NONPROVISIONAL UTILITY PATENT APPLICATION

TITLE OF INVENTION

System for Real-Time Dynamic Governance of Emerging Technologies

INVENTORS

Andrew Vance, PhD

Taylor Rodriguez Vance, PhD

5 Union Square West, Suite 1124

New York, NY 10003

ASSIGNEE

Cyber Institute

5 Union Square West, Suite 1124

New York, NY 10003

APPLICATION DETAILS

Filed: [Date of Filing]

Application No.: [To be assigned]

Art Unit: [To be assigned]

Examiner: [To be assigned]

CROSS-REFERENCE TO RELATED APPLICATIONS

This application builds upon and incorporates subject matter disclosed in U.S. Patent Application Nos. 19/004,435 and 19/045,526. In particular, it constitutes a continuation-in-part by introducing new material not disclosed or claimed in the parent filings. These novel additions include a public-facing, multilingual conversational interface for policy interaction, real-time AI model retraining triggered by semantic drift, dual-modality risk evaluation encompassing both cybersecurity and ethical dimensions, and reinforcement learning (Q-learning) for adaptive

policy optimization. The system is further extended to support transnational governance concerns and quantum-enabled technologies, including quantum encryption compliance and quantum ethics. These features are not found in the parent applications and form the basis for independent patentability.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

Not applicable

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates generally to artificial intelligence (AI) governance, cybersecurity, and dual-use risk management for emerging technologies, particularly to a real-time, cloud-native platform that integrates AI-driven policy synthesis, cybersecurity threat simulation, and ethical risk evaluation through a vendor-neutral conversational interface. It encompasses emerging concerns such as AI-enabled cybersecurity, algorithmic ethics, quantum-enabled AI, quantum encryption compliance, and quantum ethics, ensuring that governance frameworks remain responsive to transnational risk and evolving threat vectors. The invention provides a modular, extensible architecture capable of integrating new governance models as emerging technologies evolve. Its design enables jurisdiction-specific policy evaluation, multilingual query interpretation, and adaptive retraining based on live user engagement and semantic drift, making it suitable for both national and transnational deployment.

Description of the Related Art

Emerging technologies such as artificial intelligence, quantum computing, and synthetic data engines introduce unprecedented capabilities and complex risks. These include cybersecurity vulnerabilities, policy fragmentation, algorithmic bias, misuse of generative AI, and transnational cyber risks. Existing frameworks tend to be either policy repositories or internal compliance checkers that are neither real-time nor publicly accessible. There is a growing global need for a unified, intelligent, and interactive platform that offers real-time guidance and automated policy synthesis for governments, enterprises, and individuals.

BRIEF DESCRIPTION OF DRAWINGS

- FIG. 1 – System Architecture of the Cloud-Native AI Governance System.
  Depicts the layered structure of the GUARDIAN platform, including conversational interface, AI inference engines, and policy databases.
- FIG. 2 – Policy Query Evaluation and Risk Scoring Workflow.
  Illustrates the end-to-end flow from user input through natural language processing, policy synthesis, and risk computation.
- FIG. 3 – Feedback Loop and Risk Model Evaluation.
  Shows the semantic drift detection process and trigger for retraining AI models based on user interaction divergence.
- FIG. 4 – Policy Harmonization Workflow with Real-Time Feedback.
  Details the process for integrating global policy inputs and adjusting recommendations in response to user and regulatory updates.
- FIG. 5 – Threat Simulation and Model Retraining Diagram.
  Represents how cybersecurity and ethical risk scenarios are simulated to test policy resilience and trigger learning updates.

SUMMARY OF THE INVENTION

1. The invention is a real-time AI governance platform for emerging tech risk and compliance. It provides a vendor-neutral, cloud-native governance system called GUARDIAN (Governance Using AI for Risk Detection, Integration, Analysis, and Notification). The system comprises a front-end, real-time, multilingual conversational AI interface accessible via web or third-party APIs (Application Programming Interface), and a backend inference engine powered by modular AI services for policy synthesis, threat simulation, and compliance evaluation. Real-time AI model updates are supported through hybrid data sourcing methods, including structured web scraping of authoritative databases such as those maintained by the United Nations, NIST (National Institute Standards and Technology), and the OECD (Organization for Economic Cooperation and Development), and other authoritative sources, as well as through the aggregation of user-submitted queries and feedback logs that are scored based on relevance. The policy synthesis engine employs large language models and reinforcement learning techniques to extract policy entities, harmonize

regulatory language across jurisdictions, and simulate potential compliance gaps under adversarial or stress-test scenarios.

