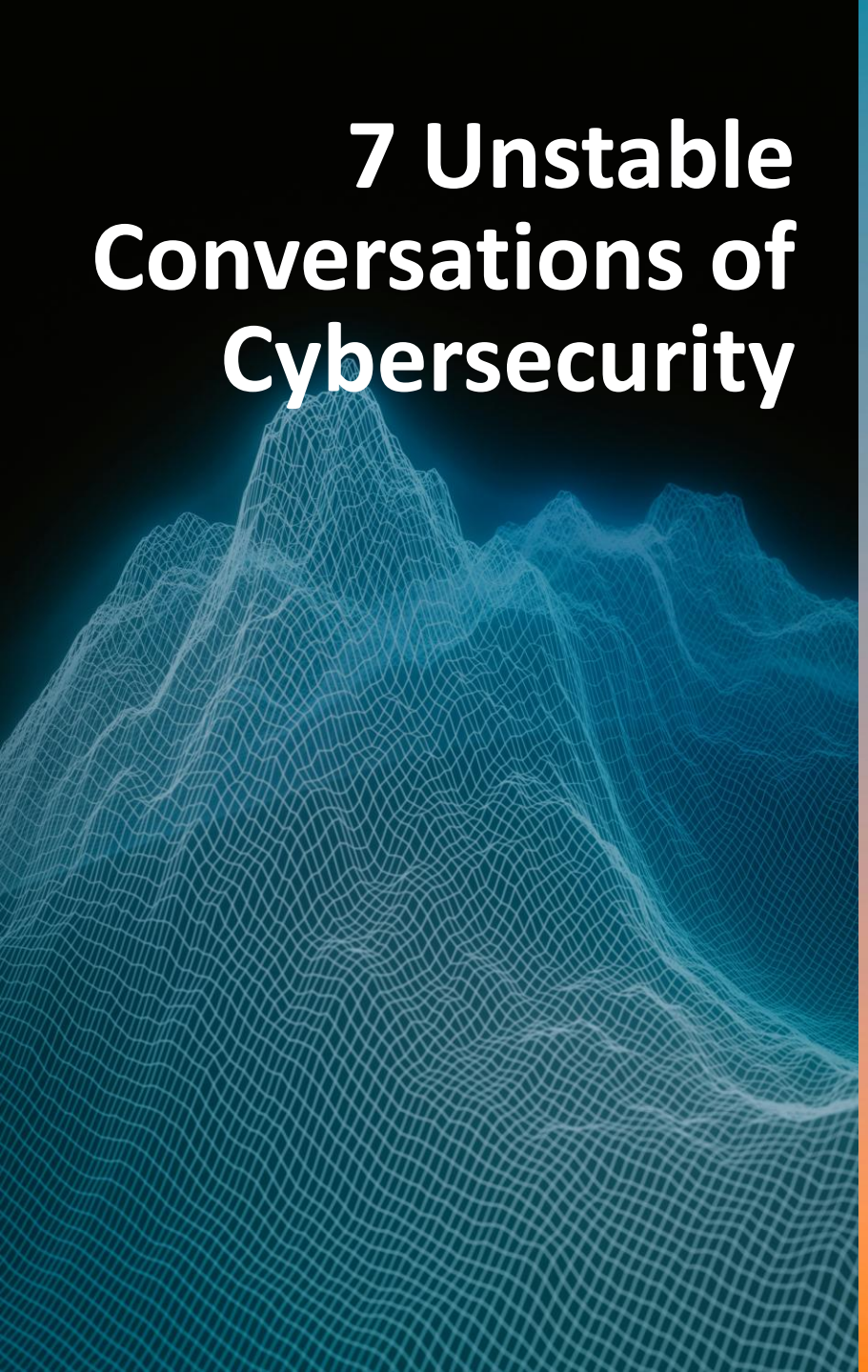
WitFoo

Empathetic Processing in Cybersecurity: A Human-Centric Paradigm

Bridging the gap between data overload and actionable insight through intelligent automation that thinks like a human analyst.

Charles D. Herring, co-Founder WitFoo
Charles@WitFoo.com
https://CharlesHerring.com

7 Unstable Conversations of Cybersecurity



1. Investigators do not understand what their tools are saying
Security analysts often face overwhelming data and lack the context needed to interpret alerts, leading to confusion and missed threats.

2. Managers cannot track security practice success
Security managers struggle to measure and communicate the effectiveness of their teams, making it difficult to supervise, improve, or justify investments.

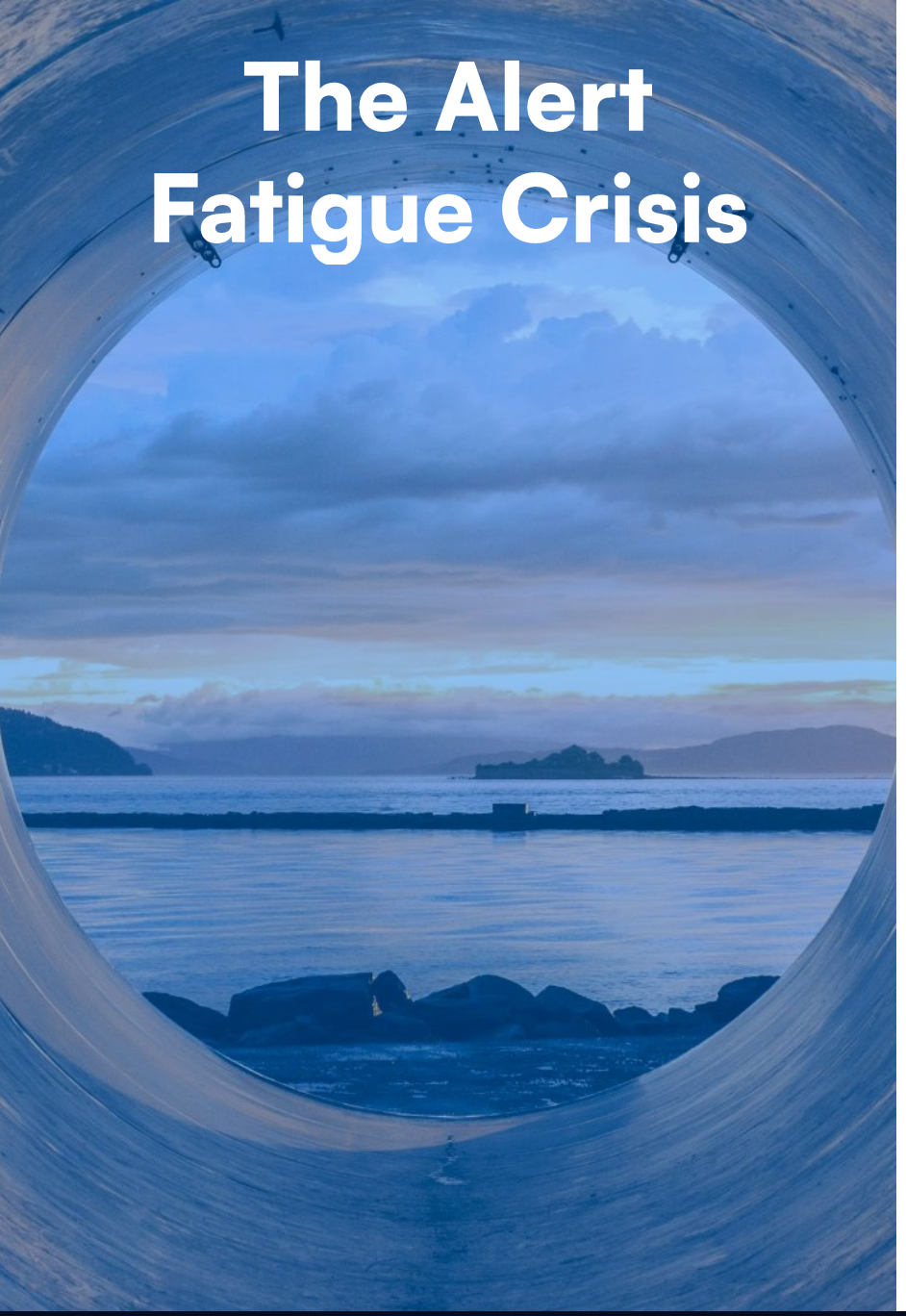
3. Security practice cannot express value to business
Security metrics rarely align with broader business goals, leaving managers unable to demonstrate the true value of cybersecurity to executives.

