International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 4, Issue 1, January-February 2022 |

DOI:10.15662/IJEETR.2022.0401006

A Reference Architecture for AI-Driven Network Assurance in Large-Scale Enterprise Networks

Manevannan Ramasamy

Software Engineering Senior Technical Leader, Milpitas, California, United States of America

manevannan3@gmail.com

ABSTRACT: Network assurance based on AI is becoming one of the most important capabilities for ensuring reliability, performance, and service-level goals in large enterprise network environments. However, automated assurance is a challenge owing to the heterogeneous nature of telemetry, dynamic network structure, time dependencies, and dependence among causes of failures. This is because the proposed AI-based network assurance engine is reliable and comprises streaming telemetry monitoring, semantic normalization, time-aware topological modeling, feature-controlled computation, multi-method feature analysis, evidence-based diagnosis, and control remediation. Observation, inference, action, and decision are well separated in the architecture so that the model prediction cannot be a free operation command. A combination of statistical detection, supervised learning, forecasting, rule-based analysis, and topology-based reasoning detects an anomaly, approximates its impact, categorizes likely root causes, and forecasts a risk underway formation. All assurance findings are linked with supporting observations, state of topology, and time frame, confidence, and data-quality indicators with an evidence layer, contributing to explain ability and audit ability. Even graduated responses are further supported with policy-managed action layer, among more diagnostics and incident creation, and the closed-loop managed and approved remediation with blast-radium limits, verification, and rollback. The architecture proposed has the potential to offer a systematic base of scalable, explainable and operationally trustworthy network assurance besides providing dynamic model testing, drifting and control in dynamic business environments.

KEYWORDS: network assurance, artificial intelligence, AIOps, anomaly detection, root cause analysis, topology, trustworthy AI

I. INTRODUCTION

Enterprise networks have now become highly distributed and continually changing resources that can support business applications, cloud resources, geographically-distant users, Internet of Things (IoT) and wireless clients, data centers and mission-critical digital services. The larger the network and the more complicated it is, the higher the number would be to ensure that the services would be provided. The methods used in the classic network monitoring systems, such as Metric collection and generation of threshold based notifications and provision of dashboards to operators are also crucial. These facilities are noteworthy yet incapable of meeting the contemporary assurance requirements since deviation detected in metrics is not always known to signify service-level failure and to recognize a cause of the failure or corrective action suited. Network assurance therefore must take a broader perspective whereby the current state of the infrastructure is continually checked against the desired state of it regarding behavior, service objectives, policies as well as performance expectations [1].

The increasing access to network telemetry has provided great opportunities to artificial intelligence (AI)-based assurance. Enterprise infrastructures can generate large volumes of flow data in streaming telemetry, device logs, alarms, configuration changes, synthetic measurements, application events, endpoint data, and identity data. The processing of these non-homogeneous sources cannot be done manually particularly in large scale events whereby a combination of a large number of symptoms may be taking place at the same time. In order to help, AI and machine learning can be used to detect anomalous behavior, and map global observations on a global basis on different infrastructure levels, predict potential failures, prioritize likely root causes, and offer suitable remediation. Nonetheless, translating AI into the direct application in the operation network control provides significant architectural and governance issues. Even an extremely accurate model will give unsafe recommendations in the case of missing input data, assumptions being broken or even the network state has changed since training the model [2] [3].

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 4, Issue 1, January-February 2022 |

DOI:10.15662/IJEETR.2022.0401006

One of the major issues arising is thus the difference between network monitoring and network assurance. The most important queries to answer with the help of the monitoring are what has grown or gone down, what device has raised an alarm, and was a threshold reached.

The issue is also complex by the network topology. Physical, logical, service and dependency relationship exist between routers, switches, wireless controllers, access point, firewalls, server, application, users and cloud services of enterprise infrastructures. When one component fails, this may be passed onto other components via these connections and create symptoms at other downstream locations. New designs that ensure that telemetry is the independent feature vector can therefore lose the critical context dependent information. Topology and dependency information may be also added and the system will give preference to the main failures and secondary symptoms and make additional efforts into root-cause analysis.

Another basic need happens to be the temporal consistency. Network conditions, configuration, topology relations and telemetry are subjected to constant changes. The diagnosis based on a current snapshot of a topology can be wrong when a failure happened prior to a configuration or routing update. Likewise, non-real-time telemetry will develop spurious causality. There should then be some powerful assurance mechanism that stores event time, collection time, configuration versions and topology versions to retain analytical conclusions and network state that will have prevailed at the time that the respective events take place.

Such as explainability, model drifting, confidence calibration and operational safety issues of using AI are also there. The fact that AI forecasts are created does not mean that it is supposed to be regarded as an authorization to modify production facilities. Assurance can be used as an alternative and divide observation, inference, decision and action into controlled architectural phases. Model results ought to be converted into evidence-based discoveries with the impacted scope, recommendations, timeline and topology ties, confidence, uncertainty, and information regarding quality data. Then we could perhaps have a policy and decision layer that can assert what it thinks is right to do is an addition of data, running some diagnostic code, creating an incident, making an incident approved by a human, or some automatic actions to rectify the situation.

What AI is not is that it is a system, that it can be seen as being a component of the overall system of assurance. The architecture should offer mechanisms to create, trace, assess and offer reversible automation. Models can not only be evaluated on the basis of whether they detect the anomalies: it is also advisable to evaluate the models based on the quality of diagnosis, religious to explanation, false-positive or drift, and operational and remediation safety.

