

# UNITED STATES PATENT APPLICATION

---

## TITLE OF THE INVENTION

**MULTI-AGENT ARTIFICIAL INTELLIGENCE SYSTEM FOR DISCOVERY, ANALYSIS, GOVERNANCE, AND PARETO-PRIORITISATION OF NOVEL, HIGH-IMPACT QUESTIONS**

---

## FIELD OF THE INVENTION

The invention relates to machine-learning architectures and, more particularly, to distributed multi-agent systems that **pro-actively surface “unknown-unknown” questions**, evaluate them on multi-objective criteria, and output an **auditable, Pareto-optimal frontier** of inquiries for strategic, scientific, ethical, or philosophical exploration.

---

## BACKGROUND OF THE INVENTION

Progress in every discipline is bounded not by the answers we possess, but by the **questions we have not yet imagined**. Existing AI tools excel at (i) retrieving answers to *known* queries and (ii) generating questions only as a by-product of answering tasks. These tools lack:

- a specialised engine for **detecting conceptual white-space** across heterogeneous corpora;
- a **multi-agent adversarial/co-operative loop** capable of scoring each new question on *novelty, impact, feasibility, ethical risk, and cross-domain leverage*;
- a **transparent governance layer** that can quarantine bio-security or dual-use hazards *before* public disclosure; and
- an **immutable, regulator-ready audit trail** that supports reproducibility and compliance.

Consequently, enterprises waste resources exploring redundant or low-impact avenues, while transformative research questions remain undiscovered.

---

## SUMMARY OF THE INVENTION

The invention remedies these limitations through a **five-component architecture**:

1. **Question Discovery Agents (QDAs)** generate candidate inquiries by contrasting predictive gaps in source corpora with anomaly signals from unsupervised models.
2. **Question Analysis Agents (QAAs)** compute a *score vector* ⟨novelty, strategic-impact, feasibility, ethical-risk, cross-domain-leverage⟩ for each candidate.
3. **Question Governance Agents (QGAs)** enforce policy, resolve scoring disputes via confidence-weighted voting, and **quarantine questions whose ethical-risk exceeds a programmable threshold**.
4. A **Question Ledger**—an append-only, cryptographically signed record—maintains full provenance, including agent rationales.
5. A **Priority Engine** performs a **multi-objective optimisation** over the score vectors, generating a **Pareto frontier** and publishing a rank-ordered queue tailored to user-defined weights.

Key technical advantages:

