Data Science @ Evolven:
Change-Centric AIOps

by Bostjan Kaluza, Ph.D., Chief Data Scientist

Enterprises today face a major challenge resulting from two competing forces: Speed and Control. The faster you move, the less control you have. Speed is critical but so is control. This situation has only become more acute in recent years with the adoption of cloud, DevOps and agile, as these technologies and practices focus on pushing changes at ever-increasing speeds.

Symptoms such as performance metrics, application traces, and log errors indicate undesirable system behavior, discrepancies between measured and expected system state, or software and hardware faults. Symptoms, however, are not the actual causes of the problem; they are only indicators. Industry experts estimate that approximately 85% of all performance and availability incidents can be traced back to changes that have been introduced to the environment. However, IT organizations have no visibility into actual changes that cause these issues. They cannot easily link the actual changes to the issues they cause. Neither they succeed to effectively evaluate the risk of the detailed changes before they cause any issue. Today Evolven is the only vendor leveraging AI-based analytics to build a complete picture of operational awareness anchored to the real root causes, changes. Correlating together detected actual changes with existing IT processes and tools, Evolven shows what changes were planned, what detailed changes were actually executed as the result of the plan, and the impact of these changes.

This paper gives an overview of Evolven's Analytics Model,. It explains how Evolven uses Machine Learning and outlines the Evolven Patent Portfolioto assist enterprises in addressing stability, compliance, security, and risk mitigation concens.

About the author

Bostjan Kaluza, Ph.D., is the Chief Data Scientist at Evolven and a researcher in artificial intelligence and intelligent systems, machine learning, predictive analytics, and anomaly detection. Before Evolven, Bostjan served as a senior researcher in the Department of Intelligent Systems at the Jozef Stefan Institute, where he led research projects on the pattern and anomaly detection, machine learning, and predictive analytics. Focusing on detecting suspicious behavior and data analysis, Bostjan has published numerous articles in professional journals and delivered conference papers. Bostjan authored several books on machine learning and data science, exploring how to leverage machine learning using Java. Bostjan is also the author and contributor to several patents in root cause analysis, anomaly detection and pattern recognition.

Contents

1. Background

In recent years Enterprises adopted technologies and practices focusing on speed such as Cloud, DevOps and Agile. This focus on speed has disturbed the *Speed-Control* balance. And as a result, stability, compliance and security are compromised. Why is it so difficult to maintain control when changes accelerate? We believe it is mainly because when it comes to changes, IT is "flying blind". In other words, IT has no visibility into the actual, implemented changes that are occurring in the enterprise environments.

The *Speed-Control* balance can be restored by detecting and understanding the actual, implemented changes. Change detection only, for instance, such as file integritiy monitoring, is simply not enough as focus on speed generates large amount of changes that simply pile. To really understand the changes, we need to understand what was the process triggering the changes, where they are coming from, what is the quality of the implemented changes, what is the impact on the environment once the changes are implemented, and if there is an issue, what are the changes causing the problem. Only this understanding brings actionable insights that stabilize stability, compliance and control.

To answer these questions Evolven leverages AI and Machine Learning to build Evolven Analytics Engine. The esential data feed Evolven collects on its own are actual changes implemented in the environment. This data is augmented by *Symptoms*, that is, external manifestations of problems such as APM metrics, log errors, and network alerts, and *Context*, indicating where the changes are coming from suc as ITSM change requests, service requests, automated deployments, etc. Evolven Analytics Engine uses AI and Machine Learning to correlate thesse data feeds, establish causal relationship between events, estimate operational and compliance risk of implemented changes, estimates impact of changes, and identifies most likely root causes in incident investigation.

In the rest of the paper we will have a deeper look at how Evolven approaches to AIOps by leveraging AI and deep, comprehensive undersdanding of changes.

2. Evolven is a Vertical AI Company

AI products fall into two main categories: horizontal AI and vertical AI. Horizontal AI products solve a broad range of problems across many different industries. For example, AI can be used to prioritize customer leads, predict which recruit will be most successful, recommend products, identify monitoring anomalies, or target advertising.