2. The system further includes a dual-modality risk evaluation framework. The invention addresses global policy compliance's growing complexity and fragmentation and the increasing documentation of acceptable societal norms, particularly in AI, quantum, and dual-use technologies. Conventional governance tools are static, jurisdiction-specific, and lack adaptive intelligence. In contrast, GUARDIAN provides real-time, conversational access to risk evaluation and policy guidance. For example, a corporate compliance officer might use the system to determine if an AI chatbot's behavior aligns with NIST guidelines, while a government policymaker could evaluate whether a quantum encryption tool introduces dual-use compliance issues. The system fills a critical gap in scalable, intelligent governance infrastructure by tailoring outputs to user roles and using reinforcement learning to adapt to real-time feedback.
3. For cybersecurity compliance, it uses metrics and controls drawn from frameworks including the NIST Risk Management Framework (RMF), International Organization for Standardization (ISO) 27001, and General Data Protection Regulation (GDPR). Wherein the system computes a cybersecurity risk score using:

$$Risk_{cyber} = \sum_{i=1}^{n}(W_i \times V_i \times C_i)$$

where *Wi* represents the weight for each identified vulnerability type, *Vi* represents the estimated likelihood of exploitability, and Ci represents the consequence or impact severity.

4. For ethics alignment, the system analyzes AI systems for bias, transparency, and human autonomy. The ethical risk is calculated using the formula:

$$Risk_{ethics} = (1 - T) \times B \times A$$

where *T* is a transparency score (0–1), *B* is the assessed bias factor, and *A* is the autonomy or control risk index, where the scores are used to refine policy recommendations via an adaptive learning engine dynamically.

5. The conversational front end enables diverse use cases across public and institutional users, including parents assessing AI software risks for children, businesses analyzing intellectual property leakage or bias exposure, and government officials reviewing AI and quantum policy compliance for dual-use concerns.

DETAILED DESCRIPTION OF INVENTION

Overview of the System for Real-Time Dynamic Governance of Emerging Technologies

6. The GUARDIAN system is designed to provide real-time, dynamic oversight and evaluation of emerging technology policies using artificial intelligence. Referring to FIG. 1, the process begins with a user 102 interacting through a multilingual conversational AI interface 104, which processes natural language processing (NLP) queries and routes them into the cloud-native architecture. Input is parsed using NLP components 106 and compared against relevant frameworks housed in a policy repository 108. The AI/scoring engines 110 apply transformer-based models to compute domain-specific risks, which are then evaluated in a cybersecurity and ethics scoring engine 112. The cybersecurity and ethics evaluator 112 receives risk classification outputs from the AI models and applies domain-specific scoring formulas for both technical compliance and ethical alignment. The final policy output is synthesized and delivered through the summary module 114, enabling immediate user insight and feedback integration.

Process Flow

7. Referring to FIG. 2, the process begins when a user submits a policy-related query through the natural language input module 202. The request is parsed by the NLP analysis engine 204, which identifies intent, entities, and jurisdictional context. Based on this interpretation, the system proceeds directly to the policy synthesis and comparison engine 208, which extracts and aligns policy content from known frameworks. The synthesized policies are then scored by two independent modules: the $Risk_{ethics}$ ethical scoring engine 210 and the $Risk_{cyber}$ cybersecurity scoring engine 212. These modules compute respective risk values using predefined domain-specific formulas. The results are merged and passed to the recommendation generator 214, which delivers actionable guidance to the user. This workflow enables real-time, AI-driven governance by integrating risk analysis into a responsive policy recommendation pipeline. Data is anonymized and securely stored, supporting privacy and long-term model refinement. This flow enables real-time AI-supported governance, where user interactions dynamically influence the evolution of regulatory alignment tools.

Conversational Interface Module

8. The frontend of the system includes a multilingual conversational interface natural language input 202, accessible through web portals, embedded APIs, or mobile devices. As illustrated in FIG. 2, this module connects to an NLP engine 204 that includes subcomponents for intent detection, named entity recognition (NER), and sentiment classification. These subcomponents form an entity extraction and query interpretation pipeline 206, which parses user input to determine actionable context. For example, a user may input, “Is my company’s chatbot NIST compliant?” The system recognizes the topic as related to cybersecurity policy, identifies NIST as a governing framework, and activates a targeted evaluation path. This process ensures that inputs are normalized, relevant frameworks are retrieved, and the query is handed off to the policy synthesis engine for further processing. In one embodiment, extracting “policy targets” refers to the system’s ability to identify specific regulatory keywords, frameworks, or policy obligations using a combination of named entity recognition (NER), jurisdictional tagging, and semantic vector comparison. These methods allow the system to classify the user’s query into governance-relevant topics such as data privacy, algorithmic bias, or encryption standards. The result is a structured query representation that can be evaluated by the downstream policy synthesis engine 208. This supports the definition used in Claim 1 and ensures interpretability for real-time AI governance.

Policy Repository and Integration Layer

9. This module aggregates structured and unstructured policy documents into a repository from global regulatory sources, including the United Nations, OECD, NIST, and ISO. The repository is populated through structured ingestion and manual curation of global frameworks. It includes AI and cybersecurity regulations and emerging policies addressing quantum encryption standards, quantum ethics, and governance for quantum-enabled AI applications such as post-classical algorithmic control or quantum-enhanced surveillance. The documents are parsed using transformer-based models and tagged with jurisdiction, enforcement level, and sector applicability metadata. A vectorized similarity engine enables fast lookup and cross-comparison between frameworks. This is particularly important when harmonizing policy expectations between conflicting transnational regulations.

