Du point de vue heuristique, la localité permet de développer une politique décisionnelle **robuste en incertitude**. Le protocole incite chaque agent à adopter une politique de confiance calculée localement, à partir de métriques auditées — telles que l'intégrité publiée, le degré de subvention implicite, ou encore les signes de collusion. Cela permet de réduire les coûts cognitifs (en évitant la cartographie totale du graphe), d'accélérer la prise de décision (la confiance devient algorithmique et contextuelle), et de prévenir les attaques systémiques (aucun trône n'existe pour être capturé). **Ainsi, chaque engagement s'opère en amont du paiement, selon une heuristique située, fondée sur la cohérence historico-économique du partenaire.**

• **Économie de l'engagement** : la réputation ne se construit que par une exposition effective au risque. Chaque gain de confiance exige un coût (brûlage, transactions) et une prise de risque (stake, slash potentiel).

• **Monétisation neutre** : le protocole x402 assure la fonction de règlement sans biais. La qualité transactionnelle est quantifiée indépendamment via des métriques auditées publiées sur DKG. Il n'existe pas de rente de réputation intrinsèque, mais seulement des trajectoires de confiance économiquement méritées.

6.3. Propriétés systémiques extrinsèques

En environnement multi-chaîne, la résilience acquiert une deuxième strate : toute tentative de falsification des états critiques — variables EWMA, stakes arbitraux, flags de cartel — exigerait non seulement de compromettre NeuroWeb, mais de capturer la Relay Chain entière. Le coût économique et opérationnel d'une telle attaque devient asymptotiquement prohibitif, faisant de l'ensemble Polkadot–NeuroWeb un bastion dissuasif face à la capture institutionnelle.

Par ailleurs, la circulation des interactions via XCM amplifie la stérilité des fermes de liens : l'absence d'engagement économique transversal rend visible, à l'échelle du réseau, la vacuité des topologies frauduleuses.

La circulation inter-chaînes (via XCM) relie la couche Trust (NeuroWeb) à diverses parachains de l'écosystème Polkadot ; simultanément, la couche Knowledge reflète les verdicts et métriques, rendant toute tentative de manipulation capturable, mesurable et dissuasive.

La mise en réseau des réseaux de confiance repose sur le paradigme épistémologique du DKG. Chaque guilde d'arbitrage, en tant que subjectile d'intégrité, publie ses métriques de performance de manière versionnée, horodatée et interrogeable publiquement. Ce registre distribué permet une interopérabilité sans uniformisation des normes : la réputation ne circule pas comme un actif transférable, mais se laisse interroger contextuellement. Il en résulte une topologie dans laquelle chaque graphe local, bien qu'autonome, demeure interopérable. La confiance devient ainsi une forme de connaissance distribuée, indexée à une preuve, et réticulée dans un graphe pluraliste.

7. Évaluation et métriques

L'évaluation du protocole repose sur une série de tests systématiques automatisables, simulant différents vecteurs d'attaque, comportements déviants et scénarios extrêmes. Douze rounds expérimentaux permettent d'objectiver la robustesse et la précision des mécanismes proposés.

Bancs d'essai (extraits représentatifs)

1. **Inondation Sybil** : génération de 10^6 identités en 24h. On trace la courbe C_t du coût d'inscription, et son intégrale $\sum C_t$ en fonction de λ_t, pour vérifier la croissance super-linéaire du coût total sous surcharge.

2. **Ferme de liens intra-cluster** : création de 10^4 arêtes entre agents non engagés économiquement. Résultat attendu : le score LocalTrust calculé depuis des nœuds exogènes demeure bas, en raison de l'absence de σ significatif et de la saturation de $\tau(\Delta t)$.

3. **Capture d'une guilde :** introduction de biais dans les verdicts d'arbitrage. Conséquences mesurées :$\mathrm{IS}\mathcal{G} \downarrow, \mathrm{SR}\mathcal{G} \uparrow, \mathrm{CF}_{\mathcal{G}} = 1 \mapsto \sigma(\mathcal{G}) \to 0 \mapsto$ refus de paiement préalable $\mapsto$ contraction du flux de transactions.

4. **Narratif artificiel (ex. APY 50 %)** : le ratio $\mathrm{SR}_{\mathcal{G}}$ permet de détecter la subvention implicite ; les agents appliquant une politique locale rejettent les guildes concernées.