Vertical AI, on the other hand, is AI that is applied to a specific problem in a specific industry and is highly optimized for that industry. Vertical AI companies use (1) industry-specific types of **proprietary data** they control to (2) train their AI algorithms developed with **subject matter expertise** that in turn (3) gain access to unique insights and **core value delivered by AI** and (4) provide a **full-stack product** designed from the interface down to APIs and data models to delivering the results.

The vertical AI approach is much more effective when analyzing domain-specific data. Changes, symptoms, and other data objects produced within IT Operations proecsses are not raw machine data; rather, they are represented by semi-structured information characterized by a set of specific domain properties. Evolven approach follows vertical AI as follows:

- **Proprietary data**: Completeness of the data increases the accuracy of the analysis. Evolven is the only technology collecting granular details of actual changes across the end-to-end enterprise cloud. Broad, in-depth collection of high-quality information allows Evolven to provide value out-of-the-box on day one, without relying on outside data feeds. However, additional external data sources can be brought into Evolven to gain additional insights.
- **Subject matter expertise**: Evolven combines the subject matter and technical expertise, understands industry workflows, common patterns, and the main pain points. The algorithm designs accumulate over 12 years of experience working with large enterprise customers in complex IT environments resulting in pre-trained AI models ready to be put at work on day one.
- **AI delivers core value**: Change monitoring and control can be a daunting task due to vast data. To make the data actionable, Evolven relies on a comprehensive AI engine designed to understand the context, impact of data, and the causal relationship between different data feeds. AI algorithms, explicitly designed for data feeds Evolven consumes, unlock core value for the primary

use cases such as root cause analysis, proactive change risk evaluation, change reconciliation, and inventory analysis.

- **Full-stack product**: Evolven provides a full-stack, fully-integrated solution to tackle customer problems. The solution is designed from the interface that drives common use cases all the way down the stack to the functionality, data models, and data collection that powers the interface.

How Evolven uses Machine Learning

Evolven Analytics Engine relies on three types of Machine Learning: **Discriminative** or *Supervised*, **Unsupervised** and **Generative**. In Discriminative machine learning, the algorithm learns how to assign a label using a dataset labeled by an expert. In contrast, unsupervised machine learning works with unlabeled data; the algorithm tries to make sense of by extracting patterns on its own. Generative machine learning models are used to simulate realistic conversations by understanding and mimicking the way humans communicate. These models can generate new text responses that are contextually relevant to the input they receive, enabling more natural and effective interactions in chat applications.

Supervised Machine Learning

Supervised learning algorithms automatically learn how to differentiate between different data types by using statistical models and induction for concept generalization. Some notable algorithms used by Evolven Analytics Engine are decision trees, naïve Bayes, Bayesian networks, neural networks, topic modeling, clustering, k-nearest neighbors, and others.

Evolven Analytics Engine uses Supervised Machine Learning for several tasks, for example:

- **Understand Change Type**: Each change is first classified into one of the broader change type categories: configuration, capacity, data, workload, code, relationship. Within each category, the change is further labeled with a more detailed subtype. For example, code sub-types include executables, scripts, web assets, raw source code, and other sub-types.

- **Classify Value Type**: The algorithm labels the value of configuration parameter, for example, IP, number, size, string, switch, and other value types.
- **Classify Impact Area**: Each configuration parameter is labeled with one or more impact area categories indicating what kind of impact might be expected from modifying that parameter.

Chapter 4 describes in detail additional supervised learning algorithms implemented by Evolven.

The algorithms are pre-trained in Evolven labs and shipped with the product ready to work out of the box. Training is based on anonymized samples of customer data labeled by domain experts. Algorithms are generalized to work with common technologies and tuned to optimize the tradeoff between recall and precision.