4. Security vendors cannot be held accountable
Organizations lack the means to accurately assess vendor performance, such as false positive/negative rates, making it hard to enforce accountability.

5. Organizations cannot safely share information with each other
Sharing actionable intelligence between organizations is risky and expensive, resulting in limited collaboration and reliance on vendors.

6. Organizations cannot safely report crimes to law enforcement
Reporting breaches often requires giving law enforcement unfettered access, which is a major deterrent and leaves many incidents unreported.

7. Law enforcement lacks evidence to prosecute criminals
Due to the above barriers, law enforcement is frequently unable to gather sufficient evidence, allowing cybercriminals to operate with impunity.



The Alert Fatigue Crisis

By the Numbers

- Modern Security Operations Centers face an overwhelming challenge. Teams receive an average of **4,484 alerts per day**, yet cannot review approximately **67% of them**. This deluge of information creates a critical vulnerability where genuine threats hide in plain sight.
- The majority of these alerts are false positives, leading to wasted effort and analyst burnout. When every alert demands attention, no alert receives proper attention.

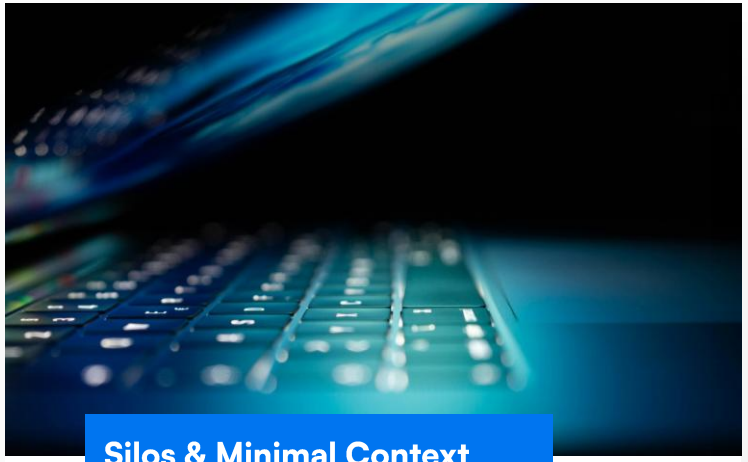
Real-World Impact: Alert Fatigue Has Severe Consequences:

- Critical attacks missed or detected too late
- Analyst burnout and high turnover rates
- Slow response times to genuine threats
- Decreased organizational security posture

The current approach creates a dangerous paradox: the more security tools we deploy, the less secure we become.

Traditional SOC Challenges

Organizations find themselves in a "data rich but information poor" situation. Despite investing millions in security infrastructure, they struggle to extract meaningful insights from their telemetry.

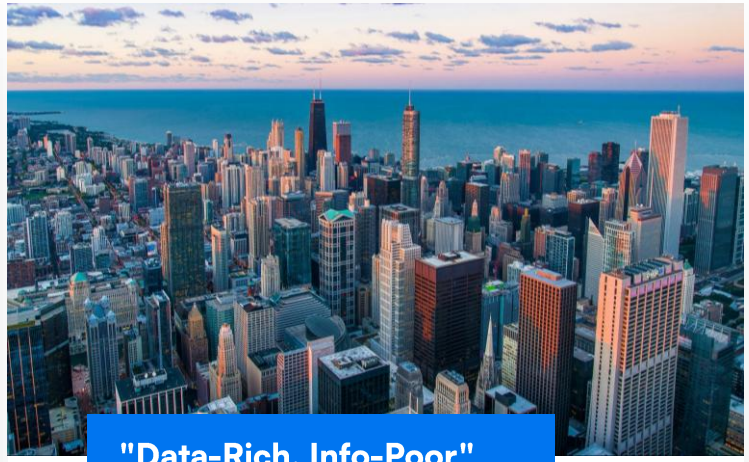


Silos & Minimal Context

Traditional SIEM pipelines ingest everything but perform minimal parsing or correlation upfront. Data is collected quickly in raw form, leaving interpretation entirely to analysts after the fact. Without context, every log line looks equally important—or equally meaningless.

Reactive & Unmanageable

Because data isn't understood at ingest, analysts must write complex correlation rules or manually piece together context long after events occur. This reactive approach results in slow, inconsistent analyses and generates excessive noise that obscures real threats.