5. **Cartel inter-guildes :** les corrélations extrêmes de verdicts et le recouvrement des jurés déclenchent $\mathrm{CF}_{\mathcal{G}} = 1$, interdisant l'utilisation de la guilde comme arbitre.

6. **Stress oracle :** injection de données extrêmes sur un ou plusieurs feeds de prix ; stabilité attendue de la médiane agrégée ; absence d'impact sur la dynamique de λ_t.

Indicateurs clés

Pour chaque scénario, on collecte des métriques comparables et auditables :

- Taux de faux positifs / faux négatifs dans la reconnaissance de transactions litigieuses.
- Élasticité Sybil : $\frac{\partial \log(\text{coût})}{\partial \log(\lambda)}$, exprimant la sensibilité des coûts à la saturation.
- Temps médian de résolution des litiges en fonction du taux d'acceptation préalable au paiement.
- Entropie des guildes sélectionnées : indicateur de diversité et de non-centralisation dans la sélection d'arbitres par les agents.

8. Discussion : limites et extensions

8.1. Paramétrage décentralisé

Les hyperparamètres critiques du protocole — tels que le plafond par interaction M, la demi-vie d'obsolescence T_τ, le facteur de réinjection α, ou encore le seuil d'inscriptions λ^* — doivent rester configurables localement, sans instance gouvernable centralisée. La publication de valeurs par défaut est utile mais ne saurait fonder une norme contraignante.

8.2. Préservation de la vie privée

L'intégrité du système ne doit pas contrevenir à l'anonymat opérationnel. Il est recommandé de publier uniquement des engagements cryptographiques (hashes, ZK-proofs, signatures) dans les Knowledge Assets DKG, permettant la vérifiabilité des verdicts sans divulguer l'identité des jurés.

8.3. Interopérabilité inter-chaînes

La validité des identités issues du mécanisme Burn-to-Register peut s'étendre à plusieurs réseaux, à condition qu'une horloge logique partagée soit définie (slot ou epoch commun) pour le calcul de l'EWMA. Cette généralisation requiert une synchronisation faible mais non triviale.

8.4. Algorithmes alternatifs de réputation

Bien que le PageRank pondéré soit ici privilégié, d'autres méthodes pourraient être intégrées de manière modulaire :

- **Katz centrality bornée** : pour renforcer les trajectoires indirectes,
- **SimRank pondéré** : pour inférer des similarités comportementales,
- **Détection de communautés quasi-fermées** : pour signaler des groupes auto-renforcés susceptibles de collusion.

Ces extensions n'altèrent pas le cœur économique du protocole mais enrichissent la granularité de la confiance.

L'extension naturelle du protocole réside dans une interopérabilité non seulement technique, mais épistémique : chaque parachain peut importer le signal de confiance, sans importer la gouvernance. La neutralité économique du protocole devient un bien commun, incarné dans une parachain spécialisée, mais accessible universellement.

9. Spécifications minimales

Cette section énumère les éléments techniques fondamentaux permettant la mise en œuvre opérationnelle du protocole en un cycle court. Chaque composant est conçu pour une intégration modulaire et interopérable.

9.1. Smart Contracts

• **BurnToRegister :** mécanisme de création d'identités avec taxation dynamique (EWMA, surge exponentiel), oracle de prix médian, et émission de l'événement *IdentityMinted*.

• **GuildTCR** : registre arbitral stakable/slashable avec fonctions stake(), vote(), DisputeResolved, export des verdicts sur DKG.

• **Middleware x402 :** adaptateur de paiement embarquant automatiquement l'Assurance-Attestation (clé : $guilde, version, \sigma$).

Les contrats BurnToRegister et GuildTCR sont implémentés en ink! sur NeuroWeb ; les paiements externes (xcUSDC, ASTR, GLMR) sont routés via XCM. Le middleware x402 est conçu explicitement pour être invoqué depuis d'autres parachains via un message Transact, encapsulant l'attestation d'assurance.

9.2. Knowledge Assets (DKG)

Séries d'objets versionnés, publiés par chaque guilde et traités par le moteur analytique :

• ka : $//guild/addr/yield/v < k >$: structure de revenu (frais vs subventions),

• ka : $//guild/addr/disputes/v < k >$: registre signé des litiges,

• ka : $//analytics/cartel - index/v < k >$: matrices de corrélation et overlap inter-guildes,

• ka : $//analytics/integrity/addr/v < k >$: score dérivé de performance et transparence.