The algorithms can be further tuned once deployed in production using two mechanisms:

- User feedback overriding algorithm decision and guiding future decision process
- Configuration tuning by Evolven administrator

Unsupervised Machine Learning

Unsupervised machine learning algorithms discover interesting patterns in data on their own, including frequent patterns, anomalies, similar clusters, suspicious system behavior, etc. Algorithms require an expert to define how patterns are encoded, including topological analysis, density estimation, distance metric, dimensionality reduction. Notable algorithms are hierarchical clustering, k-nearest neighbors, expectation-minimization, principal component analysis etc.

In Evolven, unsupervised machine learning starts to work as soon as enough data is available. The algorithms have a built-in mechanism to decide when a sufficient amount of data is available, and the resulting patterns are reliable. The algorithms start producing reliable insights within a relatively short period:

- **Day one** algorithms work as soon as the initial data collection is complete. These algorithms mine for patterns across devices, environments, and resources within the organization. Examples of these algorithms include detecting type

anomaly, detecting value anomaly, auto-benchmarking, recognition of consistency patterns, clustering, insights, and others.

- **1 to 4 weeks**. The other group of unsupervised algorithms mines for patterns in each device, environment, or resource. To establish a solid baseline and learn the range of patterns, the algorithms in this group require a couple of weeks. Detection of time anomaly, calculation of change frequency, detection of workload anomalies, event clustering are examples of unsupervised algorithms that need some time to collect the initial data.

Generative AI

Generative AI in Evolven is leveraged for implementing a chat interface, which facilitates the transformation of questions posed in natural human language into queries that can effectively interrogate the Evolven database. This enhances the system's understanding of configuration changes and enables the summarization of insights generated by the other types of AI. Evolven employs models from the GPT (Generative Pre-trained Transformer) family of algorithms, which are known for their effectiveness in processing and generating human-like text.

The algorithms are pre-trained in Evolven labs with a vast corpus of IT and operations data to ensure that they are well-versed in the domain-specific language and concepts right from the start. These models are frozen upon deployment, meaning that they do not undergo direct fine-tuning in the production environment to ensure consistency and reliability of the output. However, Evolven provides an option for customers who wish to further enhance the models' performance. Customers can choose to fine-tune the models on-premise with their own data, allowing for a more customized experience that is tailored to their specific operational context and needs.

3. Evolven Analytics Engine

Evolven Analytics Engine is designed to process actual granular changes as its primary data feed. The engine is based on several patented technologies, driving the complete process from data collection, ingestion, cleaning, and analysis to final presentation.

