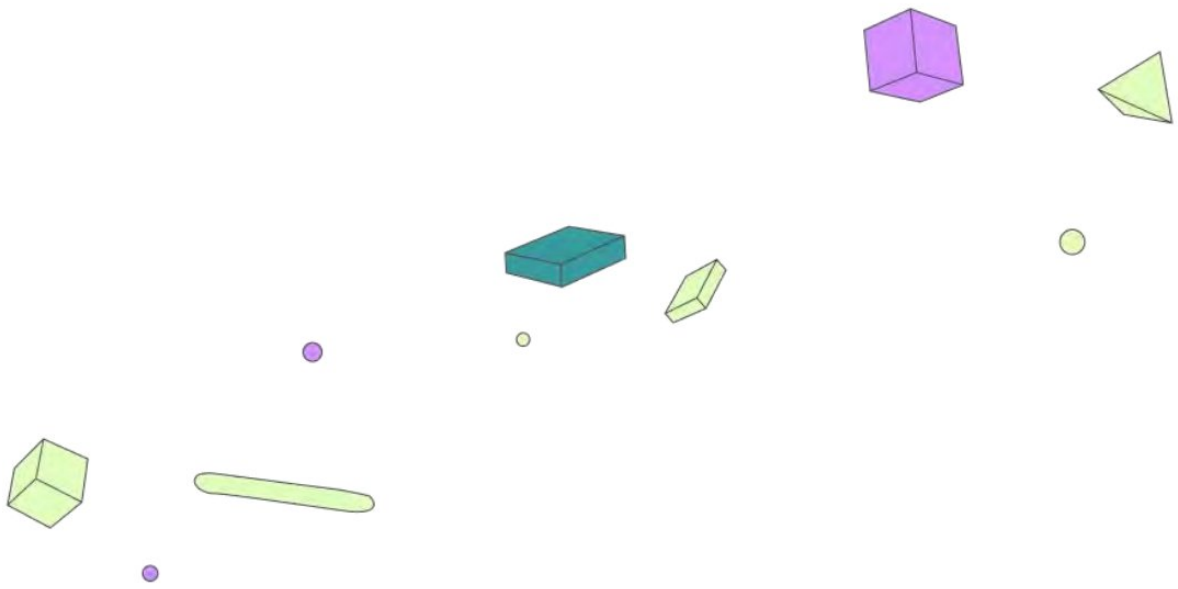
INTELLIGENCE REPORT

Data Combinations Intelligence Report

How toxic data combinations create outsized risk,
how synergistic data combinations create outsized value,
and what security teams should do about both

cyera

Table of Contents



Executive Summary

Individual data elements are routinely classified and protected based on their standalone sensitivity, but sensitivity is not additive: it's combinatorial. This report examines 24 data combinations drawn from real, publicly documented incidents and business outcomes: 12 that turn ordinary, individually low-sensitivity data into critical exposure ("toxic" combinations), and 12 that turn the same categories of data into measurable competitive advantage when properly governed ("synergistic" combinations). The throughline runs in both directions: traditional, signature-based controls evaluate data elements in isolation, and consistently miss what happens when those elements are placed next to each other.

Read together, they make the case that combinatorial risk and combinatorial value are two sides of the same governance problem.

Toxic Combinations

12	15+	7	5	$15B+
Toxic Combinations	Industries Affected	Critical Risk	High Risk	Documented Losses

Synergistic Combinations

12	15+	12/12	48	$50B+
Synergistic Combinations	Industries	High Value Impact	Real-World Examples	Documented Annual Gains

Key Insight

The combinations that create the most risk when exposed are frequently the same combinations that create the most value when properly governed. Customer and contract data, HR and compensation data, pricing and inventory data, and clinical and patient data all show up on both sides of this report. The difference is not the data; it's the governance around it, and it's exactly where Cyera helps: finding, governing, and safely enabling these high-value data combinations.

