WHITE PAPER

CYTEX™
AI-Powered Resilience Command Center

Ozempic Maker Novo Nordisk Breach

264 GB Data Breach and $25 Million Extortion Attempt

hello@cytex.io | https://cytex.io

Novo Nordisk Breach and $25 Million Extortion Attempt

Technical analysis of the FulcrumSec extortion against Novo Nordisk, with the corrected initial access vector confirmed through forensic reconstruction by an external threat intelligence team.

What Happened

A cyber extortion group operating under the name FulcrumSec has claimed responsibility for the theft of more than a terabyte of data from pharmaceutical company Novo Nordisk, including 264GB of confirmed proprietary content spanning source code, AI model assets, clinical trial information, and patient records. The group emerged in October 2025 and claims to have spent more than two months inside Novo Nordisk's networks before initiating extortion contact on June 1, 2026.

FulcrumSec demanded $25 million for non-disclosure. Novo Nordisk refused to pay. The group is now exploring private sales of subsets of the stolen data, with a stated preference for public release on the grounds that doing so creates a stronger deterrent against future non-payment.

The breach was initially reported across security media as the result of a leaked GitHub token exposed in public JavaScript files. That characterization was incorrect. Independent reconstruction by an external threat intelligence team (CybelAngel) has since corrected the public record. The actual initial access vector was an AWS account compromise, not a GitHub token leak, and the path to the data was substantially more direct than the early reporting suggested.

This brief incorporates the corrected technical findings.

What Was Actually Compromised

Per FulcrumSec's claims, validated in part through the group's response to Novo Nordisk's identity verification requests during extortion negotiations:

AI and machine learning assets:

- Trained model checkpoint (16 GB)
- Proprietary training dataset (407 MB)
- Full source code, including `modeling_novopert.py` and the associated training pipeline
- 113 complete training runs with logs
- 30 trained AI models and 70 distinct datasets

Infrastructure intelligence:

- Internal infrastructure maps covering HPC clusters, Slurm scheduler configurations, and SSH topology

- Container images totaling more than 53 GB
- Developer identities and internal hostnames
- Private GitHub repository URLs

Research and clinical data:

- 494 GB of cell painting microscopy images
- Molecular blueprints for tens of thousands of experimental compounds
- Comprehensive clinical history files for approximately 11,500 patients

The aggregate exposure spans Novo Nordisk's intellectual property layer, its research operations, its patient interaction records, and the infrastructure context an attacker would need to operate inside the environment for an extended period.

The Initial Access Correction

Early public reporting attributed the breach to a GitHub token exposed in JavaScript files crawled from Novo Nordisk's web properties. The narrative was attractive because it fit a familiar pattern: a developer accidentally commits a credential, the credential gets crawled, an attacker discovers it, the attacker uses it to access source code.

Forensic reconstruction by the threat intelligence team that initially flagged the exposure determined that the actual chain was different in material ways:

- The credential in question was not a GitHub token. It was an AWS access key embedded in client-side JavaScript served from a Novo Nordisk web property. The JavaScript file was publicly reachable.
- The credential provided direct AWS account access, not source code access. Once located, the key allowed the attacker to authenticate to AWS APIs as the identity associated with the key, with whatever privileges that identity held.
- The privileges associated with the credential were sufficient to reach high-value data. The compromised identity had access paths into S3 buckets, container registries, and other AWS services that hosted the assets ultimately exfiltrated. The attacker did not need to pivot through additional systems to reach the data.
- No GitHub repository compromise was required. The early narrative implied a chain that crossed from a leaked token through GitHub into the research environment. The corrected chain shows that AWS-resident data was reachable directly from the exposed credential.

The correction matters operationally because the defensive lessons differ. A GitHub token leak narrative points to secret scanning on commits, GitHub-side controls, and code review hygiene. The actual chain points to a broader category of failure: client-side JavaScript being treated as a low-risk surface, AWS credentials provisioned with excessive scope, and the absence of monitoring on AWS API usage from the compromised identity over the extended dwell time.