I have therefore provided reference architecture of AI based network assurance to large enterprise network in this paper. It consists of architecture of heterogeneous telemetry acquisition, semantic normalization, and temporal topology representation, controlled computation of features, evidence-based diagnosis, policy-controlled decision making, and safe remediation. The further development of the conceptual design will offer a scalable and dependable base on which AI might be put to the task of enhancing detection and diagnosis and maintaining the accountability of operations, human control and controlled automation.

II. ARCHITECTURAL PRINCIPLES

Operational control and analytical intelligence are separated by a distinct boundary of the proposed AI-driven network assurance structure which is directed by its principles. These tenets are to ensure that the assurance decisions made are context sensitive, have explanations, auditable and capable of scaling to large basis enterprise arrangements.

2.1 Separation of Observation, Inference, Decision, and Action

It consists of four logically distinct steps that are observation, inference, decision and action. During observation, telemetry and contextual data is gathered, and anomalies are detected using inference, analytical and analytical AI models. A change of configuration should not actually be directly caused by the result of the inference. Rather, a decision layer simply reflects on a model confident, quality of evidence, operational policy, scope of the area of effect and authorization. Indicatively, an AI model can detect a wireless controller as a likely cause of massive client failures, but the policy engine will decide following the adequacy of the evidence to begin diagnostics or install an incident, enlist human approval or an approved remediation. The impact of this division is supposed to ensure the operation power of prediction confidence is not boundless.

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 4, Issue 1, January-February 2022 |

DOI:10.15662/IJEETR.2022.0401006

2.2 Topology and Dependency Awareness

Network assurance is supposed to make sure that structural relationships between the infrastructure components and services are not changed. Enterprise networks are made up of physical connection, logical route, routing dependency, security dependency, application and service dependency. Failure of one component thus can result in a number of secondary symptoms. The independence of the telemetry vectors as numbers may cause the loss of these associations and the inferences made on root causes will be in mistaken beliefs. The architecture thus has a topology-sensitive representation that has the capacity to associate devices/interfaces, clients/applications, services and policies. The results of analytics can be therefore contrasted with their actual relationship of connectivity and dependency with each other rather than just statistical relationship.

2.3 Temporal Integrity and State Consistency

The networks which dominated in a particular time when an event was witnessed as indicated by the networks should determine assurance decision. Topology can be changed; configuration of the policies, measurements of the telemetry, as well as conditions of the services can be updated in less than a few minutes or seconds. The architecture is thus retaining event time, collection time, and configuration and topology snapshots. It is also free of the erroneous perception of past event by the use of the present state of topology. Root-cause analysis, reconstruction of incident, model training, and post-incident auditing are particularly useful in terms of temporal integrity.

2.4 Governed and Trustworthy Learning

AI network assurance models based on algorithms must have distinguished governance in their lifecycle. According to the NIST AI Risk Management Framework, governance, mapping, measurement, and management are complementary tasks that could be used to manage AI-related risks [3]. Each model is supposed to find its purpose in the proposed architecture, the detailed description of the scope of the training-data, the known limitations and identified failure-modes and performance attributes. Where predictions are based on confidence calibration and estimation of uncertainty, these should be provided with these confidence calculations and uncertainty estimates. The model drift needs to be also constantly checked because over time, network behavior changes, applications, population of users and infrastructure environments. Human handling is necessary on such high impact decisions, and extraordinary cases.

2.5 Evidence-Based Explainability

A probability or anomaly score should not be the only type of assurance that is provided. This means there will be a need to correlate analytical conclusion in the architecture with supporting evidences where there will be relevant telemetry, influenced entities, time dependencies, topology dependencies, data-quality indicators, and counter pointing notes. This can help the operators know why such a high rating of particular cause is noted and whether the conclusion is adequate to be used in operation. Preservation of evidence will also assist in reviewing the incident and validating the models used compliance, and continuous improvement. Explainability is then said to be a property of architecture and not a non-obligatory presentation property.

2.6 Reversible and Policy-Controlled Automation

Neither is the growth of automation to grow automatically according to confidence or evidence and operation risk. The next round of data gathering could be due to low-confidence results, whereas tickets or diagnostic advice could be obtained due to moderate-confidence results. Findings with high confidence should be predetermined with policies and automated remediation is to be provided only. It should have checks of authorization, preconditioning, blast-radar limits, rate, and post-action checks/rollback. This type of automation model can be undone and allows the AI to accelerate the reaction of the operations without an imprecise forecast causing restrained adjustments across the enterprise network.

An assemblage of these values creates a platform on which AI can improve the insights of the net, diagnosis, and response without compromising on the context, responsibility, security and the control of humans.

III. REFERENCE ARCHITECTURE: A TELEMETRY-TO-ACTION ASSURANCE FABRIC

The suggested reference architecture will provide an end-to-end promise pipeline that will map dataphorically disparate network measurements into contextualized manifestations that hold significance in functioning, evidence-driven diagnostics and systematically controlled functions. The architecture gathers together data, topology, models, evidence, and automation into coordinated layers, instead of considering AI as a distinct component of analytics. This structure

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 4, Issue 1, January-February 2022 |

DOI:10.15662/IJEETR.2022.0401006

allows a guarantee to last longer in the huge networks of businesses and not to disrupt the traceability between an operation event and the ultimate decision taken on the basis of the event.