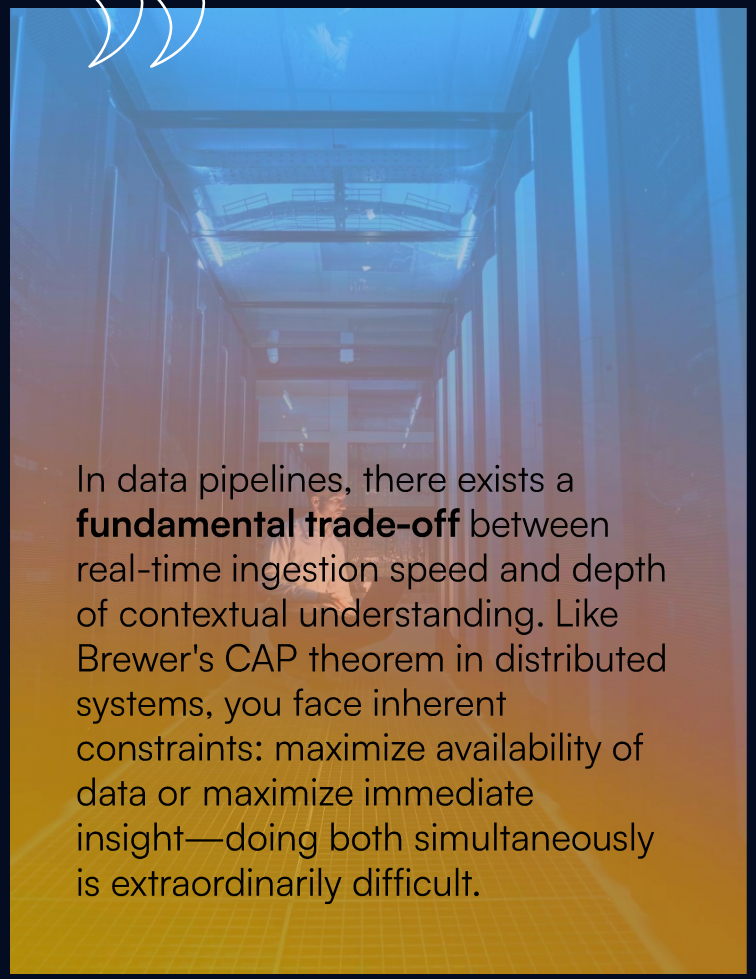


"Data-Rich, Info-Poor"

Organizations have abundant telemetry but lack actionable intelligence. Adding more security tools often worsens the problem by flooding teams with additional raw alerts without context. The volume grows, but understanding doesn't.

Current systems lack human-like understanding, creating an urgent need for a fundamentally new approach to security data processing.

The Fundamental Trade-Off



In data pipelines, there exists a **fundamental trade-off** between real-time ingestion speed and depth of contextual understanding. Like Brewer's CAP theorem in distributed systems, you face inherent constraints: maximize availability of data or maximize immediate insight—doing both simultaneously is extraordinarily difficult.

Traditional Approach

Traditional SIEMs prioritize availability: they ingest and store data as fast as possible, ensuring high throughput and data retention. However, they sacrifice "consistency of insight"—the data isn't fully understood or contextualized when stored. This results in a firehose of unfiltered alerts requiring heavy analytical lifting later, when time is critical.

The Challenge

Fully parsing and enriching everything at ingest historically seemed unworkable due to performance concerns. The computational cost appeared prohibitive, so the industry accepted the trade-off: collect now, understand later.
But what if modern technology could change this equation?

Left VS. Right Pipeline Processing

Right-heavy Processing

Conventional approach: minimal processing at ingest, just tag source and timestamp, then dump into storage. Analysis and correlation happen later, on-demand or in batches. This ensures nothing is missed initially, but leads to slow insight—context is added only after the fact.

Left-Heavy Processing

Emerging trend: invest compute upfront by normalizing and enriching data as it streams in. Events are already contextualized by the time they reach an analyst. This continuous analysis trims noise early and enables immediate understanding.

The Shift

Most legacy systems are right-heavy, but modern approaches—streaming analytics, UEBA, and now Empathetic Processing—are **shifting intelligence toward the left.** The future invests computational resources upfront for dramatically smarter data handling.

A Tale of Two Systems

Consider a concrete scenario that illustrates the stark difference between traditional and Empathetic Processing approaches.

Traditional SIEM Response

10:00 AM: Endpoint security generates alert: "Malware X quarantined on Host A"

10:05 AM: Firewall logs show Host A contacted malicious IP address

Result: Two separate, unlinked alerts appear on different consoles. An analyst might not realize they're related without extensive manual correlation. The connection between quarantine and subsequent suspicious activity remains invisible.

The malware's attempt to communicate after quarantine—a critical indicator of compromise—goes unnoticed as an isolated firewall event.

Empathetic Processing Response

10:00 AM: System notes and contextualizes malware quarantine event for Host A

10:05 AM: System recognizes outbound connection involves same host with recent malware activity

Result: Single, comprehensive incident alert: "Host A shows post-quarantine malicious activity—possible ongoing compromise despite remediation attempt."

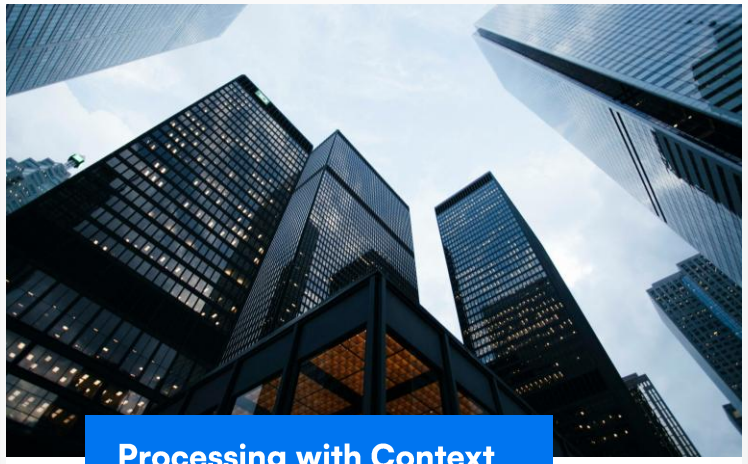
The system automatically fuses both pieces of evidence into one coherent narrative, complete with timeline and recommended next steps. The analyst immediately understands the full story.

What is Empathetic Processing?

Empathetic Processing is a paradigm that models the security data pipeline on human communication and reasoning. Rather than treating data as isolated log lines, it attempts to truly understand data in context and produce output that humans find intuitive and actionable.

Listening with Empathy

The system comprehends incoming events like a human analyst would—understanding the who, what, where, when, and why of each security event, not just mechanically parsing fields.



Processing with Context

Events are analyzed in the context of all related activity, building a comprehensive understanding of what's actually happening across the environment over time.

Speaking with Clarity

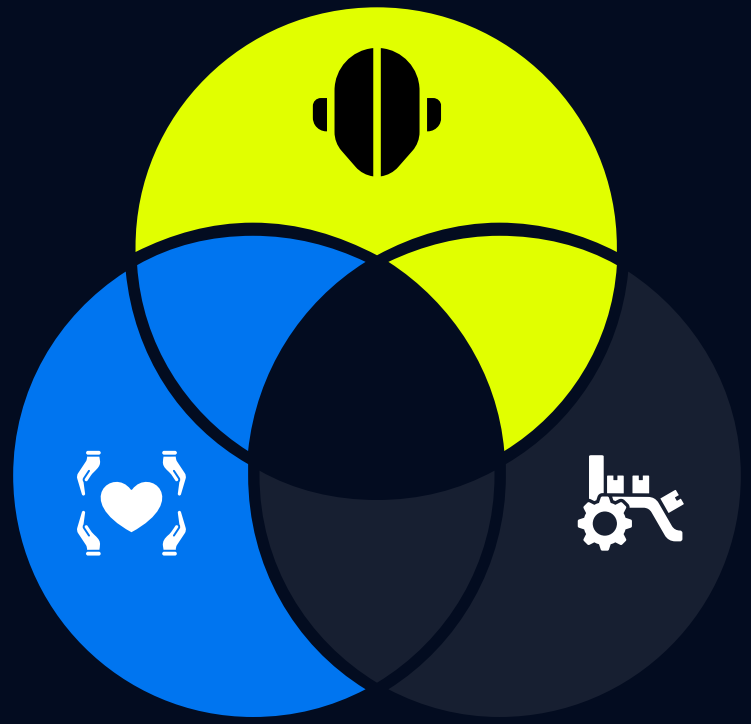
Results are delivered in relatable, narrative formats tailored to the audience—whether that's a technical analyst, executive, or compliance auditor.