How the Attack Likely Progressed

Reconstructing the full chain from public reporting and the corrected forensic findings:

Step 1. Credential exposure in client-side JavaScript. An AWS access key was embedded in a JavaScript file served from a Novo Nordisk web property. The file was publicly accessible. Standard credential-discovery tooling routinely scrapes such files looking for credential patterns.

Step 2. Credential harvesting by the attacker. FulcrumSec located the credential through automated scanning, manual discovery, or commodity dark-web feeds aggregating exposed secrets. The exact discovery mechanism has not been publicly characterized.

Step 3. Authentication to AWS as the compromised identity. The attacker authenticated to AWS APIs using the exposed key. Standard programmatic access does not require user interaction, MFA prompts, or interactive sign-in events. The authentication appeared in CloudTrail logs as legitimate API activity from the identity the key represented.

Step 4. Reconnaissance of accessible resources. The attacker enumerated the resources the compromised identity could access, including S3 buckets, ECR repositories, EC2 instances, and any other services within the identity's permission scope. This is a routine post-authentication step using `aws sts get-caller-identity`, `aws iam list-attached-user-policies`, and equivalent enumeration commands.

Step 5. Extended dwell time. FulcrumSec claims more than two months of operational presence in Novo Nordisk's environment. During this window, the attacker enumerated data assets, identified high-value targets including the AI training pipeline and clinical data stores, and staged exfiltration.

Step 6. Bulk exfiltration. Source code, model checkpoints, training datasets, microscopy images, container images, and patient records were extracted from AWS-resident locations to attacker-controlled infrastructure.

Step 7. Extortion contact. On June 1, FulcrumSec contacted Novo Nordisk executives directly. The company responded 48 hours later through a randomized Proton Mail address. FulcrumSec verified its position by producing specific files that only an actor with genuine access could have identified.

Step 8. Refusal and pivot to sale. Novo Nordisk declined the $25 million demand. FulcrumSec moved to private sales while signaling that public release remains the preferred outcome strategically.

Regulatory Liabilities

The data categories implicated in this breach trigger multiple regulatory regimes simultaneously, each with independent reporting and enforcement obligations:

GDPR and NIS2. The exfiltrated healthcare professional registry contains fully identifiable information about EU-based medical professionals. GDPR maximum penalties reach 4% of global annual turnover or €20 million, whichever is higher. NIS2 imposes incident reporting obligations on

essential and important entities, with pharmaceutical manufacturing falling within the scope of the directive.

HIPAA. If any of the clinical trial data streams contained un-scrubbed protected health information for U.S. patients, the breach falls under the enforcement authority of the U.S. Department of Health and Human Services Office for Civil Rights. HIPAA breach notification, penalty assessment, and corrective action plan processes apply.

Trade secret and intellectual property law. The molecular blueprints for tens of thousands of experimental compounds, the proprietary AI training pipeline, and the model weights represent active trade secrets. Loss of trade secret status through public release has material commercial consequences independent of regulatory fines.

Investor and disclosure obligations. Material cybersecurity incidents trigger disclosure requirements under SEC rules for U.S.-listed entities and equivalent obligations in other jurisdictions where Novo Nordisk is publicly traded.

Indicators and Detection Signals

The breach pattern produces detectable signals at multiple layers. Organizations applying the same defensive posture should instrument:

Public surface signals:

- Credential patterns in publicly served JavaScript, CSS, configuration files, and other client-side assets
- Static secret scanning across the public web footprint, not only across code repositories
- Continuous monitoring of paste sites, dark-web markets, and credential aggregation services for exposed organizational credentials

AWS API activity signals:

- CloudTrail events showing authentication from previously unseen source IP addresses for programmatic identities
- API call sequences consistent with reconnaissance (`get-caller-identity`, broad `list-*` operations, `describe-*` enumeration across services)
- High-volume S3 GetObject or container image pull activity from a single identity over a sustained window
- Programmatic identities used outside their expected operational pattern (time of day, source region, service combinations)

