Centers for Medicare & Medicaid Services (CMS) Crushing Fraud Chili Cook-Off Competition: Summary Report

January 2026

U.S. Department of Health & Human Services
Centers for Medicare & Medicaid Services
Center for Program Integrity
7500 Security Boulevard Baltimore, MD 21244

Table of Contents

List of Figures

List of Tables

1 Introduction and Challenge Overview

CMS designed the Crushing Fraud Chili Cook-Off Competition to advance its understanding of industry readiness to apply explainable artificial intelligence (AI), specifically machine learning (ML) models, to detect anomalies and trends in Medicare Fee-for-Service (FFS) claims data that may indicate fraud, waste, and abuse (FWA). Through this competition, CMS also sought to discover scalable technologies that reduce manual burden while maintaining human oversight and interpretability, further enhancing CMS' growing crushing fraud efforts.

CMS launched this competition in response to a significant increase in the private sector's ability to leverage cutting-edge technologies, particularly AI/ML, to detect FWA within federal health programs. As these advanced analytical capabilities have become more prevalent in the marketplace, CMS recognized an opportunity to systematically evaluate the industry's capacity to apply such innovations to the unique challenges of Medicare program integrity. Through this challenge, CMS aimed to explore the full range of possibilities these technologies offer and to give participants an opportunity to demonstrate the effectiveness of their proposed approaches using real Medicare claims data.

CMS executed this challenge in two distinct phases. In Phase 1, CMS invited research proposals from all interested parties outlining their proposed, explainable AI/ML techniques and approaches. A neutral third-party evaluation panel composed of subject matter experts scored all eligible submissions using metrics established by CMS. Each proposal was assessed across five equally weighted evaluation criteria (20% each): Relevance and Impact, Innovation and Technical Merit, Explainability and Interpretability, Feasibility and Implementation Potential, and Clarity and Quality of Submission. Based on the quality and relevance of their proposals, CMS selected the top 10 teams to advance to Phase 2.

In Phase 2, CMS provided finalists with access to Limited Data Sets (LDS) consisting of 2022–2024 Standard Analytical Files containing Medicare FFS Hospice, Part B, and Durable Medical Equipment (DME) claims. These data sets contain claims associated with a random 5% sample of Medicare beneficiaries. All finalists were required to execute and comply with CMS Data Use Agreement (DUA) requirements, ensuring appropriate safeguards for the protection, storage, and use of this sensitive information throughout the competition. Finalists were instructed to apply their proposed techniques to the provided data and submit a summary of their findings for final evaluation.

The neutral third-party evaluation panel reviewed the finalists' submissions and made recommendations to CMS. Following the conclusion of the Phase 2 evaluation, CMS selected **Milliman, Inc.** as the winner of the Crushing Fraud Chili Cook-Off Competition and announced the results on CMS' social media platforms on December 15, 2025.

2 Phase 1: Participant Solicitation and Results

2.1 Challenge Participation Overview

CMS announced the Crushing Fraud Chili Cook-Off Competition through Challenge.gov, as well as CMS' website (cms.gov). The competition launched on August 19, 2025, opening Phase 1 submissions to all interested parties. Participants had until September 19, 2025, at 11:59 PM ET to submit their research proposals outlining their proposed explainable AI/ML techniques and approaches.

2.2 Phase 1 Submission Details

Phase 1 of the CMS Crushing Fraud Competition Chili Cook-off drew strong interest, with 259 total submissions, reflecting substantial interest from across the data and fraud analytics ecosystem. Submissions were received from a variety of entities, ranging from technology firms to academic institutions. Moreover, there was strong national engagement as submissions were received from 34 states, plus the District of Columbia.

Participants employed a wide range of explainability approaches in their proposed fraud detection solutions. Rules-based methods were the most common, used in 53% of submissions. SHapley Additive exPlanations (SHAP)- or Local Interpretable Model-agnostic Explanations (LIME)-based techniques appeared in 50% of proposals, and 49% of teams incorporated graph- or network-based approaches. Many teams combined multiple techniques, underscoring the emphasis placed on transparent and interpretable models.