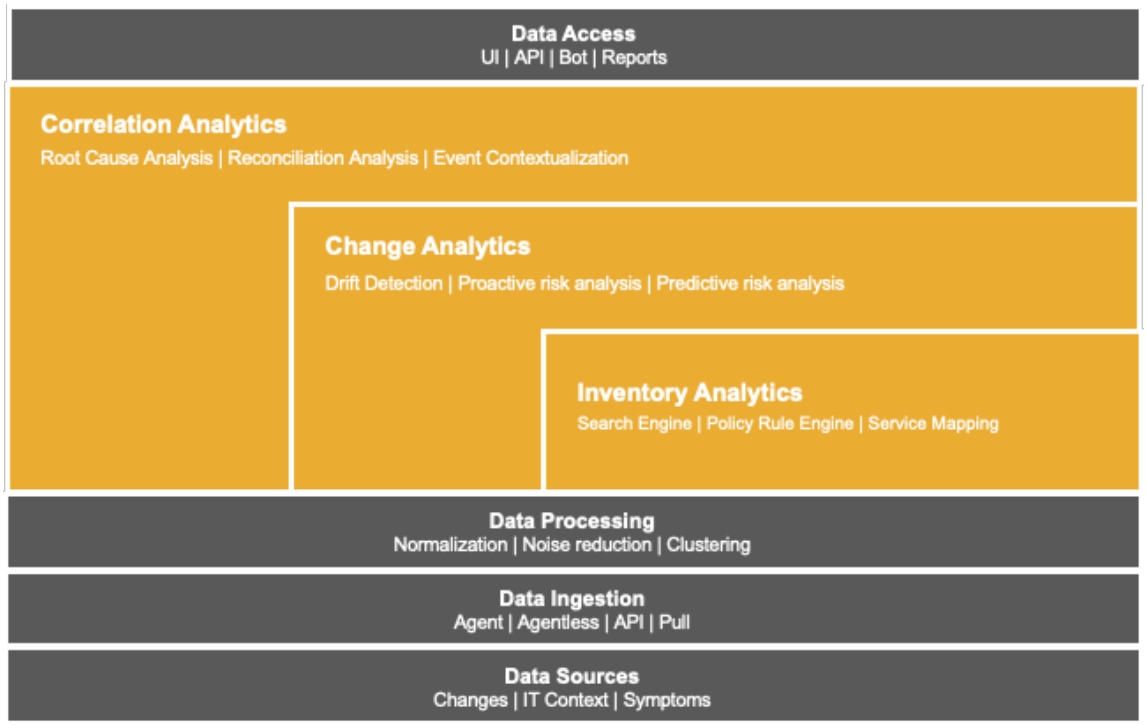

Figure 1: Evolven Analytics Engine

The architecture of the engine is organized in seven layers, as shown in Figure 1 (from bottom to the top):

1. **Data Sources:** The primary data feed Evolven collects using its own technology is the changes. Evolven collectors periodically scan configuration and inventory of the environments detecting and storing all the deltas compared to the previous scan. This mechanism allows Evolven to observe the complete current and historical configuration states of each environment. Also, Evolven has an option to ingest other data feeds that could be broadly categorized into two groups: **IT context,** indicating where the changes are coming from (ITSM change requests, service requests, automated deployments, etc.); and **Symptoms,** indicating the health of an environment or active incidents (ITSM incident tickets, APM alerts, log error events, etc.)

2. **Data Ingestion:** Our scalable data ingestion layer can support hundreds of thousands of agents reporting concurrently. Alternatively, external services can push data using API calls, or Evolven can pull data periodically from an API endpoint or a service bus.

3. **Data Processing:** Once the data is in Evolven, data is automatically cleaned, normalized, organized, and clustered. Evolven has a multi-stage noise reduction mechanism automatically detecting potentially noisy data and progressively suspending the collection of such data.

4. **Inventory Analytics:** Evolven collects near real-time granular configuration, content, and inventory of the enterprise cloud, from application to infrastructure, from the data center to the cloud. Examples of technology components scanned by Evolven include AWS/Azure cloud resources, network devices, virtualization layers, containerization layer, virtual and physical servers, containers, operating systems, application components, etc. The collected information covers each technology component's detailed configuration parameters, installed components, drivers, file systems, database tables, indices, application master data, custom application artifacts, and much more. Inventory Analytics layer provides (i) effective **search language** to query the inventory, (ii) **policy rule engine** to implement controls for policies representing compliance requirements or organization's best practices, and (iii) **service mapping** discovering technologies within host and organizing environments to applications, stages and data centers.

5. **Change Analytics:** Change analytics is based on three key technologies: drift detection, proactive risk analysis, and predictive risk analysis.
   Drift detection identifies either a **change**, which is the difference between the current configuration value and the previous value, or a **difference between several environments**, which is a difference between a configuration value in one environment and the value of the same configuration entity in other environments. Drift detection helps to analyze consistency within environments, verify alignment between production and disaster recovery environments,

investigate issues comparing working and non-working environments, assure consistency when migrating data centers or moving to the cloud, etc.

Proactive risk analysis evaluates all changes and differences to estimate their likelihood to cause issues. The resulting probability is visualized as a color-coded risk level. The calculation is based on a **decision tree** using more than 160 data points. It is tuned to calculate operational risks associated with each change. For example, did the change follow the ITSM process, is there any known impacts associated with such configuration, was the change implemented consistently, is there any anomaly associated with the actual parameter value, etc.