AI/ML Decision Layer

10. The decision layer consists of multiple AI models performing summarization, classification, and policy optimization. Large language models are fine-tuned to extract and normalize policy features across jurisdictions. As shown in FIG. 3, the feedback system begins with a user interaction module 302 that captures query data and system outputs. These interactions are passed to a vectorization layer 304, which transforms natural language into high-dimensional vectors for semantic comparison. A similarity engine 306 computes cosine similarity scores between new inputs and the existing policy knowledge base. If the semantic similarity falls below a defined threshold (e.g., 0.7), a drift detection component 308 flags the model for retraining. A retraining trigger module 310 initiates update routines using the newly annotated feedback corpus. The updated model is then redeployed through the model update controller 312, and outputs are re-routed back to the recommendation engine 314 to complete the adaptive feedback cycle. Graph neural networks assess regulatory maturity scores. This layered architecture enables policy recommendations to evolve dynamically through reinforcement signals, ethical risk scoring, and regulatory maturity assessment, creating a continuously learning governance engine.

Policy Optimization

11. Reinforcement learning using Q-learning allows the system to optimize recommendations over time. The Q-learning reward function is expressed as:

$$Q(s,a) = Q(s,a) + \alpha[R(s,a) + \gamma \max Q(s',a') - Q(s,a)]$$

where s is the policy state, a is the action, *alpha* is the learning rate, and *gamma* is the discount factor. $Q(s, a)$ is updated by adding the product of a learning rate α and the difference between the expected reward and the current estimate, where $Q(s, a)$ is the current Q-value for state s and action a, α is the learning rate (a value between 0 and 1), $R(s, a)$ is the immediate reward received after taking action a in state s, γ is the discount factor for future rewards, max $Q(s', a')$ is the maximum estimated future reward from the next state s' over all possible actions a'. This formula is novel in its application to real-time global policy harmonization. It enables adaptive tuning of policy suggestions based on real-world feedback, allowing the system to evolve alongside regulatory trends. It dynamically adapts to user behavior and global policy shifts, creating a closed-loop governance system that is

optimized for compliance accuracy, user trust, and responsiveness. It interacts directly with both the cybersecurity and ethics scoring modules, forming the decision backbone of the GUARDIAN architecture.

As shown in FIG. 4, the system begins with NLP components 402, which extract policy-relevant terms and user intent from natural language queries. These are passed to a policy query module 404 that normalizes the input and identifies applicable regulatory frameworks. The query results are forwarded to a synthesis engine 406, which analyzes, compares, and integrates relevant policy sources. The resulting policy synthesis 408 is evaluated against current system knowledge and user profile data. A real-time feedback module 410 captures user corrections or confirmations, which are used to retrain the model if semantic drift is detected. The final output is an updated policy recommendation 412, which reflects harmonized policy logic tailored to the query.

Cybersecurity Threat Simulator

12. The system includes a built-in simulator that models attack vectors and tests proposed policies under adversarial conditions. It calculates a quantitative risk score using:

$$Risk_{cyber} = \sum_{i=1}^{n}(W_i \times V_i \times C_i)$$

where *Wi* denotes the weight of the threat type, *Vi* the likelihood of exploit, and *Ci* the potential impact severity. This scoring method is unique in that it allows policies to be treated as dynamic, testable assets, not static documents. The formula integrates with machine learning outputs by accepting input from NLP-derived threat identification and reinforcement-learned policy scenarios. It forms a quantitative anchor that bridges natural language understanding with structured cybersecurity compliance outcomes. This allows users to see not only policy gaps but also the downstream technical risks of non-compliance.

Ethical Risk Evaluator

13. The ethical evaluator module benchmarks policies against ethical principles established by the OECD, UNESCO, and other governing bodies. Using machine learning model interpretability techniques such as SHapley Additive exPlanations (SHAP) and counterfactual fairness detection, the system computes a composite risk score:

$$Risk_{ethics} = (1 - T) \times B \times A$$

where T is a transparency index, B is a bias factor, and A is an autonomy or control-based risk. This risk score quantifies the degree to which a technology violates ethical governance principles. The formula is unique because it translates abstract ethical constructs, such as human autonomy and algorithmic fairness, into computable values. These values feed into the reinforcement learning engine, making ethics a core optimization goal alongside cybersecurity. These ethical insights complement cybersecurity assessments and provide multi-dimensional oversight, a feature not available in prior systems. This is not seen in conventional governance models.