Empathetic Processing shifts intelligence to the ingest phase, dramatically reducing noise and easing cognitive load on analysts. It bridges the gap between Big Data and human understanding by treating each alert as part of an unfolding narrative rather than an isolated data point.

The Three Stages of Empathetic Processing

Dissonance Resolution

Correlation and reconciliation of events over time. The system connects the dots through graph-based analysis, resolving conflicts and redundancies to form a coherent picture of security incidents.



Empathetic Listening

Deep comprehension of incoming data. The system understands what each event means in context—who's involved, where it happened, and why it might be significant—instead of performing shallow parsing.

Empathetic Speaking

Human-friendly output generation. Results are presented as narratives, summaries, and reports tailored to each audience's needs and expertise level, ensuring information is both understood and actionable.

Empathetic Processing operates through three interconnected stages that work continuously and in parallel. Together, they transform raw security telemetry into actionable intelligence. These stages don't occur sequentially—they operate as a continuous, integrated system where listening informs resolution, resolution enhances future listening, and speaking adapts based on what's learned throughout the process.

Stage 1: Empathetic Listening

Empathetic Listening transforms data ingestion from a mechanical collection process into an intelligent interpretation system that comprehends the meaning and significance of each security event.

Semantic Interpretation

The system doesn't just ingest logs—it fully interprets each message at ingest using NLP and advanced parsing. A firewall log line is read almost like a sentence, identifying the "who did what to whom, when, and why."

Context Capture

Every event receives rich context: who is the source (device/user), what happened, where and when it occurred, and why it might be significant. The system classifies event types and understands source roles—distinguishing critical servers from normal workstations.

Semantic Framing

Events are normalized into structured artifacts with common schemas: actors, actions, objects, outcomes. Disparate data sources speak a unified internal language, enabling seamless correlation regardless of log format or source system.

Stateful Memory

The system maintains short-term memory of recent events and entities. If Host A triggers ten events in succession, the system knows they're related to Host A's current state. Each new event is interpreted in light of recent history—just as a human analyst remembers ongoing situations.

Through Empathetic Listening, noise is reduced early: duplicate or trivial events are recognized and filtered automatically. By the end of this stage, we have a stream of enriched, consistent event records ready for intelligent correlation.

Predestination of Data

A cornerstone principle of Empathetic Listening is predestination—tagging and structuring data with all future analytic and compliance needs in mind at the moment of ingestion. The system anticipates what might be asked of the data later and prepares accordingly.

The Concept

- Whenever an event arrives, the system labels and stores every detail that could be relevant for investigations or audits down the road. Nothing is thrown away or left unstructured.
- Even fields not immediately used are captured with appropriate tags in case they become important during future analysis.

Forensic Completeness: All evidence is organized before incidents occur, ensuring comprehensive investigation capabilities

Compliance by Design: Regulatory logging requirements are met automatically through systematic data capture

Future-Proof: Changes in analysis needs don't require re-ingestion or data loss

Instant Context: When incidents happen, all relevant context is already structured and accessible

Predestination means **planning ahead** at ingest—the system behaves as if it knows the questions analysts or auditors will ask in the future, and structures data to answer those questions readily. This forward-thinking approach eliminates the common problem of discovering critical data wasn't captured when it's needed most.

Semantic Fingerprinting

Semantic fingerprinting is the adaptive parsing technique that enables Empathetic Processing to recognize, learn, and understand any log format—a crucial capability for handling the heterogeneous data sources in modern security environments.

Generate Unique Hash

The system creates a **fingerprint hash** for each distinct message format or pattern it encounters. This fingerprint captures the structure and key tokens of the log, so two logs with identical layouts produce the same fingerprint.

Recognize or Learn

Using fingerprints, the system quickly identifies known event types and applies correct parsing rules. If a fingerprint was seen before, recognition is instant. If new, it triggers an automated learning process.

Agentic AI Research

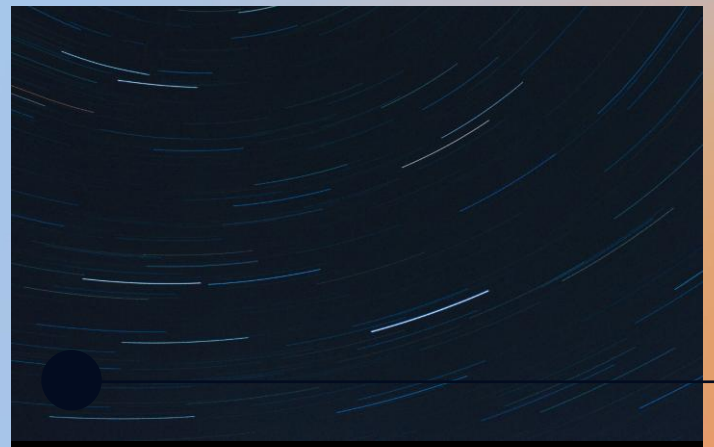
When an unknown fingerprint appears, an agentic AI component automatically researches it—searching documentation and databases to determine what the log format means, matching codes to known vendor messages.

Create Structured Artifacts

Once parsed, each event becomes a **semantic frame** with all relevant fields extracted and standardized. Context fields like log source, message type, and severity level are inferred as part of this comprehensive frame.

No Information Loss: Even unrecognized fields are included in artifacts as auxiliary attributes rather than dropped. This comprehensive, self-learning approach handles heterogeneous data sources robustly, unlike brittle regex-based parsers that require constant manual updates.

WitFoo
CONDUCTOR



Repetitive, Unstructured Inputs

Conductor takes in syslog, agent and API data in diverse formats. Natural Language Processing (NLP) delivers ***message comprehension with no parsers***.