Predictive risk analysis estimates what will be the impact of planned changes by grabbig the changes in lower enviroments and simulating what would happen if this changes were applied to production environments.

6. **Correlation Analytics:** Correlation analytics connect actual changes with IT context and symptoms building an end-to-end picture of operational awareness. This analytics layer is essential to identify a mismatch between planned and actual changes (Reconciliation Analysis) and to establish impact of the changes correlating them with issue symptoms (Root Cause Analysis). Essentially, Correlation Analytics allow Evolven to identify unauthorized changes and shortlist changes that could cause investigated Issues.

   Root Cause Analysis is based on **Bayesian network** identifying changes with the highest likelihood to cause reported issues. This Bayesian network updates the initial proactive risk estimate with the new information about the incident that actually happened. For example, when did the change happen compared to when the issue started? Is this point within the typical time frame for such a change to manifest a problem? What is the location of the change in the topology compared to the location of the issue? Is this the type of change that can cause such an issue? The final answer is calculated as a likelihood of the change contributing to the issue. This correlation score is expressed as a color-coded risk level.

   Reconciliation Analysis (a) correlates actual detected changes with change requests, service requests and automated deployments, (b) automatically identifies groups of standard or pre-approved changes, and (c) detects unauthorized and potentially unauthorized actual changes that did not follow a

formal approval process. In addition to temporal correlation, the reconciliation analysis combines a rule engine with several algorithms for pattern mining, semantic natural languange processing and association mining.

7. **Data Access:** The Evolven Analytics Engine ensures comprehensive access to all ingested and processed data as well as to the analytics results through a range of mechanisms tailored for convenience and efficiency:
    a. Interactive Analysis: Users can engage with data interactively via the Evolven UI, which allows for on-demand data analysis, exploration, and the creation of custom dashboards.
    b. Push Notifications: Reports can be automatically sent to users through email attachments or API calls to external services. This automated reporting ensures that stakeholders are kept informed with the latest analytics insights without the need for manual queries.
    c. Chat Integration: Separately, the Generative AI enables integration with collaboration tools such as Microsoft Teams or Slack. Users can interact with Evolven's chat interface to ask questions in natural language and receive summarized insights directly within their chat environment, streamlining the process of data access and decision-making.
    d. Pull Mechanism: External services can retrieve data via Evolven API calls or by querying materialized database views. This supports third-party integrations and custom workflow automations.
    e. Real-time Streaming: Data is also available through streaming to an enterprise service bus, facilitating continuous downstream processing and real-time analytics capabilities.

4. Patent Portfolio

The key components of the Evolven Analytics Engine are based on several innovative technologies awarded with patents granted by the US Patent Office. Figure 2 provides a list of the patents and areas of their application.