Feedback Loop and Continuous Model Retraining

14. Referring to FIG. 5, the feedback system begins with a threat simulation module 502 that initiates a controlled scenario to assess system robustness. The output proceeds to a risk scoring module 504, which evaluates potential exposure based on predefined criteria. Next, an identified vulnerabilities module 506 isolates weak points in the system configuration or logic. These are passed to a similarity check module 508, which compares current inputs against historical patterns or known safe states. A model drift detection module 510 then determines whether significant semantic or behavioral divergence has occurred. If model drift is detected, a model retraining module 512 is activated to update the system with new data. If not, the system resumes normal operation without retraining, completing the cycle through return path 514. The similarity is computed using:

$$\text{Similarity} = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\| \|\vec{B}\|}$$

where A and B are vectorized user interaction representations, where A and B are vector representations of the user's current query and a known policy vector, respectively, A · B is the dot product of the two vectors, ||A|| and ||B|| are the Euclidean norms (magnitudes) of vectors A and B. This similarity metric enables the system to assess how closely a new user query aligns with previously known or stored policy data. By converting textual input into high-dimensional vectors using a language model embedding (e.g., from a transformer), the cosine similarity function detects semantic drift. If the similarity score falls below a predefined threshold (e.g., 0.7), it indicates that the user query represents a novel policy concern or deviation. This triggers the retraining pipeline, ensuring the AI models remain current with emerging language, regulation, and policy patterns.

This mathematical approach ensures that only significantly novel user input leads to model adjustments, preserving model stability. It also closes the loop in GUARDIAN's end-to-end feedback structure by linking user engagement to model governance refinement and, by extension, to reinforcement learning reward adjustments. This loop is critical in transforming GUARDIAN into a living system that adapts to regulatory shifts in real time.

Deployment and Scalability

15. GUARDIAN is built using containerized microservices orchestrated by Kubernetes and deployed across cloud providers. The system operates as a modular, cloud-native Governance-as-a-Service (GaaS) platform. It is deployable in multi-tenant architectures using containerized services and exposed APIs. This model allows the governance engine to be accessed by individuals, enterprises, and institutions without requiring localized infrastructure. By separating governance logic from hardware and endpoint configurations, the GaaS model supports flexible compliance use cases across jurisdictions and rapidly changing regulatory environments. Backend functions use distributed inference APIs, and data pipelines for policy ingestion operate asynchronously via a distributed event streaming platform message queues (e.g., Kafka). The platform supports multitenancy for institutions, public-sector integration, and plugin-based extensibility. Because the system is vendor-neutral, it can operate independently of proprietary cloud infrastructures, making it suitable for sovereign deployments and sensitive policy domains.
16. Policy document ingestion is managed through scheduled scraping and feed ingestion pipelines, processed via cloud-based data lakes. Vector similarity search and retraining triggers are performed using embedded indexing engines such as Facebook AI Similarity Search (FAISS) or Elasticsearch. Model drift and feedback loops are monitored by maintaining semantic embeddings of historical user queries and comparing them against current usage using cosine similarity and classification deviation scores.
17. Role-based access control, encryption at rest, and federated data anonymization layers are employed to ensure compliance with privacy regulations and international standards such as ISO 27001 and NIST 800-53. This infrastructure allows real-time interaction with minimal latency, scalable horizontally to accommodate global user bases.

Use Cases and Role-Specific Outputs

18. The system supports diverse user personas: a government analyst querying dual-use compliance of quantum tools, a parent concerned about algorithmic bias in children's apps, a corporate compliance officer checking AI chatbot transparency In each case, the system tailors outputs using role-based templates, risk prioritization, and jurisdictional tagging. This structured architecture allows GUARDIAN to go beyond static compliance checklists by actively interpreting, updating, and recommending governance actions in real time, driven by user feedback and AI adaptability.

ENABLEMENT

The foregoing description enables a person skilled in the art to practice the claimed invention by detailing the system architecture, conversational interface, backend logic, and risk scoring mechanisms. Examples of practical deployment using cloud-based infrastructure and modular services are also included to illustrate the technical feasibility of the system.

BEST MODE DISCLOSURE

The best mode contemplated by the inventors for carrying out the invention includes implementing the system using a microservices-based backend with container orchestration (e.g., Kubernetes), transformer-based NLP for policy parsing, and Q-learning-based policy optimization (as described in claim 11). Deployment is cloud-native, and training updates are triggered through real-time user interactions and semantic drift detection. The ethics and cybersecurity scoring engines operate in tandem with the Q-learning module, allowing the system to continuously refine its recommendations based on feedback-aligned policy performance in real-world queries.

CLARITY OF NOVEL FEATURES

The invention’s novelty lies in its real-time, user-facing governance interface combined with dual-risk scoring (cybersecurity and ethics), reinforcement learning for dynamic policy optimization, and semantic-drift-based retraining. These components work together to deliver adaptive, scalable, and practical AI governance for emerging technologies.

Unlike existing AI policy systems that treat compliance scoring as a one-time output, GUARDIAN integrates reinforcement learning into the ethical risk evaluation loop. When model outputs diverge from expected ethical scores, the Q-learning function adjusts its reward function over time based on real-world query outcomes. This creates a bidirectional feedback system: ethical scoring influences policy synthesis, and policy outcomes dynamically refine ethical alignment metrics. This novel configuration enables long-term drift detection and policy optimization in a way not previously described in static governance platforms.