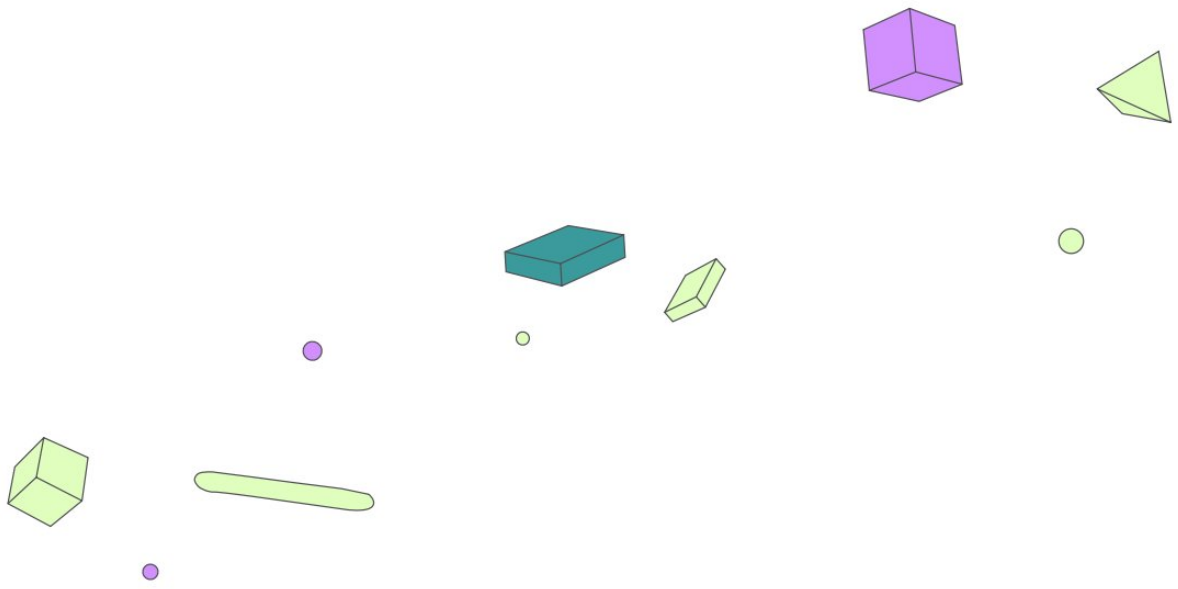
Introduction

Why Data Combinations Matter

Most data security programs are built around the sensitivity of individual data elements: a name is PII, a diagnosis code is PHI, a source code repository is intellectual property. Classification, DLP rules, and access policies are tuned to catch these elements when they appear on their own. That approach works well for single-element exposure. It breaks down the moment two or three individually unremarkable data elements land in the same place.

A price list is operational data. A warehouse manifest is operational data. Neither triggers a PII or PCI control. Together, they reveal forward-looking revenue signals that can move markets before an earnings call. A performance rating is routine HR data. A layoff list is routine planning data. Together, they can become evidence of a discriminatory pattern or an early, actionable signal of a workforce reduction. The same logic runs in the opposite direction: product usage data and support ticket history are unremarkable data sets on their own, but combined, they power the churn-prediction models that some of the world's largest subscription businesses rely on.

This report treats that combinatorial effect as the primary subject, not a footnote. The Toxic Data Combinations section catalogs 12 combinations that become dangerous when exposed or misused. The Synergistic Data Combinations section catalogs 12 combinations that become valuable when properly governed: the same underlying principle, pointed at value creation instead of risk. The sections that follow draw out the patterns that connect them, translate those patterns into what security and data teams should watch for, and show how Cyera's Data & AI Security Platform, built to understand context and proximity rather than pattern-match individual fields, closes the gap that traditional tools leave open. Each of the next two sections closes with a look at how Cyera addresses that side of the equation directly.



Toxic Data Combinations

The twelve combinations below are drawn from documented, publicly reported incidents and enforcement actions. In each case, the individual data elements were unremarkable, or at least not obviously dangerous, on their own. Combined, they enabled insider trading, IP theft, discrimination claims, physical security threats, or attacks on critical infrastructure. The real-world examples below focus on the pattern itself rather than any single organization's experience.

01 Pricing Data + Supply Chain Manifests

Critical | Retail | Manufacturing | >$100M Impact

Product List Prices + Warehouse Manifests

Why This Is Toxic

Individual elements are operational data with low sensitivity on their own. Combined, they reveal forward-looking revenue signals before earnings, enabling insider trading (SEC Rule 10b-5) and market manipulation using material non-public information (MNPI).

Detection Difficulty

Very High. Neither element carries a sensitive data pattern, so detection requires understanding business context rather than matching a known format.

Real-World Examples

A global consumer electronics manufacturer	Multiple supply chain partners leaked product specifications and inventory data **Stock volatility and competitive disadvantage worth billions**
A dispute between a semiconductor supplier and a global consumer electronics manufacturer	Intellectual property and pricing disputes tied to supply chain data **$4.5B settlement**

AI Amplification Risk

LLMs can correlate manifest volumes with pricing to generate trading signals in seconds, work that previously required analyst teams.

02 Clinical Trial Data + Patient Identifiers + Trial Sites

Critical | Pharma | Healthcare | >$100M Impact

Drug Efficacy/Adverse Events + Patient PII + Geographic Trial Sites

Why This Is Toxic

Trial results alone are confidential but not PII. Patient IDs alone are protected but not market-moving. Combined with site data, the result is identifiable patients plus tradeable intelligence, implicating HIPAA, FDA 21 CFR Part 11, and stock manipulation risk.

Detection Difficulty

High. Trial data often uses custom formats, and patient IDs may be coded but become reversible once combined with site data.

Real-World Examples

A global pharmaceutical manufacturer and its vaccine development partner	Clinical trial documents and regulatory submissions were exposed at a European medicines regulator **$10B+ market cap fluctuation during vaccine approval**
A public-sector health program	4.9 million patient records were exposed, including clinical trial participants **Regulatory delays, patient privacy violations**

AI Amplification Risk

AI can correlate adverse-event patterns across sites to predict trial success or failure, enabling trading ahead of formal announcements.