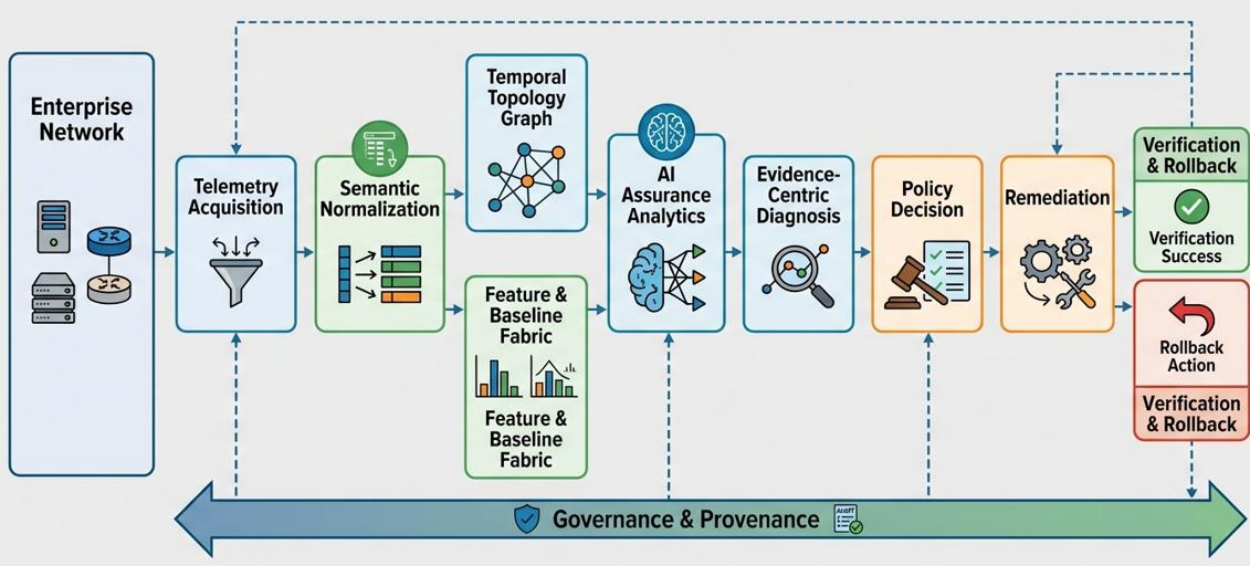

Figure 1 – Telemetry-to-Action AI Assurance Architecture

3.1 Intelligent Telemetry Acquisition and Ingestion

The assurance information is given to a system at the acquisition layer. It inputs the routers, switches, firewalls, wireless controllers, access points, servers, endpoints, applications, and cloud platforms, and network-management systems. Polling and streaming system may be embraced to realize the various capabilities of devices and telemetry demands. The streaming telemetry gives a low latency view of events that quickly evolve, and polling might be helpful to use in the presence of legacy devices and intermittently varying state.

All incoming observations contain metadata, which contains the identity of the source, date/time taken, and time of record, schema version, geographic or location, and data-quality indicators. The metadata provides the basis to subsequent temporal association and rebuilding of evidence. Simple validation, normalization, reduplication and filtering is also done at the ingestion layer followed by information being relayed to the downstream components.

Major events may cause bursts of telemetry that may overload storage and analysis capacity. Acquisition layer is therefore a part of the adaptive sampling, prioritization, buffering and load-shedding functions. Unscheduled critical events, such as interface failures or security alarms can have a higher processing priority than repetitive low-value measurements. This enables the assurance platform to be running at their optimum time when analytical demand is at its peak.

3.2 Semantic Normalization and Temporal Network Graph

As different vendors, protocols, and infrastructure locations differ, so are the characteristics of raw telemetry. The semantic layer converts these source specific records to canonical records and relationships. It is described in a typical semantic model of devices, interfaces, clients, applications, services, sites, policies and network paths. This enables logical components to reason over heterogeneous data sources, without having to repeat executing vendor specific interpretation logic.

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 4, Issue 1, January-February 2022 |