ProtoGraph Deduplication

Messages deduplicated by Six-Tuple:

- Client, Server, User, File (Relationships)
- Product Telemetry Source (Witness)
- Message Type and Intent (Purpose)

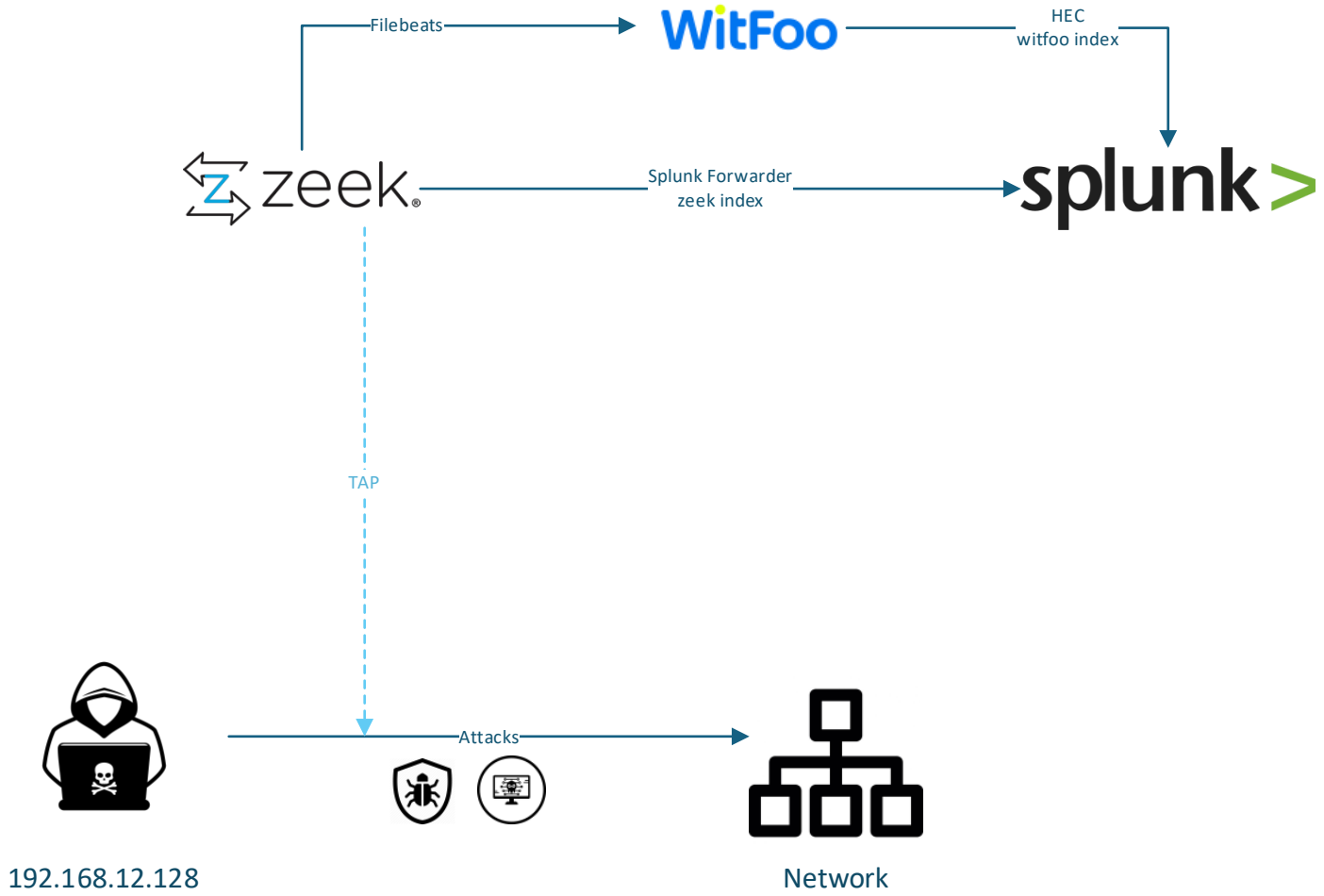
Reduced noise without evidence loss

Normalized, Structured Output

Standardized outputs allow for analytics, detection logic, and visualizations to be standardized across all data sources, ***improving understanding and minimizing SIEM/XDR upkeep.***

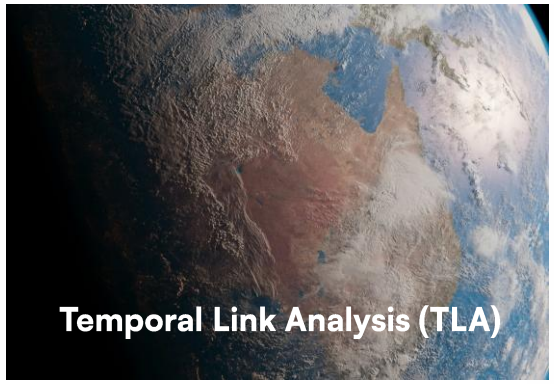
WitFoo's proprietary ETL capabilities ensure efficient, reliable, low-maintenance ingestion and prioritization of security signals at scale. The modular security architecture and automated workflows enable seamless integration and data normalization to **support rapid incident response across diverse environments and delivering unmatched TCO.**

Semantic Fingerprinting Demo



Stage 2: Dissonance Resolution

After Empathetic Listening normalizes and enriches events, Dissonance Resolution performs the critical work of correlation—building a comprehensive understanding of how events relate to each other across time and systems.

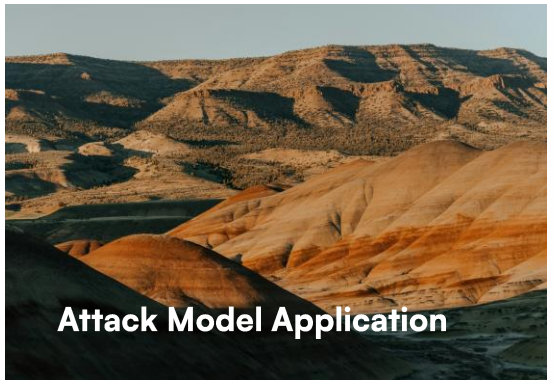


Temporal Link Analysis (TLA)

EP constructs a living knowledge graph where entities are nodes (users, hosts, IPs, files) and their relationships and events are edges. Every event updates this graph: "User X logged into Server Y" links the User X node to Server Y with a timestamped "login" edge.

Long-Horizon Correlation

TLA connects dots across extended time spans. It might link a Monday malware alert, a Wednesday suspicious connection, and a Friday data exfiltration if they involve related entities—revealing complex attacks that humans might miss when viewing events in isolation.



Attack Model Application

The system applies patterns and hypotheses—cybersecurity kill-chain models, theories of crime—to interpret the graph. When nodes and edges align with known attack patterns, it hypothesizes incidents like "Potential data exfiltration in progress."

Conflict Resolution

When data sources conflict (one says "threat removed" while another shows "threat still active"), EP identifies the inconsistency. It resolves conflicts through logic or flags them for human review, ensuring incident narratives remain internally consistent.

Through TLA, Empathetic Processing distills thousands of events into a handful of incident stories—grouping related alerts into single cases and eliminating duplicates or contradictions automatically.

Graph Enrichment

The knowledge graph is not static—it continuously grows richer with every event, creating an ever-deepening understanding of the security environment and entity relationships.