03 Source Code + API Keys/Credentials + Cloud Architecture

Critical | Technology | Fintech | >$100M Impact

Production Source Code + Embedded Credentials + Infrastructure Topology

Why This Is Toxic

Code alone is an IP theft risk. Credentials alone are an access risk. Architecture alone has reconnaissance value. Combined, they form a turnkey breach kit: full attack-chain enablement, with credentials providing access, code revealing vulnerabilities, and architecture showing lateral-movement paths.

Detection Difficulty

Medium. Credentials are detectable, but they often sit in configuration files that look routine, and architecture documentation is rarely flagged.

Real-World Examples

A widely used IT infrastructure software vendor	A nation-state threat actor combined source code, build-system architecture, build-server credentials, and code signing in a supply chain compromise **$40M+ direct remediation, $100B+ ecosystem impact, 18,000+ organizations affected**
A widely used password management vendor	A developer laptop compromise led to source code exposure, then vault architecture exposure, then a customer vault breach **Class action pending, significant customer attrition**

AI Amplification Risk

LLMs can analyze code for vulnerabilities, map credentials to services in architecture diagrams, and generate exploit chains automatically.

04 M&A Target Lists + Financial Valuations + Deal Timelines

Critical | Investment Banking | PE/VC | >$100M Impact

Acquisition Target Names + Valuation Models + Projected Close Dates

Why This Is Toxic

Target lists are sensitive but not tradeable alone. Valuations are confidential but speculative. Timelines are operational. Combined, they become actionable trading intelligence, raising exposure under securities fraud (trading on MNPI), breach of fiduciary duty, and antitrust review.

Detection Difficulty

Very High. These are ordinary business documents, not PII, often found in slide decks and spreadsheet models that look like normal work product.

Real-World Examples

A prominent hedge fund and its founder	Board-level M&A intelligence was sourced from directors at major financial and technology firms **$92.8M forfeiture, 11 years' imprisonment**
A global technology manufacturer's acquisition of an enterprise software company	Material information was withheld during an $11B acquisition **$5B writedown**
A Big Four accounting firm	Employees obtained confidential inspection data from a public company audit oversight regulator **$50M penalty**

AI Amplification Risk

AI can monitor document repositories for M&A signals, correlate across deal-team communications, and surface patterns before formal announcements.

05 Employee Performance + Compensation + Termination Plans

High | All Industries | >$100M Impact

Performance Ratings + Salary/Equity Details + RIF/Layoff Lists

Why This Is Toxic

Performance data is sensitive HR data. Compensation is confidential but not actionable alone. Termination lists are operational planning. Combined, they can evidence discriminatory patterns or provide early warning of layoffs, raising exposure under employment litigation, discrimination claims, and the WARN Act.

Detection Difficulty

Medium. HR data is typically flagged, but the combination across systems is often missed.

Real-World Examples

A global electric vehicle manufacturer	75,000+ worker records were leaked, including performance reviews, salaries, and national ID numbers **$3.3B potential GDPR fine exposure**
A major entertainment studio	Leaked salary data revealed gender pay gaps, performance reviews, and planned layoffs **$15M settlement, $35M+ remediation**
A global social media platform	Layoff lists were leaked before required notifications were issued **$500M+ ongoing litigation**

AI Amplification Risk

AI can analyze patterns to identify potential discrimination (age, gender, race correlations with terminations) faster than plaintiffs' attorneys.

06

Grid Infrastructure + SCADA Credentials + Maintenance Schedules

Critical | Energy | Utilities | >$100M Impact

Grid Topology/Substation Maps + OT System Credentials + Planned Outage Windows

Why This Is Toxic

Grid maps are sensitive but often available through public records requests. SCADA credentials are high risk but useless without context. Maintenance schedules are operational. Combined, they form an attack playbook with timing, raising national security risk, NERC-CIP violations, and physical safety threats.

Detection Difficulty

High. OT data is often siloed from IT-side DLP tooling, and maintenance schedules look routine.

Real-World Examples

A national power grid operator	Attackers combined grid topology, SCADA credentials, and operator shift schedules **230,000+ customers affected, $100M+ in damages**
A petrochemical processing facility	A safety instrumented system attack used network architecture combined with controller credentials **The attack was disrupted, but it demonstrated nation-state physical destruction capability**

AI Amplification Risk

AI can optimize attack timing by correlating maintenance windows with grid load patterns to maximize impact.

07

Customer Lists + Contract Values + Renewal Dates

High | SaaS | B2B | >$100M Impact

Named Customer Accounts + ARR/TCV Amounts + Contract Expiration

Why This Is Toxic

Customer names are business-confidential. Contract values are sensitive. Renewal dates are operational. Combined with competitor notes, the result is an actionable competitive playbook that enables perfectly timed competitive poaching.

Detection Difficulty

Very High. CRM data is rarely flagged, since it looks like normal sales operations.

Real-World Examples

A healthcare software vendor and a global IT services provider	A trade secret theft case involving customer data and contract terms **$940M jury verdict, later reduced to $420M**
An autonomous vehicle technology company and a global ride-hailing platform	14,000 files were stolen, including customer contracts and pricing **$245M settlement, 18 months' imprisonment**

AI Amplification Risk

AI can prioritize targets by value and timing, generate personalized outreach, and predict churn signals from competitor mentions.