DOI:10.15662/IJEETR.2022.0401006

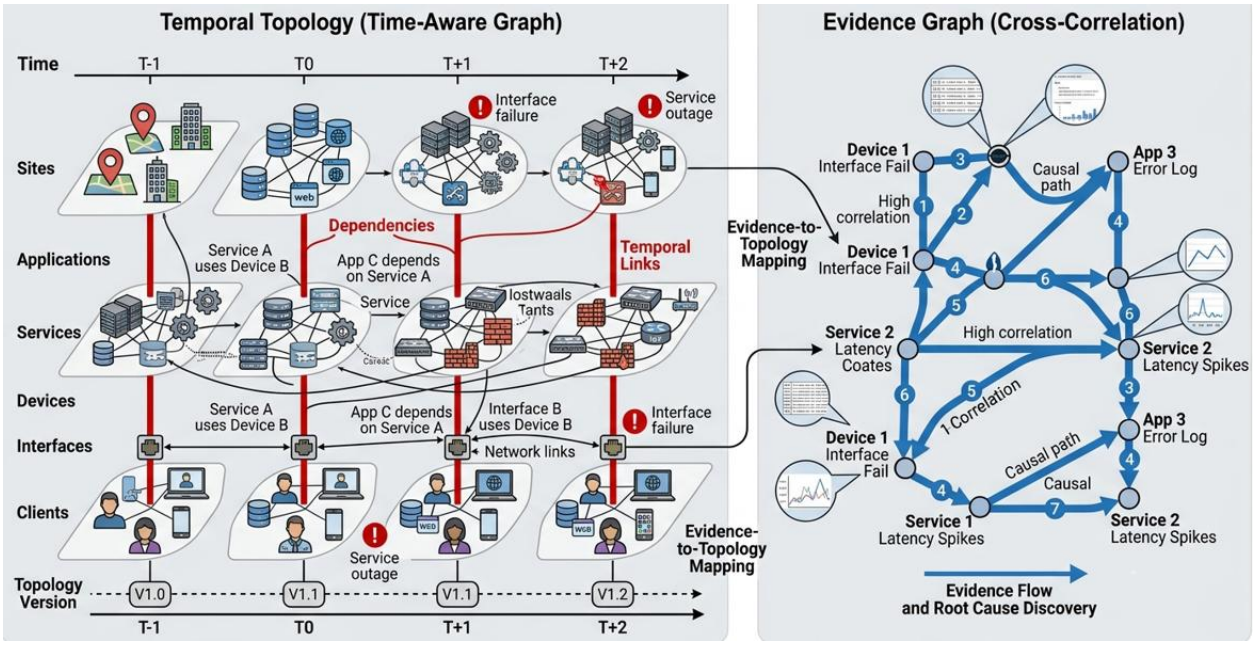

Figure 2 – Temporal Topology and Evidence Graph

A temporal topology graph is one of the most important attributes of this layer. Diagrams show relationships between physical and logical connections, routing instructions, and the dependency of services, policy associations and management details. Notably, topology is believed to change over time, but exist as a dynamical time-dependent one. Routing, the presence of devices, configuration or service dependency as a source of graph mutations the resulting new graph states can be modeled as events.

An example could be the fact that by watching several users experience application degradation in parallel graph can be utilized to conclude that they are utilizing a shared access switch, wireless controller, firewall path or upstream service. It is such relations that compose the contextual evidence to interpret the difference between a general root cause and many personal mistakes.

All analytical outcomes inferred are linked to version of observations and topology that are used in the inference. This provenance has demonstrated that operators are able to formulate the intent of an outcome that has been generated and assist in analysis of post-incidents.

3.3 Governed Feature and Baseline Fabric

Normalized observations will be coded into an analytic form that can be observed and predicted with the feature and basis layer. Calculation of features is based on more than one temporal window since network failures have varying time scales. Short windows can be used to identify sudden packet loss and spikes of latency and long windows can be used to identify gradual performance and pattern degradation which can be noticed on a daily or weekly basis.

Generation of the baselines happens at the relevant contextual levels and does not solely seem to be based on the global thresholds. As an example to this, the normal traffic that is in one of the branch offices may be very different when compared to a data center. False alarms will be reduced and more contexts of the anomalies will be increased by context sensitive baselines.

Missingness preservation is a significant architecture need. A telemetry reading which is absent (not a valid telemetry reading) should not automatically be considered a valid reading since the absence itself may be due to failure of a device to measure it, or a failure to measure it. The feature pipelines distinguish between zero, missing, stale and invalid observations as well.

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 4, Issue 1, January-February 2022 |

DOI:10.15662/IJEETR.2022.0401006

A managed feature repository is a repository of standardized definitions, ownership, lineages, and versioning and usage data. This allows training and inference systems to share features with identical name and various mathematical definitions, and helps alleviate training-serving inconsistencies and enhance model reproducibility.

3.4 Multi-Modal Assurance Analytics

The analytics layer combines complementary techniques of analysis instead of making use of single techniques of AI. Rule-based mechanisms can detect known safety conditions, and deterministic violations. A local or contextual deviation is formalized by relying on the statistical processes to discover the deviations of what is being stipulated. Patterns of recurring incidents where the use of labeled data (which is representative enough) is available are the models of supervised learning.

Another piece of analysis, toposubgraph reasoning discusses connections between the affected components by topology. Candidate causes can be ranked by the ranking of the candidates, according to their position in dependency paths, number of entities that are impacted, time dependence and actual propagation patterns. In addition the forecasting models can also predict more about the probability of an emerging condition that can lead to service-objective degradation.

They are thoughtfully considered to be complementary sources of evidence. Statistical anomaly score, by itself cannot determine root cause whereas having a relationship with a graph cannot determine causality. This may be done by mixing autonomous signals that may boost trust and reveal variations in analysis strategies. This multi-model design too offers the resilience in the event one of the models is not available, is calibrated poorly or due to distribution drift.