The GUARDIAN system addresses policy gaps surrounding quantum-enabled technologies, including ethical use cases and AI control systems that leverage quantum acceleration. This includes emerging governance considerations such as transparency in quantum-enhanced decision-making and equitable access to quantum infrastructure.

This application extends and integrates subject matter disclosed in both U.S. Patent Application Nos. 19/004,435 (Quantum Policy Framework) and 19/045,526 (AI Governance Engine). Specifically, it leverages the quantum classification engine of the former to provide regulatory signal inputs, such as jurisdictional tags, compliance categories, and threat levels, that are now routed into a newly introduced ethics scoring module. This module computes real-time ethical alignment metrics across transparency, autonomy, and fairness dimensions, creating an adaptive layer not disclosed in the earlier filing. In parallel, the policy parsing and transformer-based governance scoring features introduced in the AI Governance filing have been expanded to include Q-learning–based feedback, multilingual query intent analysis, and conversational policy synthesis for transnational risks. These integrations enable a novel, real-time governance-as-a-service (GaaS) platform that processes both technical and ethical risks using dynamic learning loops and personalized role-based output generation. None of these feedback structures or ethical modeling features were described, enabled, or claimed in the earlier applications.

CLAIMS

Independent Claims

1. A computer-implemented method for real-time governance evaluation of emerging technologies, comprising:

receiving a natural language policy-related query from a user via a conversational interface;

parsing the query using a natural language processing engine to determine intent, extract regulatory targets, and identify jurisdictional context;

retrieving relevant policy content from one or more authoritative repositories;

synthesizing the policy content using a transformer-based model to produce normalized, jurisdiction-aware representations;

computing a cybersecurity risk score using a formula that multiplies weighted vulnerability types, exploit likelihoods, and consequence severity values;

computing an ethical risk score based on transparency, bias, and autonomy indices derived from the synthesized policy content;

generating a ranked policy recommendation based on the cybersecurity and ethical risk scores; and

presenting the recommendation to the user via the conversational interface.

2. A system for evaluating policy-related risks in real time, comprising:

a frontend conversational interface configured to receive and parse multilingual natural language queries;

a natural language processing module for intent detection, named entity recognition, and sentiment classification;

a policy synthesis engine comprising transformer-based models trained to harmonize content across regulatory jurisdictions;

a dual-modality scoring engine configured to compute cybersecurity and ethical risk scores using domain-specific formulas;

a Q-learning module configured to optimize policy recommendations based on historical feedback and reward functions; and

a recommendation generator configured to output actionable guidance based on current scoring and past interactions.

3. A non-transitory computer-readable medium storing instructions that, when executed by a processor, cause the system to:

receive a policy-related query via a web-based conversational interface;

process the query using a language model to extract relevant entities and jurisdictions;

retrieve policy documents from a remote or local repository;

apply risk scoring using predefined cybersecurity and ethics formulas;

evaluate model drift based on cosine similarity between current and historical query patterns; and

output an updated policy recommendation that is stored and rendered in real time through the conversational interface.

Dependent Claims

4. The method of claim 1, wherein the cybersecurity risk score is computed using the formula:

$$Risk\_cyber = \Sigma^{n}_{i=1} (Wi \times Vi \times Ci),$$

Where *Risk_cyber* is equal to the sum, from i equals 1 to n, of the product of *Wi*, *Vi*, and *Ci*, where *Wi* represents the vulnerability weight, *Vi* the likelihood of exploit, and *Ci* the consequence score for each vulnerability *i*.

5. The method of claim 1, wherein the ethical risk score is calculated using the formula:

$$Risk\_ethics = (1 - T) \times B \times A,$$

where T is a transparency score, B is a bias factor, and A is an autonomy/control risk metric.

6. The method of claim 1, wherein user interaction logs are used to retrain AI models through vectorized input scoring and cosine similarity.

7. The method of claim 1, wherein named entity recognition is used to extract governance targets from the natural language query.

8. The method of claim 1, wherein reinforcement learning is used to optimize response recommendations via Q-learning.

9. The method of claim 1, wherein a graph neural network is used for computing policy maturity scores across jurisdictions.

10. The method of claim 1, wherein the cybersecurity risk formula is dynamically adjusted using real-time threat intelligence feeds.

11. The method of claim 1, wherein the policy engine harmonizes global regulations by comparing semantic embeddings of regulatory text.

12. The method of claim 1, wherein outputs are rendered as tiered recommendations based on user trust score or access clearance.

13. The method of claim 1, wherein the conversational interface supports multilingual input with automatic language detection.

14. The system of claim 2, wherein user roles are differentiated between public users, enterprise administrators, and government agents to generate role-specific policy outputs.