08 Patient Diagnoses + Treatment Protocols + Insurance Claims

Critical | Healthcare | Insurance | >$100M Impact

ICD-10 Diagnosis Codes + Treatments/Drugs + Billing/Claims Data

Why This Is Toxic

ICD-10 codes are clinical but coded. Treatments reflect standard of care. Claims are financial. Combined with identifiers, the result is a complete medical and financial profile of an individual, implicating HIPAA, insurance fraud detection, and discrimination risk.

Detection Difficulty

Medium. PHI detection exists, but cross-system correlation is often missed.

Real-World Examples

A regional health insurance provider	11 million records combined diagnoses, claims history, and clinical treatment data **$74M settlement, $10M in state penalties**
A European psychotherapy clinic	A breach exposed therapy session notes together with patient identities **Clinic bankruptcy, individual patient extortion**
An academic medical center	4.5 million records included diagnoses linked to treatments **$7.5M settlement**

AI Amplification Risk

AI can correlate diagnosis patterns with treatment costs to identify high-value fraud targets or predict patient outcomes for underwriting purposes.

09 Passport/Visa Data + Travel Itineraries + Executive Calendars

High | Multinational | Professional Services | >$100M Impact

Passport Numbers/Visa Status + Flight/Hotel Bookings + Executive Meeting Schedules

Why This Is Toxic

A passport number is PII but not location-specific. Itineraries are operational. Calendars are business-sensitive. Combined, they enable real-time executive location tracking, a physical security threat, kidnapping risk, and corporate espionage opportunity.

Detection Difficulty

Medium. Passport data is detected, but travel data often sits in email and calendars that aren't scanned.

Real-World Examples

Executive kidnapping risk	Executive kidnappings leveraging travel intelligence remain an ongoing risk in several regions **$500M to $1B annually in ransom, security, and insurance costs**
A high-profile technology executive	A foreign intelligence service allegedly exploited personal data to track the executive's movements **This demonstrated nation-state interest in executive targeting**

AI Amplification Risk

AI can correlate travel patterns with public announcements to predict non-public executive movements, enabling physical threats or M&A intelligence.

10

Trading Algorithms + Position Data + Execution Timestamps

Critical | Hedge Funds | Prop Trading | >$100M Impact

Proprietary Trading Algos + Current Portfolio Positions + Trade Execution Times

Why This Is Toxic

Algorithms are IP but theoretical. Positions are confidential but point-in-time. Timestamps are operational. Combined, they enable front-running or strategy replication: alpha strategies worth billions if reverse-engineered.

Detection Difficulty

Very High. Code and data resemble normal quantitative work, and no PII patterns are present.

Real-World Examples

A global investment bank	A former employee stole high-frequency trading code worth an estimated $300M in annual revenue **Criminal prosecution; code valued at $300M per year**
A major hedge fund and a competing trading firm	A former employee allegedly took proprietary trading algorithms to a competitor **$20M settlement**
A quantitative hedge fund	A former employee stole proprietary trading code **Damages in the tens of millions**

AI Amplification Risk

AI can reverse-engineer strategies from execution patterns, predict future trades, and enable systematic front-running at scale.

11 Product Formulations + Supplier Contracts + Raw Material Costs

High | CPG | Pharma | Chemicals | >$100M Impact

Manufacturing Recipes/Formulas + Supplier Agreements + Commodity Pricing

Why This Is Toxic

Formulas are trade secrets but often live in routine documents. Supplier contracts are confidential. Costs are financial. Combined, the result is a full recipe plus the economics needed to replicate a product with cost optimization built in.

Detection Difficulty

High. Trade secrets rarely follow consistent patterns, and cost data looks like ordinary accounting.

Real-World Examples

A global chemicals manufacturer and a competing manufacturer	A trade secret theft case involving a proprietary synthetic fiber **$919.9M jury verdict**
A wind energy technology company and a competing manufacturer	Turbine software and designs were stolen **$1.5B market cap loss, $59M criminal fine**
A global beverage manufacturer	An employee attempted to sell a proprietary formula to a competitor before federal law enforcement intervened **The formula, valued at $2B+, remained protected**

AI Amplification Risk

AI can optimize formulations using ingredient costs, identify supplier vulnerabilities, and predict margin impacts of commodity shifts.

12 Board Meeting Minutes + Executive Changes + Strategic Pivots

Critical | Public Companies | PE-Backed | >$100M Impact

Board/Committee Transcripts + CEO/CFO Succession Plans + Divestiture Plans

Why This Is Toxic

Board minutes are confidential but routine governance. Succession discussions are sensitive HR matters. Pivots are strategic. Combined, they represent material information ahead of public filings, creating MNPI exposure for trading, activist-defense risk, and employee morale impact.

Detection Difficulty

Very High. These are privileged documents with no PII, often stored in secure portals but synced locally.

