# Submission to the Centers for Medicare & Medicaid Services (CMS)

**Refer to File Code:** CMS-6098-NC
**Subject:** Request for Information (RFI); Potential Regulatory Changes to the "CRUSH" rules
**Date:** March 30, 2026

## I. Introduction and Statement of Interest

Claims Sciences (claimssciences.com) is a technical organization dedicated to the advancement of **Defensible Program Integrity** through new and novel AI/ML architectures. We submit these comments to offer an operational and scientific perspective on the "CRUSH" initiative by bridging the gap between mathematical anomaly detection and human-readable forensic evidence. While our current research and identification of providers committing Fraud, Waste, and Abuse (FWA) is conducted **retrospectively**, utilizing the high-fidelity historical data provided by CMS, our ultimate objective is to provide the technical foundation required to shift the industry paradigm from a reactive "**pay and chase**" model toward a proactive real-time "**detect and deploy"** strategy. By mastering the semantic patterns found in retrospective datasets, Claims Sciences aims to equip CMS with the **pre-payment controls** necessary to prevent improper payments before they occur.

Claims Sciences' current focus is the active ingestion and analysis of publicly available **CMS provider-level data** (e.g., Medicare Physician & Other Practitioners) to identify high-risk billing behaviors. Our technical leadership includes practitioners with over a decade of experience in applied AI/ML research and engineering within the aerospace and defense sectors. This knowledge base is complemented by ongoing doctoral research at the University of Denver focused on anomaly detection in program integrity for public health. This unique background allows Claims Sciences to offer "on-the-ground" technical insights into the limitations of existing rules-based systems and the opportunity to incorporate **Agentic Reasoning Pipelines.**

Our recommendations are grounded in extensive academic and experimental research including:

- **Publication in 2022 SMU Data Science Review:** Demonstrating the efficacy of initial machine learning application in detecting Medicare Fraud, Waste, and Abuse (Goodwin, B., et al., 2022).
- **University of Denver Doctoral Research (2025-2026):** Comprehensively studying 1.53 million providers to validate pitfalls in classical numerical machine learning techniques and development of foundation-model-centered anomaly detection methodologies coupling novel anomaly detection, retrieval augmented

generation, and semantic reasoning. These techniques demonstrate great promise toward identifying semantic patterns that evade numerical detection and have achieved a 71% recall rate of FWA detection.

# II. The Claims Sciences Methodological Framework

To effectively "deter Fraud, Waste, and Abuse and promote payment accuracy," CMS must augment existing rules-based models with complementary data-driven approaches. Claims Sciences advocates for a "**Triple-Pillar Architecture**" designed to navigate the complexities of provider-level claims data. This architecture serves as the "cohesive thread" for our specific recommendations in this RFI:

1. **High-Precision Mathematical Filtering**:
    - We propose purpose-built anomaly detection utilizing engineered features that normalize data across disparate provider types. This includes:
    - **The "Greed Ratio**": Isolating the markup of submitted charges vs Medicare-allowed amounts.
    - **Geographically Normalized Z-scores**: Calculating a provider's statistical distance from their specific specialty peer group to differentiate "expensive" care from "illegal" activity while controlling for geographic differences.

2. **Relational Intelligence (Knowledge Graphs)**: We recommend tracing FWA through graphically linked indicators to expose hidden networks and other schemes including:
    - **Geospatial risk**: Mapping location-specific FWA levels (Zip Code Risk) and physical address linkage.
    - **Behavioral Clusters**: Identifying HCPCS billing patterns and provider networks that indicate collusion.
    - **Temporal visualization**: Tracing provider behavior patterns over time to understand behaviors that led to exclusion.

3. **The Evidence Engine (Locally Hosted Foundation Models):** We advocate for utilizing non-cloud-based Foundation Models (FMs) to transform statistical flags into **actionable investigative narratives.** This approach provides the "**Semantic Reasoning**" required to justify enforcement actions. By leveraging program rulebooks and historical records, this engine generates explainable reports that can significantly reduce investigative workload and allow agents to more expeditiously crush FWA.