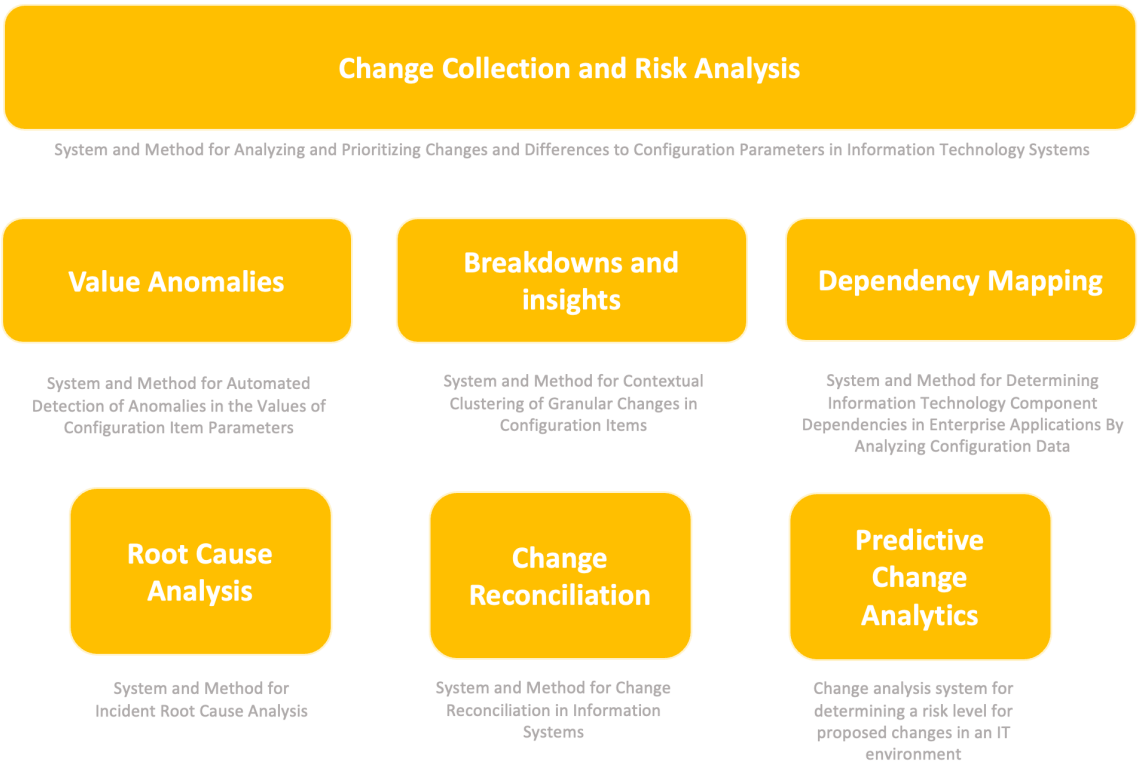

Figure 2: Evolven Patent Portfolio

1. **Change Collection and Risk Analysis**

   The patent is a cornerstone of the Evolven Analytics Engine. It covers (i) collection of configuration values, (ii) comparing them to previous values, (iii) storing the diffs to a database, (iv) calculating several risk dimensions such as consistency, operation, impact, etc. and applying user input such as policies, filters, etc., and (v) prioritizing the changes by risk.

   More details are available at https://patents.google.com/patent/US20160042285A1

2. **Breakdowns and Insights**

   The invention is about the automated grouping of change data that creates meaningful chunks labeled with an easy to understand name. More technically, the first step applies **multi-level clustering** of changes based on similar characteristics, for example, environment, action, host, impact, etc. and uses context-free grammar to create a **machine-generated descriptor** in natural

language (English). The result is automated breakdowns, putting changes in comprehensible context and translating them into actionable insights.
More details are available at https://patents.google.com/patent/US20170195178A1

3. **Value anomalies**

Evolven has a unique asset in the vast amount of diverse configuration data collected across the enterprise cloud. This data can be utilized to automatically assess the actual parameter values. Evolven has developed a unique approach that goes beyond user-defined policies, commonly applied by other vendors. The patent covers a set of algorithms that **automatically benchmark a parameter value** from different points of view. For instance, what it the type of a parameter value (Boolean, integer, IP, string, etc.) compared to the previous value or compared to the same value in other environments? What about the value magnitude? Is the analyzed value in the expected range of parameters? What about the change magnitude? Did the parameter value change for an amount within expected range? More technically, the value anomalies algorithms evaluate data across environments, thus reducing the time to learn to day one. The algorithms are based on unsupervised machine learning calculating entropy to estimate a level of surprise.
More details are available at https://patents.google.com/patent/US20180239682A1

4. **Root cause analysis**

Evolven's RCA approach calculates the likelihood of a change being the root cause of an issue under investigation. Evolven has analyzed thousands of incidents caused by changes to learn the typical **fatality rates** for various types of changes, i.e. the frequency of occurrence of incidents during a specified time interval. For example, a bad change in firewall blocking port 443 for an application in production, will result in an incident much faster compared to a capacity change in the "connection pool size" parameter of a database server. To calculate the likelihood of a change being the root cause, the approach first classifies each change into a category with an attached fatality rate model. Then it calculates **change lifetime profile**, i.e. the expected change impact over

time. Finally, it correlates the change to the incident using **Bayesian network** taking into account other change and incident properties.
More details are available at https://patents.google.com/patent/US20170213142A1