Real-World Examples

A prominent hedge fund	A case of systematic insider trading, including board-sourced tips **$1.8B in total penalties**
A major credit reporting agency	An executive sold $1M in stock after learning of a breach, before public disclosure **Criminal charges**
A global technology manufacturer's acquisition of an enterprise software company	The board withheld material information during an $11B acquisition **$5B writedown and ongoing litigation**

AI Amplification Risk

AI can analyze tone and sentiment in board materials to predict announcements, correlate with executive public activity, and identify patterns ahead of disclosure.

Regulatory Frameworks Implicated

The toxic combinations above touch at least eight distinct regulatory frameworks, often more than one at a time. The table below maps each framework to the specific risk it addresses.

SEC Rule 10b-5

Insider trading, MNPI violations

HIPAA

PHI combined with claims/billing data

GDPR

HR data exposure, cross-border transfers

NERC-CIP

Critical infrastructure protection

FDA 21 CFR Part 11

Clinical trial data integrity

WARN Act

Premature layoff disclosure

Trade Secret Law

DTSA and state trade secret statutes

SOX

Financial statement fraud, internal controls

How Cyera Detects Toxic Combinations

Traditional DLP misses sophisticated combinations because it relies on pattern matching. Cyera's Data & AI Security Platform understands context instead.

01

Deep Classification

500+ data classes with business context (M&A, clinical, HR). Understands data subject roles, geographic residency, and business associations that determine if combinations are toxic.

02

Proximity Association

Links related data elements across columns, files, and datastores. Detects when pricing data co-occurs with supply chain manifests or when board minutes reference executive changes.

03

AI Guardian

Real-time prompt blocking for AI systems. Prevents LLMs from correlating toxic combinations by validating every prompt before it reaches leading AI copilots, chatbots, or custom AI agents.

Synergistic Data Combinations

Key Insight

The same combinations that create the most risk when exposed are the same combinations that create the most value when properly governed. The 12 combinations below show what happens when organizations find, govern, and safely enable these pairings rather than leaving them ungoverned.

The real-world examples below describe verified, publicly reported programs, focused on the pattern each one demonstrates rather than the organization behind it.

01 Customer Behavior + Product Usage + Support History

High Value | SaaS | Retail | >$100M Impact

Customer Interaction Logs + Feature Usage Telemetry + Support Ticket History

Value Unlocked

Predictive Churn Modeling: identify at-risk accounts six-plus months before cancellation

Personalized Recommendations: drive upsell/cross-sell with AI-powered suggestions

Proactive Support: resolve issues before tickets are filed

Real-World Examples

A global credit reporting and data analytics company	Combines credit behavior, payment history, and demographic data **Drives up to 15% more revenue for clients through better targeting**
A global music streaming platform	Listening behavior, skip patterns, and playlist creation **40% of all listening comes from algorithmic recommendations**
A global video streaming platform	Behavioral, viewing, and engagement data **$1B+ in annual savings from reduced churn**

AI/BI Use Cases

Churn prediction models, next-best-action engines, automated health scores, personalized onboarding flows, proactive outreach triggers.

02 Sales Pipeline + Marketing Attribution + Win/Loss Analysis

High Value | B2B | SaaS | >$100M Impact

CRM Opportunity Data + Campaign Touchpoints + Deal Outcomes with Reasons

Value Unlocked

Marketing ROI Optimization: know which campaigns drive closed revenue, not just leads

Forecast Accuracy: predict close rates by source and segment

Rep Coaching: identify winning behaviors vs. losing patterns

Real-World Examples

A conversation intelligence platform	Call data, deal outcomes, and rep behavior analysis **27% increase in win rates for customers**
A predictive B2B targeting platform	Intent data, pipeline, and outcomes **2x pipeline conversion through AI targeting**
A revenue operations platform	Pipeline, activity, and outcomes **20% more accurate forecasts**

AI/BI Use Cases

Multi-touch attribution modeling, lead scoring, deal scoring, forecast modeling, pipeline coverage analysis, campaign optimization.

03 Inventory Levels + Sales Velocity + Supplier Lead Times

High Value | Retail | E-commerce | >$100M Impact

Real-time Inventory Positions + Historical Sales Rates + Supplier Delivery Windows

Value Unlocked

Perfect Demand Forecasting: predict stockouts before they happen

Working Capital Optimization: reduce overstock by 20–30%

Supplier Negotiation: data-driven terms based on actual lead-time performance

Real-World Examples

A global e-commerce and logistics company	Combines inventory, velocity, and supplier data at massive scale **35% of revenue from recommendations plus industry-leading 1–2 day delivery**
A multinational retail chain	Point-of-sale, inventory, and supplier data platform **16% reduction in out-of-stocks on $600B in revenue**
A global fast-fashion retailer	Sales velocity, inventory, and manufacturing lead times **Two-week design-to-shelf cycle vs. an industry average of six months, on $30B in revenue**

AI/BI Use Cases

Demand forecasting, automated reorder points, safety stock optimization, supplier scorecards, markdown optimization, seasonal planning.