Long-dwell signals:

- Programmatic credentials that authenticate continuously from a single source IP for weeks without corresponding application deployment events
- Identities accessing resources outside the scope of their documented purpose
- New IAM policy attachments or session policy modifications associated with programmatic identities

Extortion-stage signals:

- Inbound communications referencing specific internal files, identities, or systems
- Verification requests from external parties that demonstrate knowledge of internal-only content

Defensive Actions

Eliminate long-lived programmatic credentials. The AWS key that initiated this breach was a static credential. Replace static keys with short-lived, identity-bound, and scoped tokens issued through federated identity, IAM Roles Anywhere, or equivalent mechanisms. Where static keys cannot be eliminated, rotate them aggressively and scope them narrowly.

Scan the entire public web footprint for credentials, not only code repositories. Client-side JavaScript, configuration files served from CDNs, mobile application bundles, and any other public asset where developers might inadvertently embed credentials should be in scope for continuous secret scanning.

Apply least-privilege rigor to programmatic identities. The compromised identity in this breach had access to substantially more than its documented operational scope required. Audit programmatic identity permissions across the environment and reduce to the minimum required for the documented use case.

Reclassify AI assets as crown-jewel data. Model weights, training configurations, training data, and pipeline code represent intellectual property at the same security tier as active product patents. Apply data classification policies, access controls, and monitoring at that tier. Treat model registries with the same controls applied to production databases.

Embed fingerprinting and watermarking in proprietary model weights and training datasets. Where models are exfiltrated, watermarking enables attribution if the stolen assets surface in downstream products, academic publications, or competitive offerings. Standard ML watermarking techniques apply.

Engineer SIEM and UEBA alerts for anomalous programmatic credential usage. Programmatic identities should have documented operational baselines, and deviations from those baselines (new source geographies, unusual API combinations, sustained high-volume activity) should generate high-priority alerts. The 60-day dwell time in this breach indicates that anomaly detection on programmatic identities was either absent or insufficiently tuned.

Apply conditional access policies to code repositories, container registries, and AI asset stores. Restrict access by source network, device posture, and authenticated identity. Default-deny external access to these stores and require step-up authentication for administrative operations.

Establish cross-functional governance over AI assets. Security, data science, and legal teams should jointly define classification, access control, retention, and incident response procedures for AI assets. The historical pattern of treating ML pipelines as research infrastructure outside the standard security governance perimeter produces the exact gap this breach exploited.

Plan for the data-release downstream impact. The data is now in the hands of an extortion group that has stated a preference for public release if private sales do not materialize. Organizations that may have been counterparties in clinical trials, supply chain partners, or collaborators with Novo Nordisk should treat their own data exposure as a function of what was in the stolen archives, not only what Novo Nordisk discloses.

What This Disclosure Pattern Reveals

Two structural observations apply beyond this specific breach.

The early public narrative was wrong. A GitHub token leak is a familiar and easily communicated story. The actual chain was an AWS account compromise through credential exposure in client-side JavaScript. Defensive programs that calibrate against the early narrative will harden the wrong surface. The corrected version reframes the lesson around AWS identity hygiene, scope discipline, and detection coverage on programmatic identities.

AI assets are now a primary extortion target. The most distinctive feature of FulcrumSec's claimed haul is not the patient records or the source code, both of which appear in many breaches. It is the model checkpoints, training datasets, and pipeline code. The 16 GB model checkpoint and the 70 distinct datasets represent commercial value an attacker can sell to competitors, foreign state-aligned research programs, or other actors interested in AI capability acquisition without the underlying research investment. Organizations operating proprietary AI training pipelines should treat them as crown-jewel assets at the same governance tier as core IP, with the access controls, monitoring, and incident response procedures that tier implies.