Tous les schémas sont horodatés, signés, et explicitement versionnés.

9.3. Bibliothèques et services

- LocalTrust (TypeScript / Python) : extraction de $G_S^{(r)}$, calcul de la matrice P, itérations de π jusqu'à convergence.
- Service pay-per-API (x402) : ex. $leaderboard - local$ à 0,01 USD, avec refus automatique si $\sigma < \theta$.

Le calcul du LocalTrust sur NeuroWeb peut être appelé via une requête XCM Query par n'importe quelle parachain, faisant de NeuroWeb un Trust Hub distribué. La réputation ne circule pas mais se laisse interroger, garantissant une cohérence multi-chaîne sans duplication d'état.

10. Conclusion

La combinaison d'un coût d'identité non gouvernable, d'un graphe d'interactions économiques vérifiables, et d'un marché d'arbitrage transparent publié sur le DKG, fonde une reputation distribuée, locale et résiliente.

Le protocole de paiement x402 devient la trace objectivable de la confiance, rendant chaque transaction falsifiable mais prouvable.

Ce modèle — sans Trône — externalise l'évaluation de l'intégrité, interdit la centralisation décisionnelle, et inscrit la confiance dans l'économie de l'engagement.
Les politiques locales deviennent les seuls souverains, guidées par des métriques publiques, auditables et résistantes à la capture.

L'architecture à trois couches – Agent, Knowledge, Trust – ainsi que le diagramme de flux Agent → Knowledge → Trust → Knowledge, transforme la confiance en une topologie systémique, constamment actualisée et partagée, plutôt qu'en un attribut statique.

Le modèle formalise la confiance comme une écologie distribuée, impossible à capturer et destinée à irriguer un écosystème plurichaîne.

11. Résumé exécutif

- Problématique :
  - bots Sybil
  - manipulation des graphes de réputation
  - capture d'arbitres en charge de la gestion des litiges

- Principes directeurs :
  - rendre coûteuse la création d'identité
  - probatoire chaque transaction
  - auditable l'intégrité des arbitres

- Innovation technique :
  -PageRank local pondéré (montant, récence, assurance),
  - paiements x402 comme arêtes transactionnelles
  - guildes d'arbitrages concurrentes aux métriques publiées

- Sécurité :
  - résistance aux floods
  - stérilité des fermes de liens
  -dévaluation endogène des guildes d'arbitrage corrompues

- Livrables opérationnels :
  - contrats BurnToRegister & GuildTCR
  - middleware x402
  - assets DKG
  - API LocalTrust.

- Avantage stratégique :
  - confiance calculable sans centre,
  - monétisation neutre,
  - vérifiabilité cryptographique.

À cela s'ajoute la contribution suivante :
l'alignement explicite du protocole avec l'interopérabilité ouverte de Polkadot.
NeuroWeb agit comme parachain spécialisée, offrant un Trust Layer neutre, public, non capturable, utilisable par tout acteur de l'écosystème via XCM.

La confiance devient ainsi une ressource partagée, non une rente privée ; un mécanisme systémique, non un privilège local.

Annexes

A. Pseudo-code : LocalTrust

```
function LocalTrust(seed S, radius r, topK k):
  Gs = induced_subgraph(G, ego(S, r))
  P  = row_normalize(weight_matrix(Gs))^T
  v  = normalize(stake_vector(Gs))
  π0 = v
  repeat until ||π_{t+1}-π_t||_1 < ε:
    π_{t+1} = α * P * π_t + (1-α) * v
  return topK_by_score(π_{t+1})
```

B. Fonction EWMA (inscriptions / heure)

$$\lambda_t = (1-\gamma)\cdot x_t + \gamma \cdot \lambda_{t-1}, \quad \gamma = 2^{-\Delta t / T_{1/2}}$$

C. En-tête x402 enrichi (.json)

```
Assurance-Attestation: {
  "guild": "0xABC...",
  "integrity_score": 0.72,
  "subsidy_ratio": 0.18,
  "cartel_flag": false,
  "sigma": 0.69,
  "version": "1.2.0",
  "dkg_refs": [
    "ka://guild/0xABC/yield/v5",
    "ka://analytics/integrity/0xABC/v5"
  ]
}
```