04 Employee Skills + Project Outcomes + Engagement Scores

High Value | All Industries | >$100M Impact

Skills Assessments + Project Performance Metrics + Employee Satisfaction Data

Value Unlocked

Optimal Team Composition: staff projects with the right skill combinations

Succession Planning: identify high-potential leaders early

Retention Modeling: predict flight risk before resignation

Real-World Examples

A global technology and consulting company	Skills, performance, and engagement prediction **95% accuracy predicting flight risk, an estimated $300M saved annually in retention**
A global technology company	Performance, engagement, and team-composition analysis **Identified eight habits of effective managers, improving management quality across 37,000 employees**
An enterprise HR software vendor	Skills, performance, and engagement combination **40% reduction in time-to-hire through internal mobility**

AI/BI Use Cases

Team assembly optimization, internal mobility matching, flight-risk prediction, compensation benchmarking, diversity analytics, skills-based hiring.

05

Financial Actuals + Operational Metrics + Market Data

High Value | All Industries | >$100M Impact

P&L / Balance Sheet Data + Operational KPIs + External Market Indicators

Value Unlocked

Driver-Based Forecasting: understand what operational levers move financial outcomes

Scenario Modeling: stress-test strategies against market conditions

Performance Attribution: separate execution from market tailwinds and headwinds

Real-World Examples

A data analytics platform deployed at a global aerospace manufacturer	Merged operational, financial, and external data **30% reduction in production time, billions of dollars in savings**
A global asset management firm	Portfolio, market, and risk data integration **Manages $21.6 trillion in assets with integrated analytics**
A process mining software vendor	ERP financials combined with operational process data **15–20% working capital improvement for customers**

AI/BI Use Cases

Rolling forecasts, variance analysis, what-if modeling, capital allocation optimization, pricing optimization, M&A target evaluation.

06

Customer Feedback + Product Roadmap + Competitive Intel

High Value | Technology | Consumer Products | >$100M Impact

NPS/CSAT/Reviews + Planned Feature Priorities + Competitor Feature Matrices

Value Unlocked

Feature Prioritization: build what customers want and competitors lack

Market Gap Identification: find unmet needs before competitors do

Churn Prevention: address complaints before they become cancellations

Real-World Examples

A product analytics platform	Product usage, feedback, and roadmap integration **28% increase in feature adoption through data-driven prioritization**
An experience management platform	Customer feedback, operational data, and competitive benchmarks **25% improvement in customer retention**
A competitive intelligence platform	Market data, product positioning, and win/loss data **15% win-rate improvement against tracked competitors**

AI/BI Use Cases

Feature request clustering, sentiment analysis, competitive gap analysis, roadmap prioritization scoring, launch timing optimization.

07

Pricing History + Customer Segments + Competitive Rates

High Value | Retail | Travel | SaaS | >$100M Impact

Historical Transaction Prices + Customer Cohort Data + Market Pricing Intelligence

Value Unlocked

Dynamic Pricing: optimize price by segment, time, and demand in real time

Margin Optimization: identify where revenue is being left on the table

Competitive Positioning: price strategically against alternatives

Real-World Examples

A global e-commerce and logistics company	Demand, competitor, and segment data, repricing millions of times per day **25% margin improvement on optimized items**
A global ride-hailing platform	Demand, supply, and time data **$2B+ in additional annual revenue during peak times**
A global online travel booking platform	Inventory, demand, and competitive rate data **90%+ gross profit margins through dynamic pricing**

AI/BI Use Cases

Price elasticity modeling, discount optimization, promotional effectiveness, competitive price monitoring, revenue management systems.

08 Claims Data + Treatment Outcomes + Cost Benchmarks

High Value | Healthcare | Insurance | >$100M Impact

Insurance Claims History + Clinical Outcome Measures + Regional/Provider Cost Data

Value Unlocked

Care Pathway Optimization: identify the most effective treatments per condition

Fraud Detection: spot anomalous billing patterns at scale

Value-Based Care: tie reimbursement to outcomes, not volume

Real-World Examples

A large health insurance and health services company	Claims, outcomes, and cost data across 100M+ covered lives **10–15% medical cost reduction for clients, hundreds of millions saved**
A health insurance technology company	Claims, outcomes, and engagement data **20% lower hospitalization rates vs. the Medicare average**
A precision medicine data company	Clinical, genomic, and outcomes data **Better outcomes through data-driven treatment matching**

AI/BI Use Cases

Care gap identification, prior authorization optimization, provider quality scoring, population health management, drug utilization review.

09

Manufacturing Telemetry + Quality Data + Maintenance Logs

High Value | Manufacturing | Energy | >$100M Impact

IoT Sensor Data + Quality Inspection Results + Maintenance History

Value Unlocked

Predictive Maintenance: prevent failures before they happen; 30–50% downtime reduction

Yield Optimization: identify root causes of quality defects

Asset Lifecycle Management: optimize replacement timing

Real-World Examples

An industrial predictive analytics vendor	Sensor, maintenance, and performance data **36% reduction in unplanned downtime plus a 9% production increase**
A global aerospace engine manufacturer	Engine telemetry, maintenance, and performance data **99.9% on-wing availability, billions of dollars in service revenue**
A global electric vehicle manufacturer	Real-time telemetry, quality, and maintenance data **10x production efficiency vs. traditional auto plants**

AI/BI Use Cases

Predictive maintenance models, root cause analysis, OEE optimization, spare parts forecasting, quality prediction, energy optimization.