# III. Response to Specific RFI Questions

## A. Modifications to Program Integrity Requirements

Question: *What changes could CMS or its contractors make to existing processes to promote their ability to effectively deter fraud, waste, and abuse and promote payment accuracy and efficiency, including by more expeditiously gathering actionable information?*

### Recommendation 1: Transitioning from disparate tabular data to relational knowledge graphs

**The problem: Temporal Fragmentation and Data Silos**
Current CMS data dissemination (via data.cms.gov) relies on disparate temporal snapshots. While these datasets are robust, they function as disconnected "flat files" that allow bad actors to operate within the gaps of regularly spaced updates.

For an analyst to investigate a single provider in a specific geography (e.g., Gary, Indiana), they must manually aggregate the **Medicare Physician & Other Practitioners by Provider and Service (MPOPPS)** file, the **Geography and Service**, and various cost-calculation datasets. Navigating all ten years of currently available data means managing some 100 million rows of tabular data indexed by HCPCS codes but lacking internal linkages is computationally expensive and creates a significant barrier to "expeditiously gather actionable information."

**The Proposed Solution: Knowledge Graphs (KGs) as the Relational Backbone**
Claims Sciences recommends a transition toward representing CMS data as **Knowledge Graphs**. KGs move beyond tabular structures by defining the relationships (edges) between entities (nodes), enabling meaningful spatial and temporal linkages across all 173+ unique CMS datasets.

- **Computational Efficiency:** Unlike traditional SQL (Structured Query Language) joins on massive datasets, KGs utilize efficient traversal algorithms. Research has shown that graph databases like **Neo4j** can outperform relational databases by orders of magnitude for complex relationship queries (Kotiranta, P., Junkkari, M., & Nummenmaa, J. (2022)).
- **The "Detective vs. Prosecutor" Model:** The Claims Sciences framework utilizes purpose-built XAI detection models as the "detective" to identify statistical anomalies, while the KG serves as the "prosecutor," providing the relational evidence (i.e., shared addresses) required for a forensic narrative.
- **Visualization of Fraud Networks:** As a simple example, which is demonstrated in our research **Goodwin, B. (2026)**. *provider_forensic_graph.html: A*

*Force-Directed Knowledge Graph for Latent Fraud Detection* (Version 2.0)., a KG can instantly flag “Phoenixing” schemes by identifying “clean” NPIs that share physical practice locations or billing patterns with entities currently on the **List of Excluded Individuals and Entities (LEIE)**.

These capabilities are particularly relevant for identifying coordinated fraud schemes and provider network behaviors that are difficult to detect using isolated, claim-level analysis.

## Recommendation 2: Enhancing LEIE Accessibility and Real-Time Synchronization for Unified Provider Identity

**The problem: The “Window of Opportunity for Fraud”**

A critical operational bottleneck exists in determining a provider’s exclusionary status. Currently, this requires cross-referencing CMS data with the **Office of the Inspector General’s LEIE database**, where identifiable keys (like NPIs) are frequently unavailable.

As of February 2026, the LEIE contains **82,749 entries**, yet only **8,482 (10.25%)** contain an NPI. This lack of a unified identity layer limits the ability of machine learning models to study historical billing data to unravel long-term behavioral patterns of FWA. Furthermore, the current monthly update cadence creates a “**Window of Opportunity**”; if a provider is excluded on the 5th but the file is not updated until the 25th, bad actors have a 20-day window to maximize fraudulent billing before detection. Additionally, delays in claims data availability further compound this gap.

**The Proposed Solution: A Unified Provider Identity and Behavioral Layer**

CMS should collaborate with OIG to establish a **Unified Provider Identity Layer** with real-time LEIE synchronization.

- **Real-Time Deterrence:** Transitioning from monthly updates to a real-time API-driven LEIE would allow future Claims Sciences’ models to conduct real-time analysis on incoming claims and immediately flag suspect claims at the pre-payment stage.
- **Behavioral Data for Model Fine-tuning:** The LEIE should be expanded to include the specific “reasoning” for exclusion (i.e., HCPCS codes associated with the abuse). Our research indicates that even a small amount of specific behavioral data can significantly increase the ability of language models to generalize and identify previously unseen exclusionary patterns (Chung, H. W. et al., 2024**).**
- **Defensible Enforcement:** Including known dates of FWA and specific exclusionary behaviors within the identity layer allows for the creation of “Semantic Reasoning” pipelines that have potential to reduce the administrative burden of investigations on state and federal investigators.