Continuous Enhancement

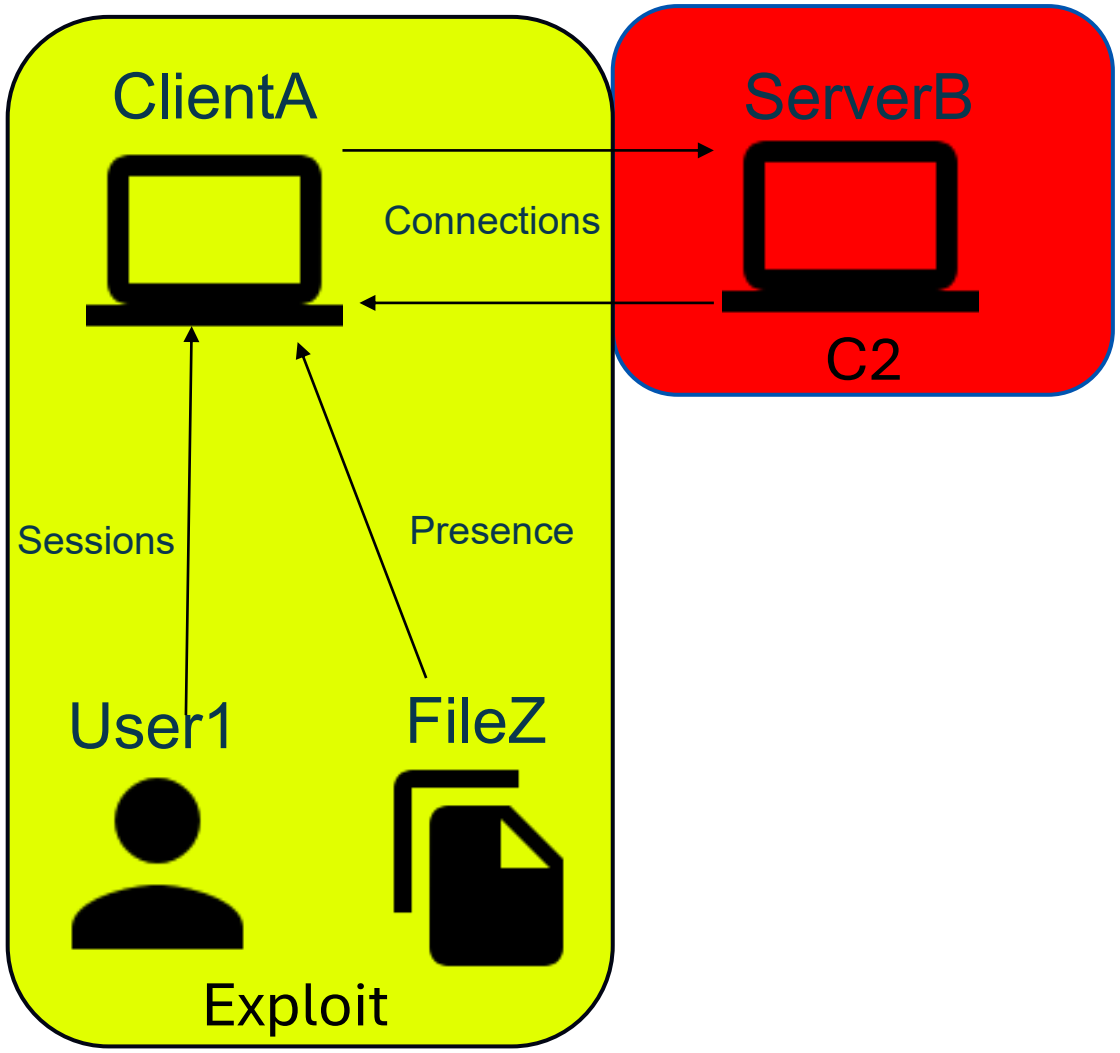
- Every new event enriches the graph by creating or updating nodes and edges for involved entities. If an event mentions an IP address, user, file, or other entity, the system ensures there's a node for it—creating one if needed—and links it appropriately.
- **Graph IDs in Artifacts:** Each event artifact from stage 1 receives tags referencing the graph entities it involves. An alert artifact carries IDs of related device and user nodes, directly linking raw data to graph context.

Powerful Capabilities

- **Graph-Enhanced Queries:** Find all events connected to a specific IP over six months
- **Path Discovery:** Trace relationships between any two entities
- **Pattern Recognition:** Identify recurring behaviors or anomalies
- **Temporal Analysis:** Understand how relationships evolve over time

Addressing Long-Term Attacks: Graph enrichment handles eventual consistency and slow-developing attacks brilliantly. Even when malicious activity is spread across weeks or months, the graph accumulates context so patterns can emerge after the fact. Attacks unfolding slowly are caught because the graph remembers and connects earlier steps that would be long forgotten in traditional systems.