10 Code Commits + Bug Reports + Customer Escalations

High Value | Technology | SaaS | >$100M Impact

Git Commit History + Issue Tracker Data + Support Escalation Tickets

Value Unlocked

Engineering Prioritization: fix bugs that actually impact customers, not just the loudest ones

Technical Debt Quantification: measure the cost of legacy code in customer impact

Release Quality Prediction: forecast post-release issue volume

Real-World Examples

An engineering analytics platform	Git, issue, and review data **30% improvement in cycle time plus a 50% reduction in bugs shipped**
A software delivery intelligence platform	Deployment, incident, and change data **2x faster mean time to resolution through correlated analytics**
A global enterprise technology company	Work items, commits, and test results **50% improvement in lead time for internal teams**

AI/BI Use Cases

Bug triage automation, commit risk scoring, release readiness prediction, developer productivity metrics, technical debt tracking.

11 Call Transcripts + Deal Outcomes + Competitor Mentions

High Value | B2B Sales | SaaS | >$100M Impact

Sales/Success Call Recordings + CRM Win/Loss Data + Competitive Intel Mentions

Value Unlocked

Winning Messaging: know exactly what language wins deals

Objection Handling Playbooks: learn from successful responses

Competitive Battlecards: real-time intel on competitor positioning

Real-World Examples

A conversation intelligence platform	Calls, outcomes, and competitor mentions **27% increase in win rates plus a 50% reduction in ramp time**
A sales conversation intelligence platform	Call, outcome, and engagement data **20% increase in quota attainment**
An AI-guided sales engagement platform	Engagement, call, and outcome data **40% more meetings booked through AI guidance**

AI/BI Use Cases

Call scoring, coaching recommendations, battlecard generation, talk-pattern analysis, objection clustering, rep benchmarking.

12 Web Traffic + Content Engagement + Conversion Data

High Value | E-commerce | Media | D2C | >$100M Impact

Website Analytics + Content Interaction Metrics + Form Fills/Purchases

Value Unlocked

Content ROI Measurement: know which content drives revenue, not just traffic

Journey Optimization: identify friction points in conversion paths

Personalization: serve the right content to the right visitors at the right time

Real-World Examples

A global digital experience software company	Traffic, engagement, and conversion personalization **30% increase in conversion rates through personalization**
A digital product analytics platform deployed at a consumer wellness app	Product, engagement, and conversion analytics **40% increase in subscription conversions through journey optimization**
A digital experience analytics platform	UX analytics, engagement, and conversion data **35% increase in conversion through experience optimization**

AI/BI Use Cases

Attribution modeling, content scoring, A/B test analysis, personalization engines, journey analytics, SEO ROI measurement.

The Value Enablement Story

The same data combinations that create the most risk when exposed are the same combinations that create the most value when properly governed.

Without Governance: Risk

- Customer + Contract data exposed to competitors
- HR + Compensation data leaked = discrimination lawsuits
- Pricing + Inventory data enables insider trading
- Clinical + Patient data = HIPAA violations

With Cyera Governance: Value

- Customer + Contract data powers churn prediction
- HR + Compensation data powers retention modeling
- Pricing + Inventory data powers demand forecasting
- Clinical + Patient data powers care optimization

Cyera's role: find where high-value combinations exist, ensure proper access controls, enable safe AI/BI consumption, and monitor for misuse, so you unlock the value without the exposure.

Common Patterns Across Both Types of Combinations

Read side by side, the 24 combinations above point to four underlying dynamics that hold regardless of whether the outcome is toxic or synergistic.

Sensitivity is combinatorial, not additive

In nearly every combination in this report, each individual data element is described as low-to-moderate sensitivity on its own: operational, routine, or confidential but not actionable. The risk, or the value, only appears once two or three elements are placed in proximity. A price list plus a warehouse manifest is nothing alone; together, they are a trading signal. Product usage data plus support tickets is nothing alone; together, they are a churn model.

Traditional detection is built for single elements, not proximity

Every toxic combination in this report carries a detection-difficulty rating of Medium, High, or Very High, and the stated reason is consistent: no single element trips a PII, PCI, or keyword-based control. Signature-based tools were built to catch a social security number or a credit card number, not to notice that a spreadsheet of customer contract values happens to sit next to a file of renewal dates.

AI and BI tooling amplifies whichever direction the data is pointed

Every combination in this report, toxic or synergistic, carries an AI callout. The same correlation capability that lets a large language model turn manifest volumes and pricing into a trading signal in seconds is the capability that lets a churn model turn usage telemetry and support history into a retention play. The technology is neutral; governance determines the outcome.