15. The system of claim 2, wherein the ethics risk score incorporates SHAP values to assess model explainability.

16. The system of claim 2, wherein user interactions are logged using a differential privacy mechanism.

17. The system of claim 2, further comprising a user feedback scoring module that ranks queries for inclusion in retraining datasets.

18. The non-transitory computer-readable medium of claim 3, wherein the instructions further enable real-time policy adjustment based on live user feedback.

19. The non-transitory computer-readable medium of claim 3, wherein the similarity threshold for retraining is set to a cosine score below 0.7.

20. The non-transitory computer-readable medium of claim 3, wherein generated recommendations include references to governing bodies from which policy rules were derived.

ADVANTAGES OVER PRIOR ART

The GUARDIAN system introduces several significant advancements over existing AI governance and cybersecurity frameworks, including those described in the applicant's prior filings. First, it offers a dynamic, public-facing, multilingual conversational interface for real-time interaction with policy guidance. This represents a shift from the static internal dashboards described in earlier systems and allows engagement by a broad range of stakeholders, from individual parents to transnational policy agencies. Second, the invention integrates a dual-modality risk scoring architecture, combining cybersecurity and ethical risk assessment within a single platform. These parallel risk scores are actively used to adjust the behavior and outputs of the system's AI models. Third, GUARDIAN applies reinforcement learning, specifically Q-learning, to refine policy recommendations based on live user queries and contextual feedback. This enables adaptive optimization, which is not described in existing governance models. Fourth, the system features a novel retraining mechanism triggered by semantic drift, using cosine similarity scoring on user input vectors. This departs from traditional retraining schedules by enabling real-time, feedback-driven learning updates only when the underlying language or context has materially changed. Finally, GUARDIAN is designed as a scalable, cloud-native Governance-as-a-Service (GaaS) solution, deployable in multi-tenant environments and capable of integrating with external platforms for practical civic, corporate, and governmental use cases. These claims demonstrate novelty, non-obviousness, and technical distinctiveness sufficient for independent patentability under 35 U.S.C. §§ 101, 102, and 103.

Unlike the parent applications, this continuation-in-part introduces key novel features, including a public-facing conversational interface, real-time AI retraining triggered by semantic drift,

integrated cybersecurity and ethical risk scoring, and Q-learning-based policy refinement. These enhancements enable transnational governance and quantum-related risk analysis, none of which were disclosed or claimed in the prior filings.

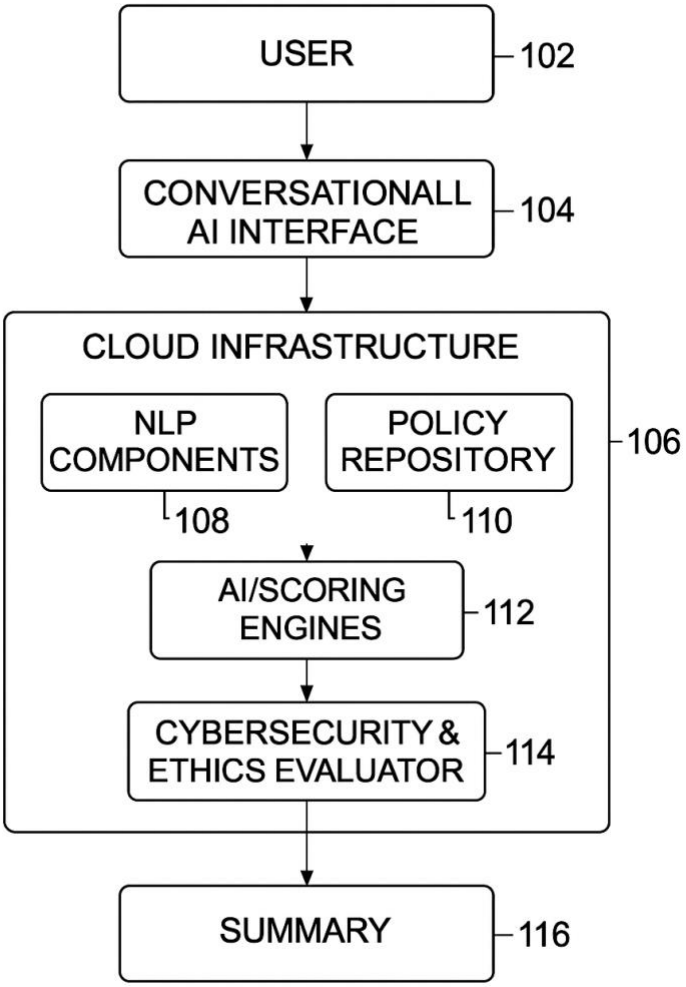
ABSTRACT OF THE DISCLOSURE

A cloud-native governance system for evaluating and optimizing policy alignment in emerging technologies using artificial intelligence. The system features a multilingual conversational interface that enables real-time risk queries from individuals, enterprises, and governments. It computes both cybersecurity and ethical risk scores using domain-specific formulas, and integrates reinforcement learning to simulate, adjust, and recommend compliant policy actions. Input is gathered from global regulatory databases and continuously updated through semantic similarity scoring and feedback-based retraining. The platform operates as a scalable Governance-as-a-Service (GaaS) solution and enables dynamic interaction with AI and quantum risk policies across diverse jurisdictions, providing actionable compliance guidance and threat simulation at scale.