5. **Dependency mapping**
From the beginning Evolven performed host based discovery detecting key applications installed on operating system to crawl granular configurations in-depth. This invention presents the next step, analyzing collected detailed configurations to establish a dependency map between components. Compared to the common network-based dependency mapping, this approach can detect non-active dependencies, which cannot be observed in network traffic, for example, a dormant dependency to an external service that is triggered under specific conditions only. The approach crawls configuration data to find references to other environments and configuration tems. Then it uses machine learning to determine the type of dependency. Finally, the identified dependencies are mapped to common architectural patterns to establish a dependency map.
(This approach is not yet fully implemented in the current version of Evolven's platform)
More details are available at https://patents.google.com/patent/US20190081861A1

6. **Change reconciliation**
At the core of Evolven's patented technology lies the capability to precisely identify actual changes within an environment and meticulously contrast these against scheduled change requests and deployment actions through advanced correlation techniques. The system harnesses the power of machine learning and AI, including the use of natural language processing to comprehend the extent of modifications stipulated by change requests. It then aligns the identified changes with those that were planned, culminating in the generation of an authorization score. This score aids in determining whether a detected change corresponds with a planned or previously approved action. The calculation of the authorization score incorporates principles from game theory, balancing the cost associated with misclassifying a change as either authorized or

unauthorized. More details are available at https://patents.google.com/patent/US11290325B1

7. **Predictive change analytics**
   This patent details the development of a predictive change analytics system designed to enhance the stability and security of IT environments. The system functions by capturing a configuration snapshot of the intended target environment. It then generates a manifest detailing the impending changes. Utilizing this manifest, the system simulates the post-change state of the environment, enabling it to assess the potential impacts on stability, compliance, and security. This simulation process facilitates the accurate forecasting of the effects of proposed changes, providing valuable foresight at various stages of the change management process.
   More details available at: https://patents.google.com/patent/US11894976B1

Evolven continues to innovate developing many more patents for the new technologies upcoming in the future versions of the Evolven's platform.

5. Summary

Analysts, industry experts and enterprise IT staff agree that changes are the most likely cause of performance and availability incidents. Today, Evolven is the only vendor that detects actual changes at the most granule level across the end-to-end enterprise cloud, reconciling them into operational awareness picture, prioritizing them by the risk and correlating with issues they are causing. Complete granular visibility is essential not to miss any change but it creates volumes of data. AI-based analytics is the only effective way to highlight a needle in a haystack, pinpointing the changes that carry the highest risk or are the most likely incident root causes.

Evolven has developed an analytics technology that leverages numerous data sources, thus creating a cross-silo perspective. The core component is Evolven Analytics Engine based on a combination of machine learning, anomaly detection, causal correlation, and expert knowledge specifically focus on all the unique properties and attributes of the semi-structured change data This technology introduces a number of unique

patented capabilities required to investigate incidents, prevent incidents, and maintain the highest quality of service across the enterprise cloud.

About Evolven

Change propels businesses forward but is also the primary source of errors that disrupt the services our businesses rely on.

Evolven's AI-Powered Change Control & Analytics technology is helping ITOps, DevOps and SRE teams reduce the risk of stability, compliance & security issues stemming from changes. Evolven lets you finally know all actual changes carried out in your environment and uses machine learning to detect and prioritize the riskiest changes, resulting in fewer incidents, faster MTTR and improved productivity.

Evolven is a recognized AIOps leader and was selected by Gartner as a Cool Vendor in IT Operations. Evolven is also the winner of the Red Herring Top 100 North America. TiE 50 Top Startup, 20 Most Promising Data Center Solution Providers, Banking CIO Outlook and ITOA50 awards.

For a free demo call 1-888-841-5578, or for all the latest updates from Evolven follow us on Twitter@evolven