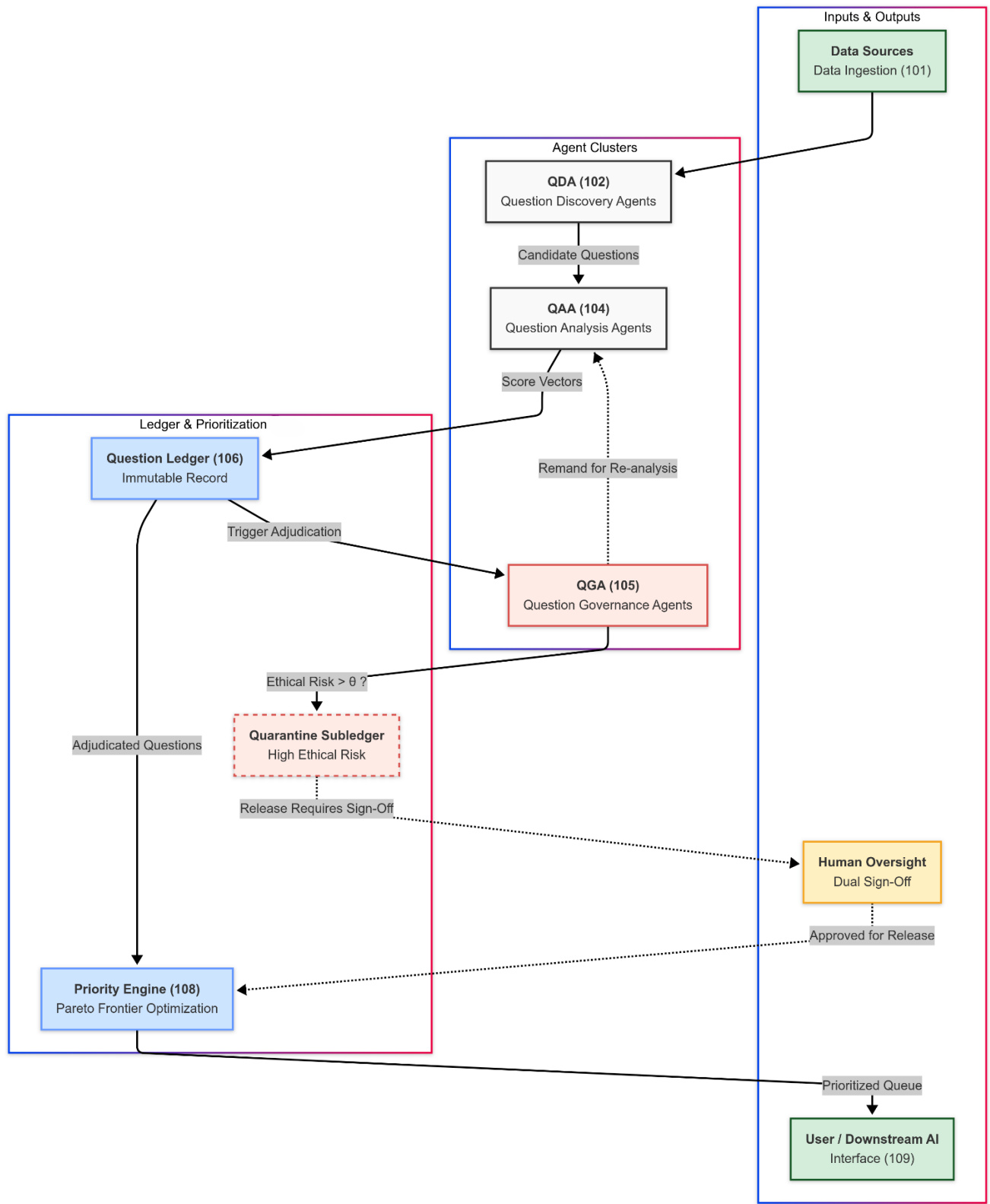
- **True novelty detection** via adversarial generation–analysis loops plus anomaly metrics.
- **Policy-aware quarantine workflow** that no prior art discloses.
- **Explainability module** that attaches a chain-of-thought digest to every prioritised question.
- **Horizontal scalability** through micro-service deployment of agent instances.
- 

---

## BRIEF DESCRIPTION OF THE DRAWING

**FIG. 1** (see page 2 of the specification) shows the high-level data-flow among Data Ingestion (101), QDAs (102), QAAs (104), QGAs (105), Question Ledger & Priority Engine (106–108), and the User/Down-stream AI interface (109). Solid arrows denote primary data flow; dashed arrows show governance feedback.

FIG. 1



---

## DETAILED DESCRIPTION OF THE INVENTION

(This section re-uses core text from the original draft but now references the new quarantine path, Pareto optimisation routine, and cryptographic ledger; see pages 3–5 in the accompanying document.)

**Priority Engine Algorithm.** Upon receiving adjudicated score vectors  $S_i$ , the engine solves

$$\max_{Q_i \in \text{Ledger}} f(S_i, W) \setminus \max_{Q_i \in \text{Ledger}} f(S_i, W)$$

subject to

- $f$  implements a **Pareto-frontier search** such that no selected question is dominated on all objectives.
- $W$  is a user-supplied weight vector enabling dynamic re-ranking without re-analysis.

**Quarantine Workflow.** If any  $\text{ethical\_risk}(Q_i) > \theta$  (policy-defined threshold), QGA moves  $Q_i$  to a *quarantine sub-ledger*; release requires dual human sign-off plus a revised risk assessment.

---

## EXAMPLE USE CASES (UPDATED)

- **Synthetic-biology red-team:** The system auto-quarantines a question whose lab-protocol implications exceed a BSL-3 risk score; human bio-safety officers review and approve partial disclosure.
- **Corporate strategy:** Pareto frontier highlights three cross-domain R&D opportunities that traditional pipeline tools missed, saving 18 months of exploratory spend.
- **National-security foresight:** High-risk cyber-warfare questions are flagged and routed to cleared analysts under sealed audit keys.