DRAWINGS

FIG. 1

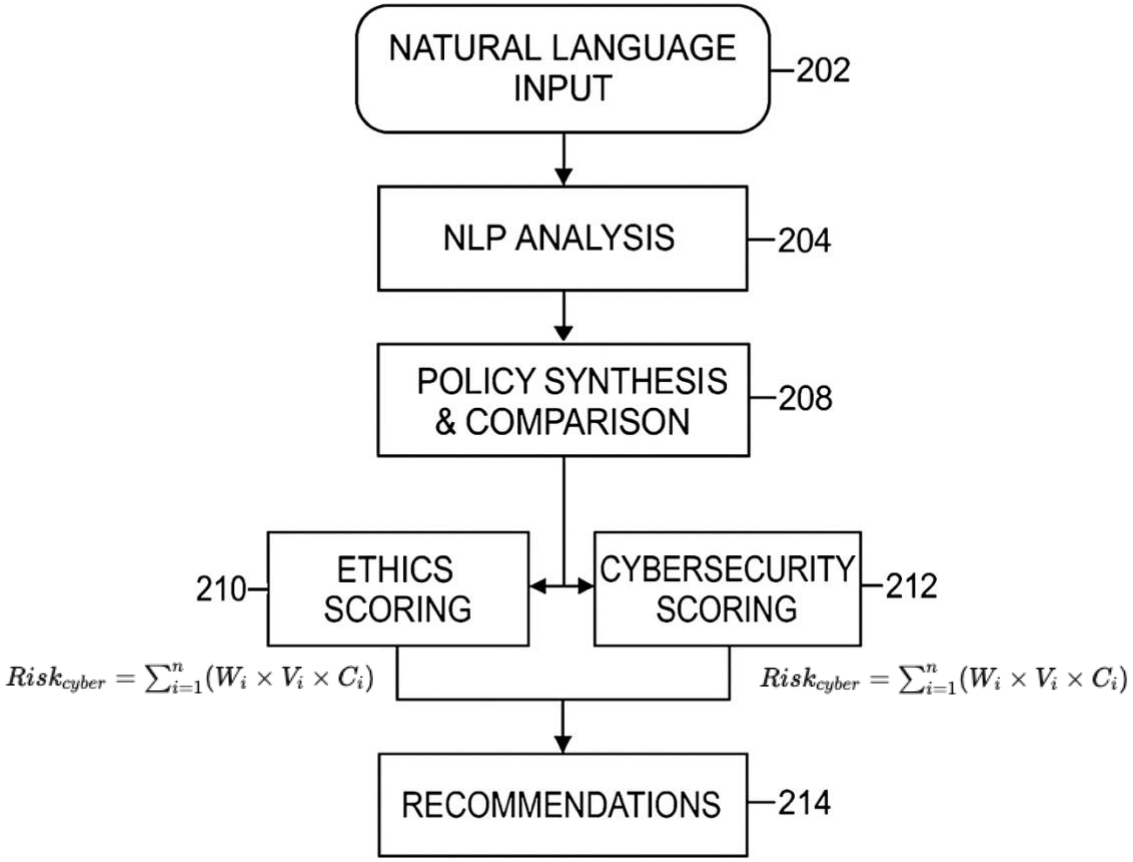
SYSTEM ARCHITECTURE OF THE CLOUD-NATIVE AI GOVERNANCE SYSTEM



DRAWINGS

FIG. 2

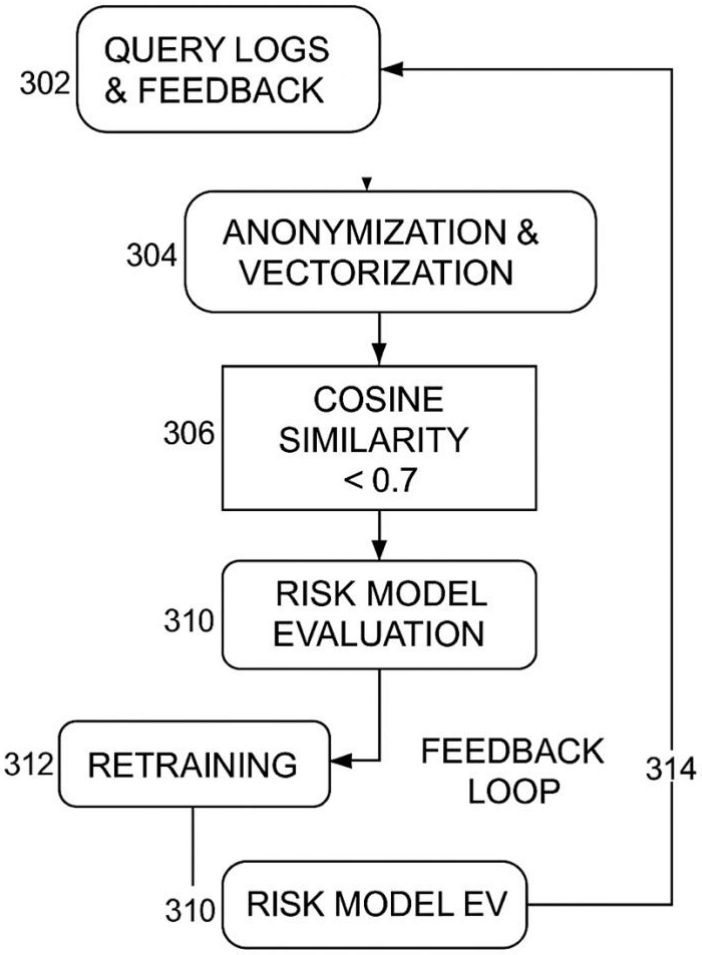
POLICY QUERY EVALUATION AND RISK SCORING WORKFLOW