Question: *What types of analytics, methodologies, or data-driven approaches would be most effective in identifying indicators of potential fraud, waste, or abuse?*

### Recommendation 1: Shifting the Paradigm from “Pay and Chase” to “Detect and Deploy”

A primary friction point for State Medicaid agencies and Program Integrity teams is the reactive nature of the “pay and chase” model. Claims Sciences recommends a shift toward a “**detect and deploy”** strategy. By leveraging high-fidelity historical data, modern machine learning can pivot from merely calculating historical losses to identifying active billing patterns in real-time. This allows CMS to trigger payment suspensions or pre-payment edits far more expeditiously, stopping the flow of funds before a provider can disappear. While real-time prevention represents the definitive goal of payment integrity, Claims Sciences is executing a phased deployment that begins with rigorous retrospective validation. This methodology ensures that our architectures are benchmarked against historical ground truth, allowing CMS to maintain its commitment to prompt payment while reserving intervention for only the highest-confidence FWA signals.

### Recommendation 2: Transitioning from Numerical to Semantic-Enabled Foundation Models

**The Failure of Traditional Numerical Models:**
Our research at the University of Denver confirms that traditional “numerical” ML models (such as Logistic Regression and Random Forest) miss the vast majority of FWA because they lack semantic context. In a Fall 2025 evaluation of 1.5 million providers, we benchmarked two “traditional” models against the OIG List of Excluded Individuals/Entities (LEIE) with the following results (i.e., given MPOPPS data, can a model flag providers who later became excluded?):
**These findings are detailed in: Goodwin, B. (2025). *The use of foundation models to detect Medicare Fraud, Waste, and Abuse: Experimental Results.***

- **Experiment I (Linear Baseline)**: A Logistic model achieved a ROC - AUC score of 0.4871, performing slightly worse than random chance. Without peer-group benchmarks, it incorrectly flagged high-cost specialists due to volume.
- **Experiment II (Non-Linear Architecture)**: A Random Forest model achieved 95% overall accuracy, but it had a recall of only **0.08**. In other words, it **missed 92% of actual fraud cases**. Because the minority class (fraud) represents <0.1% of the dataset, numerical models often “play it safe” by predicting nearly every provider as “clean” to maximize accuracy.

**The “Semantic Bridge” Breakthrough:**
To solve the “contextual blindness” identified in Experiments I and II, Claims Sciences developed the **Semantic Bridge**. This proprietary data engineering framework translates raw numerical claims into a context-rich “Provider Knowledge base,” giving the AI the clinical intuition required to evaluate data. This framework is detailed extensively in:
**Goodwin, B. (2026). Technical Appendix: Provider Profile & Peer-Group Database [Technical Report].**

The Semantic Bridge consists of three critical layers:

1. **High-Density Dimensionality Reduction:** Aggregating millions of granular HCPCS codes into multi-dimensional **NPI Profiles**. This allows the model to evaluate the "total output" and longitudinal behavior of a provider.
2. **Contextual Peer-Group Baselines**: Segmenting providers by specialty and calculating specialty-specific Z-scores. This provides a "yardstick" for statistical norms, preventing the model from flagging providers for simply being "expensive."
3. **Forensic Feature Enrichment**: Engineering signals such as the "**Greed Ratio**" (the markup of submitted charges versus Medicare-allowed amounts) and geographic risk scores based on historical exclusion rates.

**Experimental Results: Foundation Model Assisted Anomaly Detection (FMAAD)**
By contrast, our third experiment utilized a Foundation Model paired with the Semantic Bridge. When tested on a class-balanced validation set (to enable meaningful recall measurement under extreme class imbalance) of providers, the **FMAAD framework achieved a 71% recall** (compared to the 8% recall seen in the Random Forest). While this model identified more "suspect" cases (precision of 0.46), it provides a far superior "net" for proactive detection. In practical terms, **for every two providers the model flags, one is a legitimate FWA lead**.

**Outcome: Eliminating Alert Fatigue via "Prosecutorial Narratives"**

Instead of investigators manually validating hundreds of leads against complex manuals, this framework performs **Retrieval-Augmented Generation (RAG)** to create a "**Prosecutorial Narrative**.**"** This plain-English case brief pre-validates findings against state-specific policy documents, providing the "Semantic Reasoning" required to justify enforcement actions while significantly reducing investigative time and administrative burden associated with provider investigation and appeals. By combining anomaly detection with XAI, systems can generate:

- Structured case summaries
- Policy-aligned reasoning
- Supporting evidence references

This reduces manual validation effort and improves consistency across investigators.