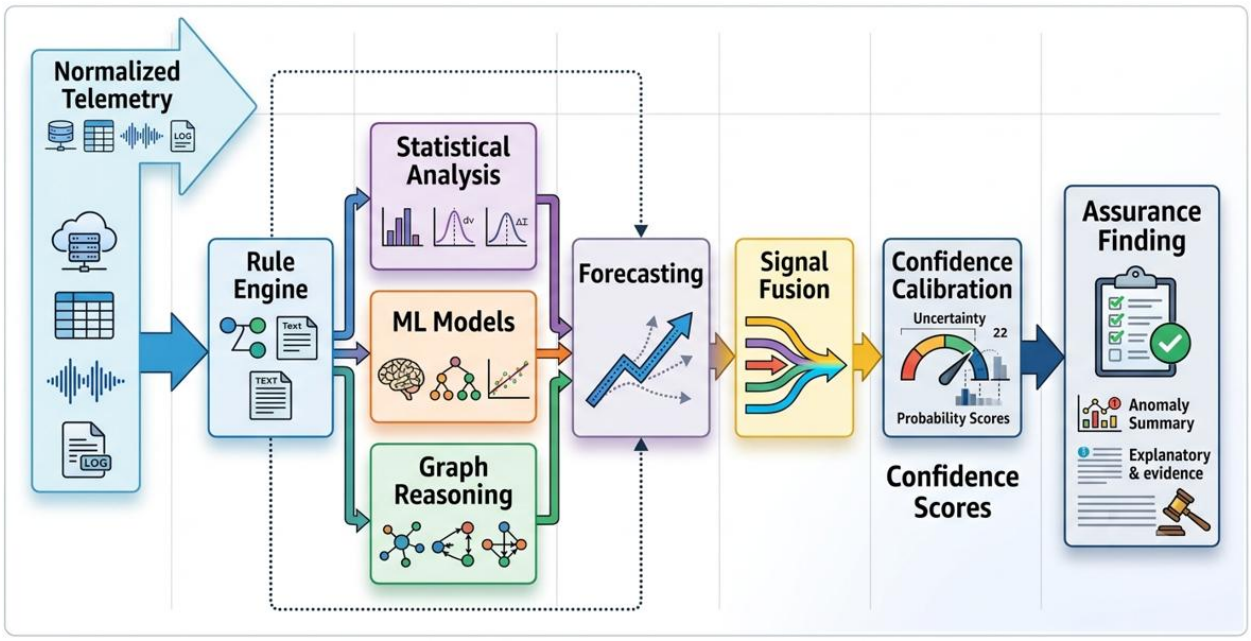

Figure 3 – Multi-Modal AI Assurance Analytics Pipeline

3.5 Evidence-Centric Diagnosis and Assurance Findings

Individual model outputs are then transformed into structured assurance findings by the evidence layer. Rather than contracting the raw predictions to operators or automation systems it makes decisions based on the level of confidence and coverage to the area of impact, time series, support of topology, conflicting evidence and quality of the data it receives.

Evidence must be correlated with preliminary observations, version of features, version of models, version of state of topology and versions of pertinent policies. This form of provenance aids in explaining and enables operators to fault or override an analytical conclusion. It also creates a trail on auditing that can be utilized again after analyzing an incident to enhance models and gauge governance.

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 4, Issue 1, January-February 2022 |

DOI:10.15662/IJEETR.2022.0401006

3.6 Policy-Governed Decision and Action Plane

The last level of architecture changes the findings of assurance into work answers. Notably, analytical inference is distinctly divorced by the action decision. The outcome of low confidence can lead to a further set of telemetry data or order further diagnostic. An incident, a ticket, or a remediation can be created based on a moderate-confidence discovery. Only in the case of an approved action pattern, a high-confidence finding can be considered to be automated and remedial.

Subsidy: operating policies are considered at the decision engine which in turn authorizes decision to proceed. Some of the controls that are of interest are maintenance windows, authorization requirements, frequency of change and the extent to which that scope is affected, the significance of a service, blast-radius limits and the availability of rollback. These checks guard against the fact that an otherwise accurate AI recommendation has unwarranted operational impacts.