DRAWINGS

FIG. 3

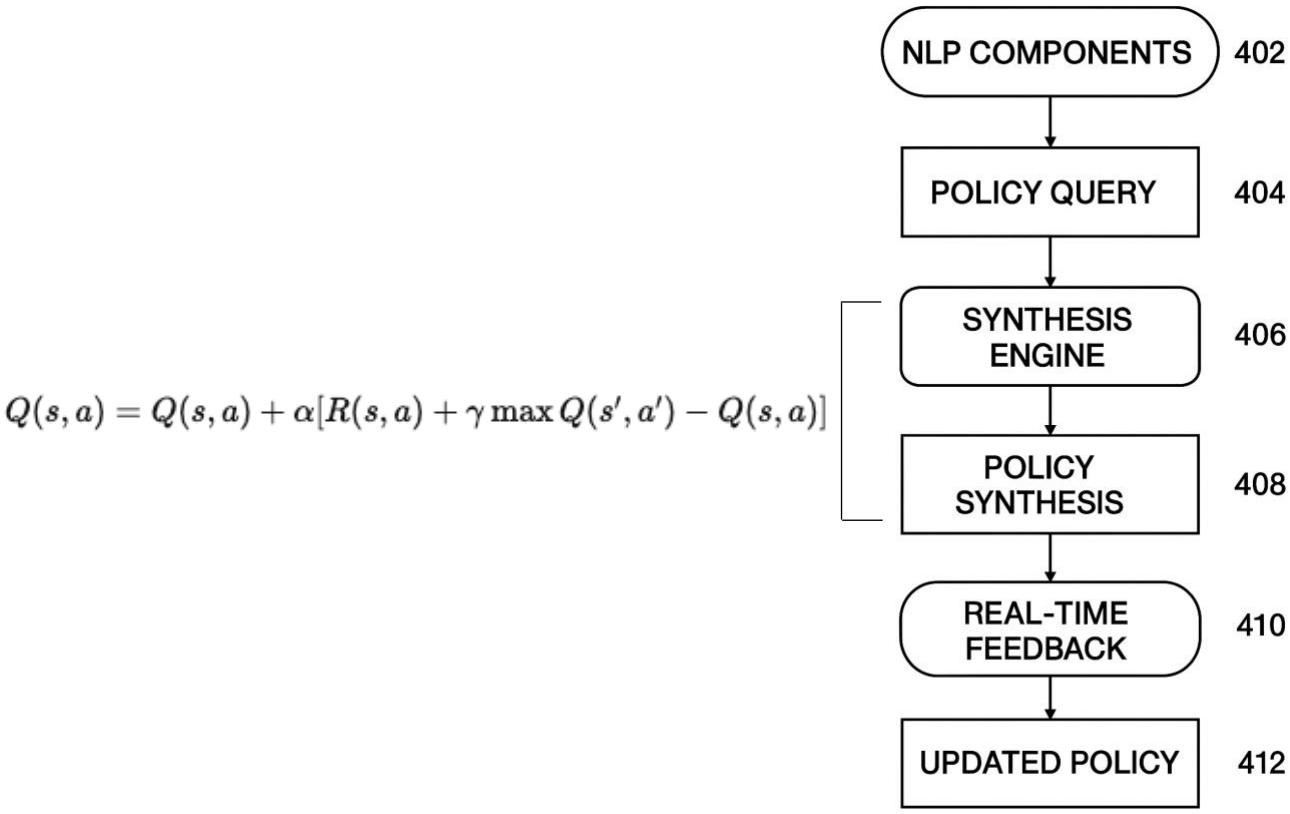
FEEDBACK LOOP AND RISK MODEL EVALUATION



DRAWINGS

FIG. 4

POLICY SYNTHESIS WORKFLOW WITH REAL-TIME FEEDBACK



DRAWINGS

FIG. 5

THREAT SIMULATION AND MODEL RETRAINING DIAGRAM