We note that these capabilities can be implemented incrementally, beginning with retrospective analysis and pilot programs before progressing to pre-payment interventions once sufficient precision thresholds are validated.

# IV. Risks and Safeguards

Claims Sciences recognizes that advanced analytics and AI-driven program integrity approaches introduce important considerations, including the potential for false

positives, provider administrative burden, and algorithmic bias. To ensure responsible deployment, we recommend the following safeguards:

- **Human-in-the-loop review**: AI systems should augment, not replace human investigators, particularly for adverse actions such as payment suspension or revocation.
- **Transparency and explainability requirements**: Model outputs should be auditable and interpretable to support defensible decision-making.
- **Routine auditing of model performance and bias:** Models should be continuously monitored for drift, bias, and real-world effectiveness.

These safeguards are essential to maintaining provider trust while enabling CMS to responsibly expand the use of advanced analytics in program integrity operations. These recommendations align directly with CMS's objectives under the CRUSH initiative by enabling faster, more accurate, and more defensible identification of FWA while maintaining fairness, transparency, and trust in program integrity operations.

# V. Authors & Technical Leadership

This submission represents the collective expertise of the **Claims Sciences** engineering team. Our organization bridges the gap between high-consequence AI/ML engineering and academic research in healthcare program integrity.

- **Benjamin Goodwin - Lead Researcher/Founder:** A doctoral student in Computer Science at the University of Denver. His research, "*Novel Anomaly Detection Methods for Identifying Fraud, Waste, and Abuse in U.S. Healthcare Systems,*" focuses on the intersection of anomaly detection and healthcare administration. Ben is a professional AI/ML engineer in the aerospace and defense sector.
- **Ashton Teimoori - Lead ML Engineer:** Holds an M.S. in Computer Science from Georgia Tech with research specializing in machine learning. Ash is a professional AI/ML engineer with an extensive background in GN&C and engineering within the aerospace and defense industry, bringing technical rigor from a fault-tolerant systems perspective.
- **Samuel Richardson - Strategy & Operations**: Specializes in technical sales, project management, and product scaling. Sam brings an extensive background in navigating the growth of technical startups and shaping operational opportunities within complex, high-stakes regulatory environments.

# VI. Core Research and Technical Documentation

The methodologies and experimental results presented in this RFI response are supported by the following literature:

1. **Goodwin, B., Canton, A., & Olanipekun, B. (2022).** *Detection of Fraud, Waste, and Abuse in Medicare Public Datasets.* SMU Data Science Review. **(Submitted as an attachment with this RFI response)**

2. **Goodwin, B. (2025).** *The Use of Foundation Models to Detect Medicare Fraud, Waste, and Abuse: Experimental Results.* Unpublished technical report, University of Denver. **(Submitted as an attachment with this RFI response).**

3. **Goodwin, B. (2026).** *Technical Appendix: Provider Profile & Peer-Group Database.* [Technical Report]. University of Denver. **(Submitted as an attachment with this RFI response)**

4. **Goodwin, B. (2026)**. *provider_forensic_graph.html: A Force-Directed Knowledge Graph for Latent Fraud Detection* (Version 2.0). University of Denver. **(Submitted as an attachment with this RFI response)**

5. **Chung, H. W., Hou, L., Longpre, S., Zoph, B., Tay, Y., Fedus, W., Li, Y., Wang, X., Dehghani, M., Brahma, S., Webson, A., Gu, S. S., Dai, Z., Suzgun, M., Chen, X., Chowdhery, A., Castro-Ros, A., Pellat, M., Robinson, K., Valter, D., Narang, S., Mishra, G., Yu, A., Zhao, V. Y., Huang, Y., Dai, A. M., Yu, H., Petrov, S., Chi, E. H., Dean, J., Devlin, J., Roberts, A., Zhou, D., Le, Q. V., & Wei, J.** (2024). Scaling Instruction-Finetuned Language Models. *Journal of Machine Learning Research, 25*(70), 1-53.

6. **Kotiranta, P., Junkkari, M., & Nummenmaa, J.** (2022). Performance of Graph and Relational Databases in Complex Queries. *Applied Sciences*, *12*(13), 6490. https://doi.org/10.3390/app12136490