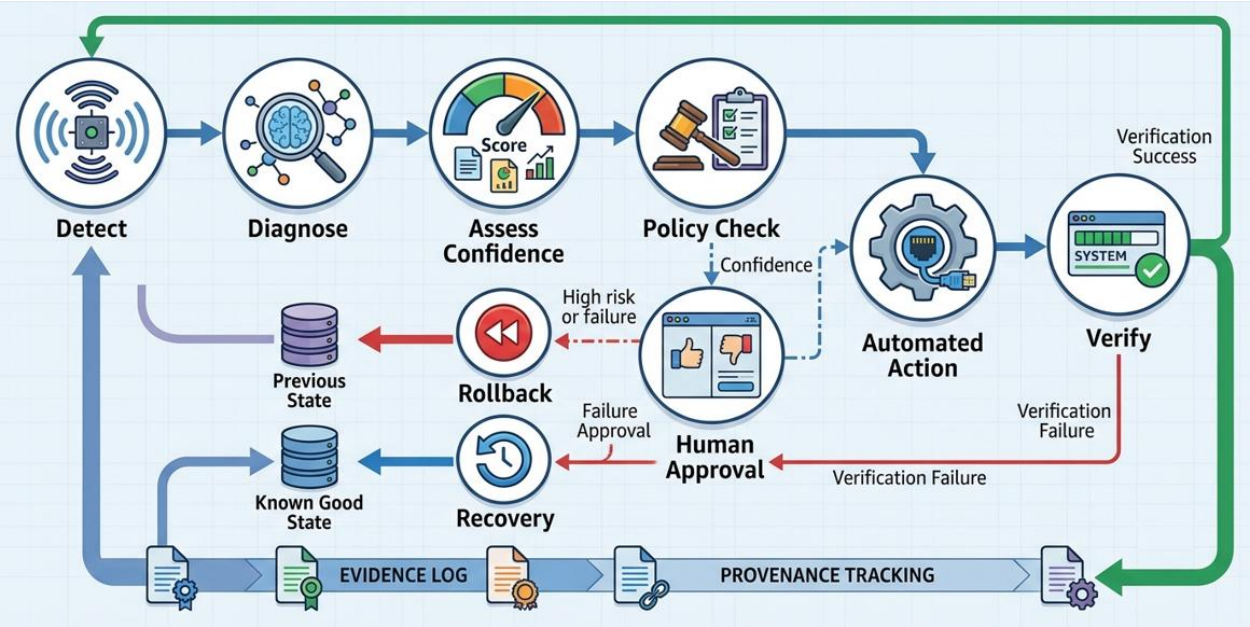

Figure 4 – Evidence-to-Action Closed-Loop Assurance

Automation in a closed-loop way is thus pursuing a sequence that is controlled; that is detect, diagnose, authorize, execute, verify and recover. The system has been applied to the establishment of reinstatement of the desired outcome after remediation. To prevent failure of the process, it could be halted or countermanded and the human operators could be notified of the circumstance.

3.7 Cross-Layer Governance and Provenance

The layer of AI model is not the only one that the governance is bound to; instead, it is all across the architecture. The data lineage, information in the assurance lifecycle, model versions, topology versions, policies, analytical evidence, decision, actions, and verification results ought to be traceable. This forms a successive chain of provenance between observation to inference, inference to decision as well as a decision to an outcome in operation.

This type of architecture can assist AI to be integrated into the enterprise network assurance without having to be an unmanageable working force. The reference architecture offers a scalable platform that helps offer reliable, explainable, and more autonomous network services via seamless real-time telemetry mixed with temporal topology-wise intelligence, managed analytics, backed by evidence-based reasoning and policy-directed automation.

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 4, Issue 1, January-February 2022 |

DOI:10.15662/IJEETR.2022.0401006

IV. EVIDENCE-TO-OUTCOME EVALUATION AND OPERATIONAL SAFETY FRAMEWORK

AI guided network assurance architecture must go through an additional test of the conventional model accuracy. The other useful evaluation system should be based on automatically measuring the anomaly detecting, root-cause diagnosing, confidence, quality of explanation, resilience to changing network conditions, and the safety of operation. Such differences between these dimensions make sure that a model is not determined a successful one only due to the fact that it detects anomalies and can generate the unreliable diagnoses or unsafe remediation options.

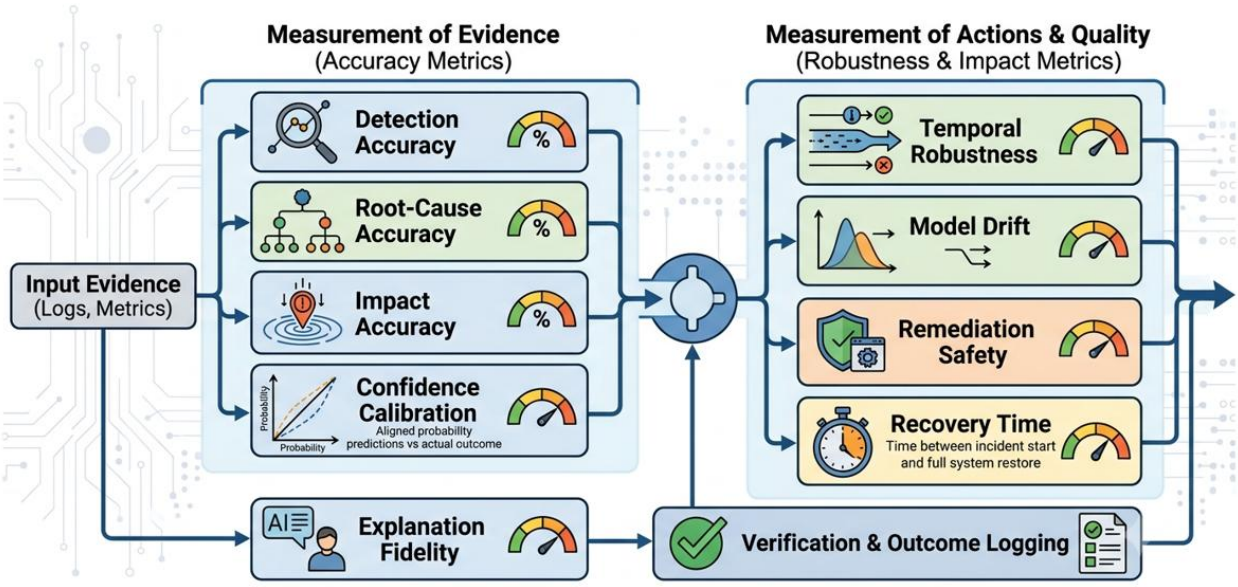

Figure 5 – Evidence-to-Outcome Evaluation Framework

4.1 Anomaly Detection Performance

The initial evaluation dimension is how well the architecture can understand abnormal network behavior in the quickest time achievable. This metric measure must involve the accuracy, recall, F1-score, turnaround time to detect and false alarms per entity under control. Precision is the ratio of the deviations (warned) which capture the relevant abnormalities and recall is a property that looks at capturing real events. Time to detection will be used to establish how fast the system will pick up a problem that has occurred when a problem has actually occurred.