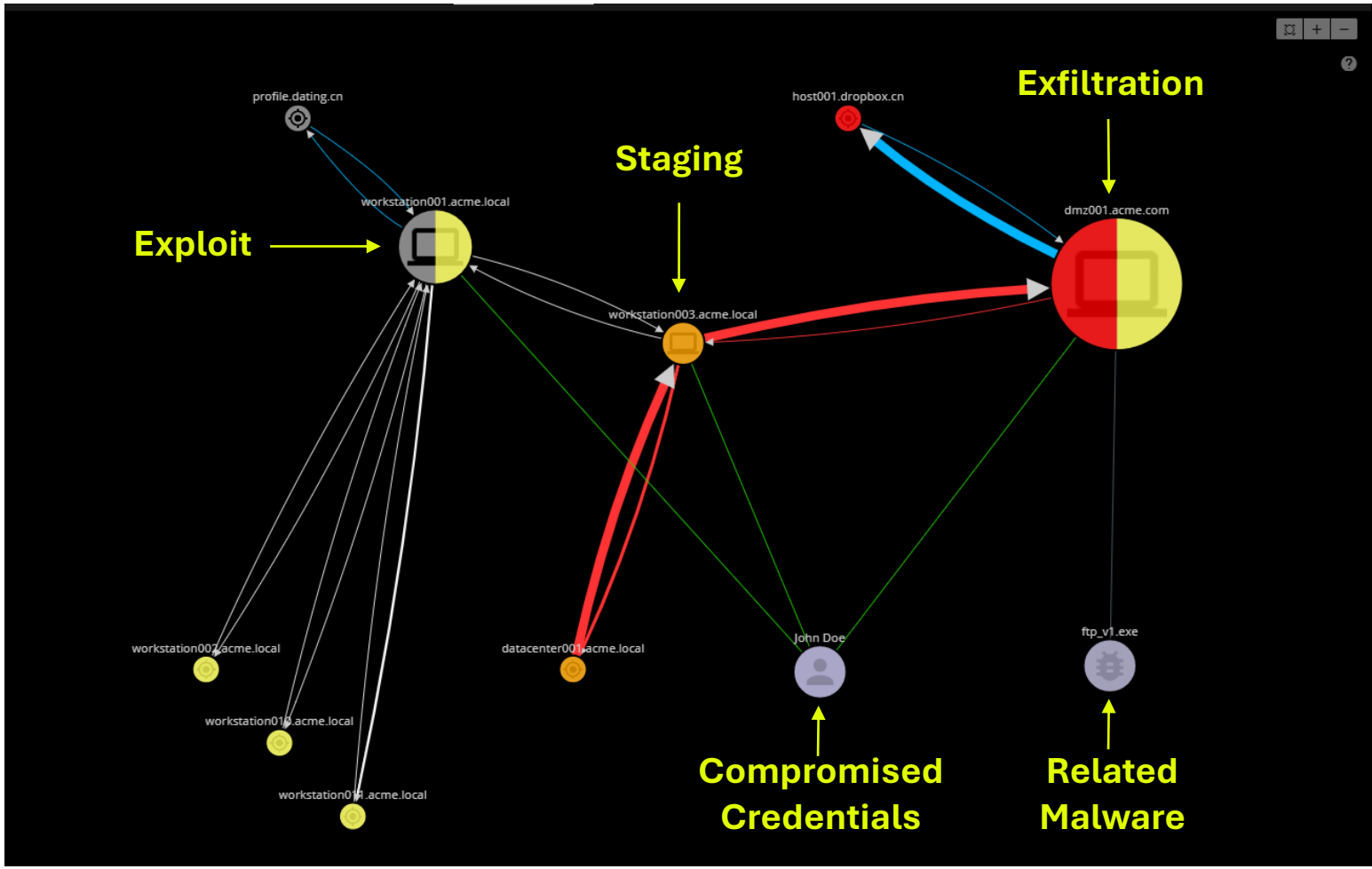
Graph Nodes & Edges



Artifacts
• ClientName: ClientA • ClientIP: 10.10.10.43 • ClientMAC: 00-DC-EF-23-15-12 • Product: MS DHCP • MessageType: DHCP Lease • Intent: Asset Info
• ClientName: ClientA • User: User1 • File: FileZ • Product: Crowdstrike Falcon • MessageType: Malware Detected • Intent: Exploit Detection
• ClientIP: 10.10.10.43 • ServerName: ServerB • Product: Cisco Firepower • MessageType: C2 Detected • Intent: C2 Detection

WitFoo Incident (*Temporal Link Analysis*)

- Meaningful Graph Relationships
- Modus Operandi of Attacker
- Combines, standardizes diverse data
- Learns from Investigator
- Consolidated Investigative unit
  - *Increases Clarity*
  - *Reduces Investigations*
  - *Reduces Time per Investigation*
- ***Unit of Work***



Advanced Analysis Techniques

Dissonance Resolution employs sophisticated methods to prioritize threats and model security incidents with remarkable precision and adaptability.

Suspicion Scoring

EP assigns **dynamic suspicion scores** to incidents and entities, gauging risk in real-time. Each event can increase or decrease the score: a malware detection adds high suspicion, while benign events add little. Scores decay over time (like a half-life) if no supporting evidence emerges—mimicking how human suspicions naturally fade.

Objective vs. Subjective Data

The system distinguishes between raw facts (objective observations) and calculated risk levels (subjective inferences). Machine learning helps adjust weighting: how quickly scores should decay, which event combinations are truly suspicious, and how to balance different indicators.

Object-Oriented Modeling

Incidents and entities are treated as objects with properties (suspicion level, current state) and behaviors. Using "game-like physics," incidents can escalate when threats "collide" with them or cool down over time. This simulation approach models evolving scenarios dynamically, adapting to new information in real-time.

Once data is structured and scored through these techniques, it becomes ready for advanced AI analysis. Large language models can review fully built incident objects to provide natural language assessments or answer complex questions—all grounded in factual, structured data that prevents hallucination.

Dissonance Resolution Demo

Stage 3: Empathetic Speaking

Empathetic Speaking transforms complex security analyses into clear, actionable communication tailored to each audience's needs and expertise level. The same incident data generates vastly different outputs for different stakeholders.

For Security Analysts

- Detailed incident reports with narrative timelines: "At 08:30 UTC, user JohnDoe had 15 failed login attempts from IP X (possible brute force); at 08:31 a login succeeded from that IP, followed by access to sensitive files and an outbound FTP transfer to known malicious server..."
- All evidence is compiled chronologically with context, enrichment data, and recommended next steps—enabling analysts to understand and respond immediately.

For Executives & Managers

- High-level dashboards and summaries with one line per incident: "Compromised account led to data exfiltration—contained; ~50MB of finance data at risk." Includes metrics like incident counts this week, average response times, and false positive rates.
- Decision-makers receive clear security posture overviews without technical noise, enabling informed strategic decisions.

For Compliance & Audit

- Complete evidence logs and audit trails showing who did what during response, with full linkage between processed results and raw data. Automated documentation proves events were handled correctly.
- Predestination ensures nothing was dropped—every regulatory requirement is met by design, with instant report generation capabilities.

Empathetic Speaking reduces back-and-forth communication, smooths hand-offs during shift changes or escalations, and ensures everyone—regardless of technical expertise—understands the security story being told.

Reporting Demo

Benefits of Empathetic Processing

Empathetic Processing delivers transformative improvements across every dimension of security operations, fundamentally changing how organizations detect, analyze, and respond to cyber threats.

>90%

Alert Reduction

Orders of magnitude reduction in alerts presented to humans through intelligent correlation and duplicate elimination

50%+

Faster Response

Dramatic decrease in Mean Time to Resolution through ready-made incident narratives with full context

24/7

Consistent Quality

Automated correlation ensures thorough, accurate analysis regardless of time, shift, or analyst workload

Operational Excellence

- **Eliminate Alert Fatigue:** Analysts focus on real incidents, not noise, dramatically improving job satisfaction and retention
- **Catch Complex Threats:** Graph correlation detects multi-stage and stealthy attacks that siloed alerts would miss
- **Reduce Human Error:** Consistent logic application prevents oversights and ensures thorough investigation

Strategic Advantages

- **Superior Forensic Readiness:** Predestined data organization makes investigations and compliance audits effortless
- **Scale Effectively:** Handle increasing data volumes without proportional increases in staff
- **Future-Ready Platform:** Structured data enables advanced AI integration and continuous improvement

Organizations implementing Empathetic Processing report not just incremental improvements, but fundamental transformations in their security operations—doing more with less while catching threats that previously went undetected.

Implementation: The WitFoo Precinct Platform

Empathetic Processing is not theoretical—it has been successfully implemented in the **WitFoo Precinct** security platform, validating these concepts in real-world production environments.

Conductor: Empathetic Listening Engine

Adaptive ingest engine performing Empathetic Listening through **adaptive context parsing** with semantic fingerprints. Automatically normalizes any log format it encounters—deployments have identified thousands of unique log formats via fingerprinting, all handled without manual parser configuration.

Precinct: Correlation & TLA Engine

Builds and maintains the knowledge graph, running continuous Temporal Link Analysis. Forms incidents by matching graph patterns—brute force followed by successful login triggers "Credential Compromise" incident. Calculates suspicion scores to intelligently prioritize alerts for analyst attention.