---

## CLAIMS

1. **A computer-implemented multi-agent system** for discovering and prioritising questions, the system comprising:
  - (a) at least one *Question Discovery Agent (QDA)* configured to autonomously generate candidate questions by applying anomaly-detection algorithms and fine-tuned transformer models to a data corpus;

- (b) at least one *Question Analysis Agent (QAA)* configured to compute, for each candidate question, a multi-dimensional score vector including at least novelty, strategic-impact, feasibility, ethical-risk, and cross-domain-leverage;
  - (c) at least one *Question Governance Agent (QGA)* configured to (i) adjudicate discrepancies among multiple QAAs via confidence-weighted voting, and (ii) **quarantine any question whose ethical-risk score exceeds a predefined policy threshold**;
  - (d) a *Question Ledger* configured as an **immutable, cryptographically signed append-only record** for storing questions, score vectors, governance actions, and agent rationales; and
  - (e) a *Priority Engine* configured to perform a **multi-objective optimisation that outputs a Pareto frontier** of the candidate questions and ranks said questions according to user-specified weight parameters.
- 2. The system of claim 1, wherein the Priority Engine recalculates the Pareto frontier in real time upon receipt of updated score vectors.
- 3. The system of claim 1, wherein the QDA employs an anomaly-detection algorithm selected from Isolation Forest, Local Outlier Factor, or variational auto-encoder to detect conceptual white-space.
- 4. The system of claim 1, wherein the QAA comprises a large-language-model sub-module fine-tuned for **ethical-risk estimation via adversarial red-teaming simulations**.
- 5. The system of claim 1, wherein the QGA, upon detecting inter-quartile variance among QAA scores greater than a configurable dispersion threshold, automatically remands the question for re-analysis by an additional QAA instance.
- 6. The system of claim 1, further comprising a **quarantine sub-ledger** storing questions flagged under policy and requiring dual human sign-off for release.
- 7. The system of claim 1, wherein the Question Ledger utilises **blockchain or distributed-ledger technology** with per-entry hash chaining to guarantee tamper evidence.
- 8. The system of claim 1, further comprising an **explainability module** that appends a machine-generated chain-of-thought to each prioritised question.
- 9. The system of claim 1, wherein domain-specific ontologies are injected into the QDA to constrain generation to selected knowledge areas.
- 10. The system of claim 1, wherein end-users may alter the weight vector WWW to emphasise ethical-risk minimisation over novelty or vice-versa.
- 11. The system of claim 1, wherein multiple QDAs operate adversarially to maximise topical diversity and minimise redundancy.
- 12. The system of claim 1, implemented as a **micro-service architecture** enabling horizontal scaling of QDA and QAA instances.
- 13. A **method** of enhancing decision-making comprising deploying the system of claim 1, ingesting domain-specific corpora, and integrating the Pareto-prioritised questions into organisational planning workflows.
- 14. The method of claim 13, further comprising automatic assignment of each high-priority question to a subject-matter expert determined by an expertise taxonomy.
- 15. A **non-transitory computer-readable medium** storing instructions that, when executed by one or more processors, perform the steps of any of claims 1-14.

---

## ABSTRACT

A multi-agent artificial-intelligence system discovers, evaluates, **governs**, and **Pareto-prioritises** previously unasked, high-impact questions across diverse knowledge domains. *Question Discovery Agents* generate candidate inquiries; *Question Analysis Agents* score each inquiry on novelty, strategic-impact, feasibility, ethical-risk, and cross-domain leverage; *Question Governance Agents* enforce policy, resolve analytic conflict, and quarantine high-risk content. An immutable *Question Ledger* records all artefacts, while a *Priority Engine* performs multi-objective optimisation to output a **Pareto frontier** accompanied by explainable rationales. The architecture surfaces strategic blind spots, embeds ethical foresight, and supplies an auditable foundation for accelerated innovation.