The frequency of false-alerts is of great concern especially when dealing with large business environments. When issued on thousands of devices, interfaces and services and a false-positive is fairly probable, thousands of alerts will not be necessary. A measurement of performance in terms of alert volume, at both model and infrastructure level should then be done. The incident extent of the level of severity also needs to be known concerning how the architecture prioritizes the critical failures relative to the performance of the identified performance.

4.2 Root-Cause and Impact Diagnosis

They are supposed to be specifically inquired about diagnosis and detection since it would not be to show that something has gone wrong but rather it would be to show that the system knows the reasons behind its happening. Top-1 and top-3 root-cause accuracy, causal ordering accuracy, affected-scope accuracy and time to a useful explanation are the measures of diagnosis.

Top-1 is a measure of the correlation of highest rank cause to the actual cause and Top-3 is a measure of whether or not the correct cause is in the top candidates. Causal ordering assessment of the system is to check if the system is differentiating correctly between primary failures and downstream symptoms. Affected-scope accuracy determines the availability of an accurate identification of impacted users, devices, services or locations in the architecture. All these steps are crucial towards making the decision on the usefulness of AI-generated results in the operation and not the statistical plausibility.

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 4, Issue 1, January-February 2022 |

DOI:10.15662/IJEETR.2022.0401006

4.3 Confidence Calibration and Explanation Fidelity

Comparison of confidence with the correctness has to be made. Neither is it a system that can be regulated to 95% certainty in the predictions which are only correct 60 percent of the time. The predicted levels of confidence should thus be compared with the actual results in various incident types and operating conditions by the calibration analysis.
Fidelity of explanation should also be experimented upon. In all findings, the system has to find out how the observations and relationships will be the supporting evidence. Then the identified evidence can be either disappeared or can be utilized to modify the conclusion and whether or not the conclusion is modified as per expectation. An explanation might be just a description of how things are but not taken causally into consideration in the case where removal of the allegedly crucial evidence may not make any difference in making the prediction. The technique provides a more rigorous explainability test as compared to reporting feature importance or textual explanations.

4.4 Scenario-Based and Temporal Evaluation

Normal traffic traces and frequent cases of incidents should be re-created using the operative tests. The failure points are to be in failure link, device saturation, authentication failure, routing modification, telemetry failure and correlated failure. Simultaneous and cascading failures should be part of these experiments since enterprise failures are seldom one-shots.

Topological changes should not be arbitrary, but strategic to take place during fault events. This validates that the architecture has been exploited appropriately in the use of historical state of topology, rather than taking into account modifications subsequently in a past diagnosis. There should also be inter-family, site to site, inter release, varying time performance and cross-traffic. Comparisons of this kind reveal the concealed dependence which in general will not be evident in aggregate measures of accuracy.

4.5 Model Drift and Generalization

Network settings are constantly changing due to replacement of hardware, software upgrades, configuration, and addition of new applications as well as varying user behavior. Explicit drift testing should then be a part of model evaluation. It is essential to measure the performance, both in the historical and recent time windows, and compare well know and unfamiliar device families or devices.

The variation in precision, recall, calibration error, false-alerts and diagnostic accuracy are significant aspects. The extent of degradation will be massive resulting into scrutiny of the model, retraining, feature estimation, or an interim curtailment of automated behavior. Drift management is, thus, included in the operational assurance, as opposed to a distinct machine-learning practice.

4.6 Remediation Safety and Closed-Loop Evaluation

The accuracy of analysis is not an essential attribute compared to operation safety. Whichever automated remediation it nehayevy, the assessment structure needs to achieve the precision of preconditions, adherence to the blast- radius, and rollback success, time to a demonstrably recovered status, and the actions blocked by policy. Precondition accuracy is used to derive the precondition that remediation is only run when it is obligatory. The blast-radius tests will ensure that an operation does not have an impact on infrastructure beyond its location.

Rollback success tests to find out whether or not the system can safely undo the unsuccessful intervention, time to a verified recovery tests how long time elated since an action has been started up to a point where the service outcome desired has been restored. There are as well policy-blocked actions, which reveal that governance systems can be used to guarantee that the recommendations that are unsafe do not go to the implementation.

Lastly, there should come into being systems of assurance in the context of the entire process of detection -diagnosis-explain already- decide-remedy-verify. Root-cause explanations cannot be trusted to be accurate enough to guarantee high anomaly-detection accuracy, and automated actions can add operational risk of their own. Suggested evaluation framework thus addresses analytical, explainable, robust, and safety as interrelated prerequisites of trustworthy AI-based frameworks of network assurance.

V. DISCUSSION: FROM AI-CENTRIC AUTOMATION TO GOVERNED ASSURANCE

A key principle realized by the proposed architecture is that AI is not eliminating deterministic network engineering; it shifts the deterministic controls in the assurance lifecycle. Enterprise networks cannot lose the intentionality about data by contract, validation, configuration constraints, access control, rate limits, change policy and rollback. This is because

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 4, Issue 1, January-February 2022 |

DOI:10.15662/IJEETR.2022.0401006

these controls allow them to offer some forms of predictability within which AI-based elements can be allowed to act. This places AI in the most favorable usage as a veil of augmentation, but not in the sense of an unbridled means of control.