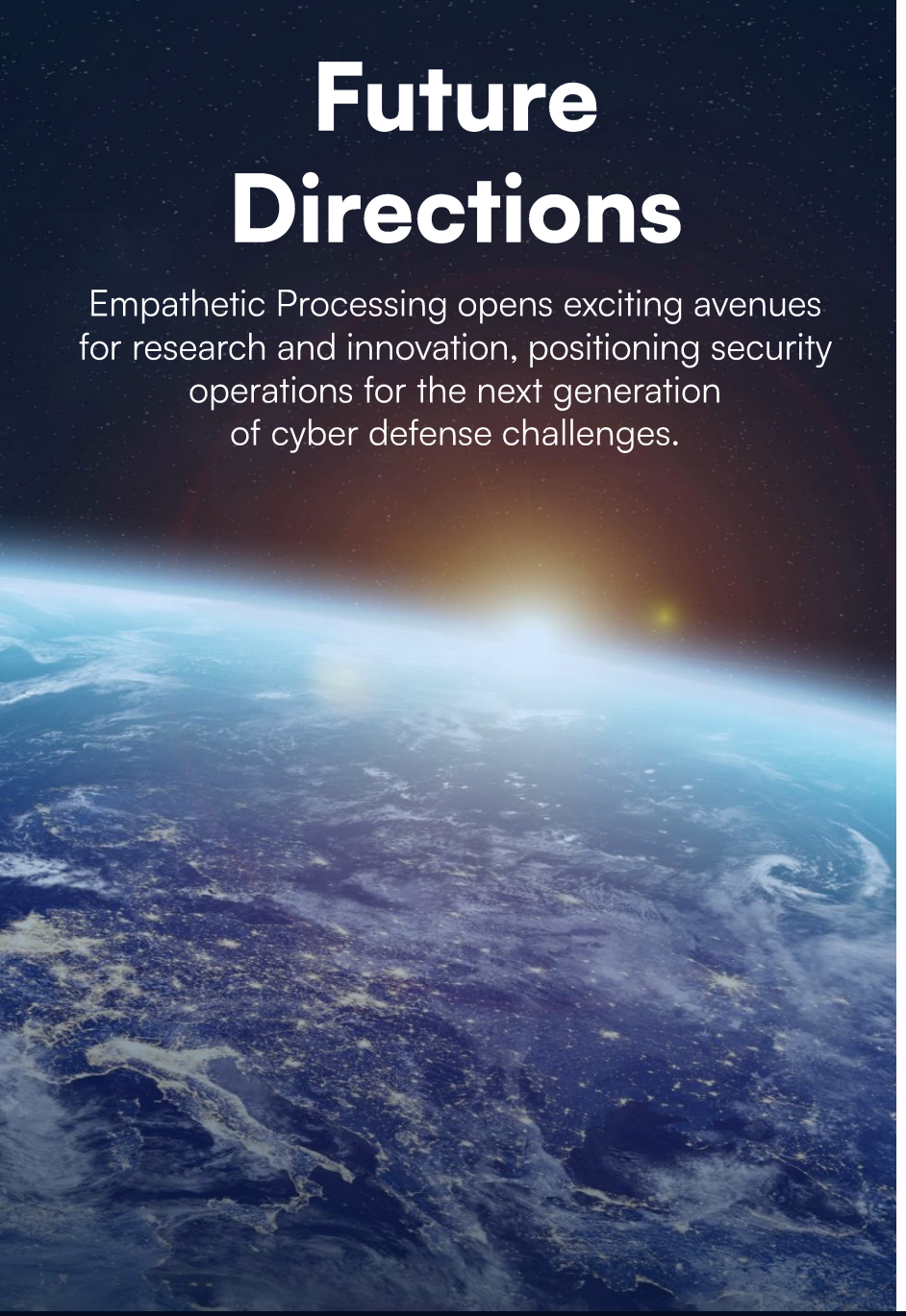
Reporting: Empathetic Speaking Module

Automates output generation with narrative timelines for analysts, executive summaries for management, and compliance reports for auditors. Uses template-driven narratives customized for each audience while maintaining consistency across the organization.

Proven Results: Side-by-side testing showed the EP-driven system greatly reduced alerts analysts see—clustering millions of raw events into handfuls of high-fidelity incidents. All genuine threats were caught. Some deployments observed over 90% reduction in alerts presented with significantly faster investigation times.

Future Directions

Empathetic Processing opens exciting avenues for research and innovation, positioning security operations for the next generation of cyber defense challenges.



- **LLM INTEGRATION:** Large Language Models working with EP's structured data to answer complex questions in natural language—grounded in facts, preventing hallucination.
- **CROSS-DOMAIN FUSION:** Applying EP principles to other data-intensive domains: fraud detection, network operations, IoT security.
- **AUTONOMOUS RESPONSE:** AI-driven remediation actions based on EP's comprehensive understanding—from containment to evidence preservation.
- **CONTINUOUS LEARNING:** Feedback loops where analyst corrections refine correlation patterns and suspicion thresholds—systems that learn from every interaction.
- **BROADER CONTEXT:** Integration beyond IT logs: HR events, physical security, global threat intel—enabling insider threat detection and predictive defense.
- **COLLABORATIVE DEFENSE:** Organizations sharing anonymized patterns and fingerprints for collective threat intelligence—federated learning for security.

These future directions maintain EP's core philosophy: keeping humans at the center while leveraging technology to amplify human capability rather than replace human judgment.

Conclusion: A New Paradigm for Security

Empathetic Processing represents a fundamental shift in how we approach cybersecurity analytics. By modeling the data pipeline on human communication and reasoning, we bridge the critical gap between overwhelming data volumes and actionable intelligence.

Reduction in Alert Fatigue

Analysts focus on genuine incidents rather than drowning in noise. The system does the heavy lifting of correlation and context-building, freeing human experts to apply their judgment where it matters most.

Human-Centric Automation

AI that thinks like an analyst—not just a log aggregator. Empathetic Processing doesn't replace human expertise; it amplifies it by handling the mechanical, time-consuming work that computers do best while preserving the critical thinking humans do best.

Reactive & Unmanageable

A platform for integrating advanced AI capabilities (NLP, LLMs, machine learning) and scaling security operations as threats evolve. The structured, semantic approach enables continuous innovation and adaptation.

As cyber threats grow in volume and sophistication, approaches like Empathetic Processing offer a path to scale defenses in a human-aligned way.
We can finally escape the "data-rich, information-poor" trap and build security operations that are both more effective and more sustainable.

The future of cybersecurity is not about collecting more data—it's about understanding the data we have.

Q&A

Resources

WitFoo Provides the following Educational Resources

- CharlesHerring.com – This talk, deck & whitepaper
- Huggingface.com/witfoo – OpenSource Models, Datasets
- ArtiFish.dev – Python toolkit for GenAI in cybersecurity
- WitFoo Educational Initiative – No-cost software, demo environments, datasets and training resources
- WitFoo.Zendesk.com – Support, Training and Demo resources

Charles D. Herring, WitFoo co-Founder
Charles@WitFoo.com
https://CharlesHerring.com