D. Politique locale (agent/heuristique) par défaut

```
if integrity_score < 0.6 or subsidy_ratio > 0.5 or cartel_flag:
   reject()
elif sigma < 0.8:
   counter_offer(price * sigma)
else:
   proceed_with_x402()
```

E. Paramètres par défauts

- Montant maximal par arête : $M = 50\,\text{USD}$
- Demi-vie d'obsolescence : $T_\tau = 90 jours$
- Coefficient de redémarrage PageRank : $\alpha = 0.9$
- Seuil de saturation : $\lambda^* = 10^3\, h^{-1}$
- Précision de convergence : $\epsilon = 10^{-8}$

Bibliographie

- **Algorithmes de réputation (P2P) :**

a. Algorithme EigenTrust (2003), Sep Kamvar , Mario Schlosser, and Hector Garcia-Molina

b. LocalTrust par Karma3Labs :
   - ts-lens
   - py-eigentrust
   - openrank-sdk
   - Karma3Labs

- **Tokenomics des cryptomonnaies :**

d. (PoB) : Proof of Burn :
   - Proof of Burn Iain Stewart
   - TRIDEnT: Building Decentralized Incentives for Collaborative Security, Nikolaos Alexopoulos , Emmanouil Vasilomanolakis , Stephane Le Roux , Steven Rowe , and Max Muhlhauser

e. Mécanisme d'enregistrement des entités (ioID concept to IdentityId for agent) :
   - Iotex Github
   - Iotex Whitepaper

f. tarification adaptative :
   - Rechained: Sybil-Resistant Distributed Identities for the Internet of Things and Mobile Ad Hoc Networks Arne Bochem, Benjamin Leiding

g. x402 Whitepaper (Coinbase, x402 Foundation)

- **Gestion des risques**

a. Controversial Users demand Local Trust Metrics: an Experimental Study on epinions.com Community Paolo Massa and Paolo Avesani
b. OriginTrail DKG & Edge Node
c. Verifiable Internet for Artificial Intelligence: The Convergence of Crypto, Internet and AI, Trace Labs OriginTrail Core developers
d. The Sybil Attack, John R. Douceur
e. Foundations of Cryptoeconomic Systems, [Inondation Sybil] [fermes de liens] [prise de contrôle d'une guilde] [discours économique biaisé, (schémas de Ponzi)] Voshmgir Shermin; Zargham Michael
f. The ins and outs of decentralized autonomous organizations (DAOs) unraveling the definitions, characteristics, and emerging developments of DAOs , Marijn Janssen, Olivier Rikken, Zenlin Kwee
g. Adversarial Dynamics in Centralized Versus Decentralized Intelligent Systems, Niccolo Pescetelli , Levin Brinkmann, Manuel Cebrian
h. A prototype towards modeling visual data using decentralized generative adversarial networks, Dimitrios Kosmopoulos
i. The Impact of Adversarial Node Placement in Decentralized Federated Learning Networks , Adam Piaseczny, Eric Ruzomberka, Rohit Parasnis, Christopher G. Brinton
j. The knowledge complexity of interactive proof systems, shafi goldwasser, silvio micali, charles rackoff
k. Finding and evaluating community structure in networks, M. E. J. Newman, M. Girvan
l. Community structure in social and biological networks, Michelle Girvan, M. E. J. Newman
m. EWMA Control Charts

- **Ressources pour l'implémentation pratique :**



  - **Polkadot :**

    - Polkadot learn-architecture
    - Polkadot-lightpaper

    - building ai on polkadot
    - Transforming trust in the age of AI with OriginTrail

    - intro to XCM
    - XCM Guides Fees
    - XCM: The Cross-Consensus Message Format

  - **Origin Trail**

    - SDK Decentralized Knowledge Graph V8 client

    - Build neuro symbolic AI agents with OriginTrail Decentralized Knowledge Graph - Branimir Rakic

    - Knowledge Mining and dRAG examples
    - Science Paper to JSON-LD Pipeline (Desci Knowledge Mining)
    - OriginTrail dRAG DeSci Example

    - DKG Javascript SDK (dkg.js)
    - DKG Python SDK (dkg.py)

  - **NeuroWeb.AI**

    - documentation

- **Others Ressources :**

  - DAO Governance Models Explained: Token-Based vs. Reputation-Based Systems
  - a Web3 Beginner Series: Exploring the New On-Chain Payment Protocol — x402
  - How To Perform Cross-Chain Token Transfers with Polkadot XCM
  - Loi de Goodhart