The highest-risk pairings and the highest-value pairings are frequently the same pairings

Customer and contract data appears as a vector for competitive poaching and trade secret theft, and elsewhere as the foundation of predictive churn modeling and marketing ROI analysis. HR and compensation data appears as a discrimination and WARN Act exposure, and elsewhere as the foundation of retention modeling and succession planning. Pricing and inventory data appears as an insider-trading enabler, and elsewhere as the foundation of demand forecasting. Clinical and patient data appears as a HIPAA exposure, and elsewhere as the foundation of care pathway optimization. The data itself is not the variable. Governance is.

The practical implication follows directly: a governance program that only reduces risk, without also enabling the legitimate high-value uses of the same data, solves half the problem. A governance program that only accelerates AI and BI access, without controlling for toxic proximity, solves the other half badly. The next section translates that into what security and data teams should actually watch for, in both directions.

What Security Teams Should Look For

The patterns above point to a consistent set of capabilities that matter whether the goal is shutting down a toxic combination or clearing the way for a synergistic one. Drawing on the detection-difficulty and value-unlocked detail across all 24 combinations, the following reflects where most programs have gaps today:

- [ ] **Cross-system correlation, not single-system scanning.:** Nearly every toxic combination in this report involved data that lived in two or three different systems: CRM plus email, OT plus IT, board portal plus local sync. Detection that only scans one system at a time will miss the pairing by design.
- [ ] **Business-context classification, not just PII/PCI pattern matching.:** M&A models, trading algorithms, board minutes, and pricing data carry none of the regex-matchable patterns that traditional DLP looks for. They require understanding what a document is about, not just what strings it contains.
- [ ] **Proximity and co-occurrence monitoring.:** The toxic effect in this report is triggered by data elements sitting near each other, in the same folder, the same dataset, the same conversation thread, not by any single element's sensitivity. Monitoring needs to flag when otherwise-benign elements co-occur.
- [ ] **OT/IT boundary visibility.:** The energy-sector combination in this report was rated High detection difficulty specifically because operational technology data is commonly siloed from IT-side DLP tooling.

- [ ] **AI prompt and agent activity monitoring.:** Every AI amplification risk in this report describes a scenario where an LLM or AI agent does in seconds what previously required a specialized analyst team. Any AI copilot, chatbot, or custom agent with broad data access should be treated as a new correlation surface, not just a productivity tool.

- [ ] **Regular review of routine business documents.:** Several of the highest-impact combinations in this report, including M&A models, board minutes, and supplier contracts, were stored in ordinary business tools, precisely because they didn't look sensitive enough to warrant special handling.

- [ ] **The same visibility applied in both directions.:** The context and proximity monitoring that catches a toxic pricing-plus-manifest combination is the same capability that should be clearing pricing-plus-inventory data for an approved demand-forecasting model. Treat governance as a lever for enabling safe, fast access to synergistic combinations, not only for blocking risky ones.

Conclusion: Turning Combinatorial Risk into Combinatorial Advantage

The 24 combinations in this report make one point in two directions. Data sensitivity is not a property of a single field; it is a property of context, proximity, and combination. Every security program built exclusively around single-element pattern matching will keep missing the exact category of exposure that has already produced more than $15B in documented losses across the examples in this report. Every data program that treats governance purely as a cost center will keep leaving more than $50B in documented annual value on the table, unrealized, in those same datasets.

The organizations that get ahead of this are not choosing between locking data down and opening it up. They are building the visibility to know, continuously, which combinations exist, where they live, and what they mean, so the same program that closes off toxic exposure is the one that safely unlocks synergistic value. That is the gap traditional tools cannot see, and it is the one Cyera was built to close.

Schedule your private demo to see the Cyera platform in action, and learn more about how we can set you up to reduce toxic combinations and optimize synergistic combinations.

Book a demo →

About Cyera

Cyera is the Data and AI Security Platform built for the age of agents. Named a Leader in The Forrester Wave™: Sensitive Data Discovery & Classification Solutions, Q2 2026, Cyera helps enterprises like Paramount, Chipotle, and Valvoline control exactly what data their AI can reach, and govern what happens next. The platform secures data at rest, in motion, and in use.

Learn more at www.cyera.com, or follow Cyera on LinkedIn.

Appendix: Methodology & Definitions

This report consolidates two prior Cyera intelligence briefs, a Toxic Data Combinations report and a Synergistic Data Combinations report, into a single document, organized by combination type and by the patterns that connect them. No combinations, statistics, risk ratings, or dollar figures were added or altered in the process; the source material was reorganized, and duplicate framing between the two source reports was consolidated for readability.

Real-world examples throughout this report reflect verified, publicly documented outcomes, with organizations described by industry, scale, and role so the focus stays on the underlying pattern. Figures and outcomes are reported as documented in the original source material.

Term	Definition
Toxic combination	A pairing or grouping of data elements that, individually, carry low-to-moderate sensitivity, but that create critical or high business/security risk when combined.
Synergistic combination	A pairing or grouping of data elements that, individually, carry limited standalone value, but that unlock significant business value when combined and properly governed.
Detection difficulty	An assessment of how readily traditional, signature-based security tooling (DLP, PII/PCI scanning) would catch the combination in question, based on the presence or absence of pattern-matchable data.