The best AI would be that region where the ambiguity just as it is difficult aptly transfers lent him-or-herself to a collection of definite rules. Minors can be identified using complex and mutating patterns of behavior as anomalies, but machine learning and using graphs can prioritize likely causes in the dependency graphs of large scales. It is also possible to predicatively model congestion, service degradation or resource exhaustion, before they occur as traditional threshold violations. Yet, it is required that such skills can give evidence based recommendations as compared to blanket mandates. The architecture thus isolates inference and decision and action and, consequently, to the policy and operational constraints these, to regulate the effect of AI outputs on the network.

One of the implications, in this case, is that doubt and potential impact should drive maximum conservatism of automation. Where the scope is small and high-confidence results are found, it can help automated remediation to do so provided that the appropriate policies are met. Every other method leads to inconclusive conclusions or steps that have an enormous impact on service and are to be followed up by some consequential observation, diagnostic test taking, human assessment or experimental therapy. This introduces a sense of risk sensitized assurance that is proportional not just to the assurance but also operational impacts.

General descriptions of the modern-day enterprise assurance systems show that there otherwise exists a similar pipeline where heterogeneous telemetry and contextual information is gathered, advocated, analyzed and disclosed to a visualization or operational intervention [4]. The architectural design is based on this pattern, with evidence provenance and confidence calibration, time context, topology relationships, and action governance considered as a first-class architectural concern, and not a second-class implementation consideration.

Very imperative to healthy enterprise deployment is the variance. An assurance system is not that any such condition of a network is abnormal, but a description of how such a finding was obtained, what underlay's it, what service and entities are impinged, how it is persuasively demonstrated and what is operated is goal-oriented. Hence, AI is clothed in a controlled assurance outfit in the architecture whereby, intelligence, deterministic engineering, governance, and operational safety do not compete but rather work together.

VI. CONCLUSION: EVIDENCE-CENTRIC AND GOVERNED AI ASSURANCE

The AI network assurance will not be created as a prediction service, rather a system that generates evidence and delivers a decision. Predictive functionality in big enterprise arrangements is not only limited in its off-putting skillfulness but also in being answerable to substantiate working choices, audits, responsibility, and to be consistent with company policies. The proposed reference architecture has met this requirement well, as it has incorporated streaming acquisition of telemetry, semantic normalization, modeling of topology over time, controlled features computational, complementary analytical approach, evidence-based diagnosis, and policy-controlled remediation into a single assurance fabric.

One of the most significant contributions of architecture is the separation of observation, inference, decision, and action. This difference disavows the right of model output to be free-floating operational commands and sets up the right control boundaries of AI-produced recommendations. Telemetry: to stream telemetry to provide the most up-to-date operational evidence and to estimate the relationship between infrastructures, services, users, and dependencies, temporal topology. This can be done by reducing the reliance of a single analytical method through the use of statistical, rule-based, machine-learning, forecasting, and graph-based methods, one at a time, being guided by features and contextual baselines can enhance the quality of the analytical process.

It is even highlighted in the architecture, which emphasizes the factor of measured confidence and provenance of evidence. The oppressed parties, demonstration of observations, linkage of time, probable cause, lowliness, and topology setting should be established on all the assurance findings. This evidence can help network operators to confirm AI-based conclusions and aid post-incident analysis, model testing, and regulating.

Lastly, there is a possibility that the automation may be safe and risk-sensitive when the action that the automation undertakes is policy-regulated. Findings with high levels of confidence might be used in automated remediation when there are predefined conditions that would have already been satisfied, where no doubtful or high-impact cases would

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 4, Issue 1, January-February 2022 |

DOI:10.15662/IJEETR.2022.0401006

need further elucidation in the data gathering, diagnostic response, or human consent. Verification and rollback functions complete the assurance loop that ensures that the remediation is tested to the desired operation attained.

In general, the suggested architecture offers a justifiable basis for reliable AI-oriented network assurance. It integrates smart analytics and approachable determinism to apply engineering controls, such that AI can be scaled and flexible, without the need to erode the traceability, accountability, safety or control of enterprise networks.

REFERENCES

1. Clemm et al., Intent Based Networking Concepts and Definitions, RFC 9315, October 2022. https://doi.org/10.17487/RFC9315
2. H. Song et al., Network Telemetry Framework, RFC 9232, May 2022. https://doi.org/10.17487/RFC9232
3. E. Tabassi, Artificial Intelligence Risk Management Framework AI RMF 1.0, NIST AI 100 1, January 2023. https://doi.org/10.6028/NIST.AI.100-1
4. Cisco Systems, Cisco Catalyst Assurance User Guide Release 3.3.1, 2026. https://www.cisco.com/c/en/us/td/docs/cloud-systems-management/network-automation-and-management/catalyst-center-assurance/3-3-x/cisco-catalyst-assurance-user-guide-3-3-x/b_cisco_catalyst_assurance_3_2_x_ug_chapter_01.html
5. OpenConfig, gRPC Network Management Interface Specification. https://openconfig.net/docs/gnmi/gnmi-specification/
6. J. O. Kephart and D. M. Chess, The Vision of Autonomic Computing, Computer, vol. 36, no. 1, pp. 41 to 50, 2003. https://doi.org/10.1109/MC.2003.1160055