Figure 1: Phase 1 Submissions by State

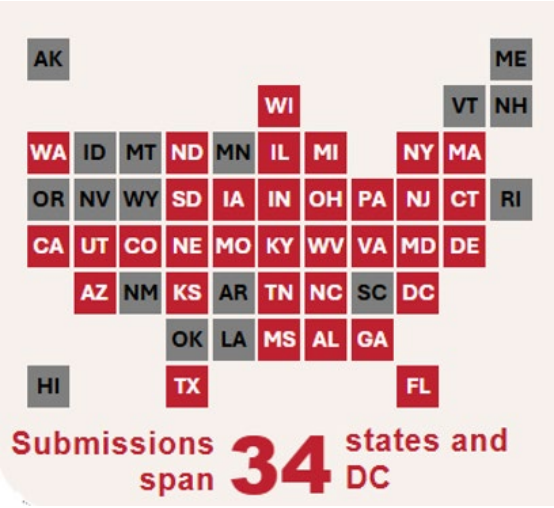

Figure 2: Types of Explainability Approaches Proposed in Phase 1

3 Phase 2: Approaches and Finalist Outcomes

3.1 Phase 2 Submission Details

Phase 2 of the competition provided the 10 finalist teams with the opportunity to apply their advanced analytic techniques to the 5% LDS datasets made available by CMS. Each team used their own unique approach to uncover different suspicious billing scenarios, with several groups uncovering multiple unique patterns. Traditional machine learning approaches were the most common, with many teams focusing on anomaly detection algorithms and clustering. Graph-based methods were nearly as prevalent among the solutions, reflecting a strong focus on detecting relationships among entities. Timeseries and other data analytic techniques accounted for the remaining portion of solutions.

To make results understandable for end users, teams incorporated several explainability approaches. Many provided interactive dashboards, while others used graph visualizations to show networks of suspicious actors. Teams leveraged Large Language Models (LLMs) to help interpret model outputs, and SHAP values were used to quantify which features most heavily influenced predictions.

Overall, seven types of anomalous and suspicious behaviors were identified across the submissions. The most frequently detected pattern was billing for unnecessary services. Collusion, self-referrals, upcoding or modifier abuse, and duplicate or phantom billing were also commonly identified. Less frequent, but still significant, were patterns suggestive of beneficiary identification fraud, and shell provider fraud.

3.2 Top 10 Finalist Solutions

Table 1 provides an overview of each finalist team along with a brief synopsis of their technical solution, as described by their respective team leader.

Figure 3. Top 10 Finalists Teams

Table 1. Top 10 Finalists and Solution Synopses (alphabetical order)

Finalist Team	Solution Synopsis
ABT Global	“The core part of our Phase 2 submission was around Quantile Random Forest and Generalized Low Rank Models. We built in K-means clustering and a Gaussian Mixture Model to verify outliers. We wanted to not just detect the fraud, but detect these patterns that were becoming suspicious for fraud - to flag fraud before claims were being paid, ultimately.”
Amida	“Our solution takes in claims data that is fundamentally semantic and subjective and transforms it into a mathematical and geometrical representation through vectorization. Once in the vector space, we're able to apply systematic and explainable approaches to cluster legitimate claims from claims that are anomalous in some way.”
Basys.ai	"We offer three unique value propositions. One is how we've trained our platform… we bring unique and complete data to the table. Second, we know from our work with health insurance companies, both commercial and Medicare Advantage, that certain providers are using ChatGPT or even Claude to cause fraud, waste, and abuse, and we knew how those fingerprints could be figured out. The third part was, based on the usual suspects like phantom billing, excess utilization, or billing for those that cease to exist, we were able to figure out those instances within the data as well. We tried to focus our solution on scalability as well to make sure we were able to deploy in the real world.”

Finalist Team	Solution Synopsis
Cormac	"We designed our solution to scale up to CMS's environment. We looked into technologies we can lift and shift into a secure environment at CMS, and we came up with a multi-model approach to tackle all aspects of fraud. We used a peer model that will look at all of the provider's peers and see who the outliers are. We also used a statistical outlier model to look into claims. We consolidated those outputs and used a couple of models to explain these results, such as a basic explanation in human language, along with a score associated with that. We also describe what fraud mechanisms are used and compile that information into a dashboard that can be used by investigators."
Hilltop Institute	“We thought about this idea of finding fraud like detecting fish in the water, with 'bad fish' or bad actors mixed in with other fish. We titled our solution 'Fraud Sonar.' The idea is that, using machine learning techniques against the Medicare Fee-For-Service claims, we can send out sonar pings in the data to be able to detect anomalous behavior. We focused specifically on looking at provider billing patterns over time to detect outliers.”
KPMG	“With our approach, we took the data, organized it, and applied it to a policy graph that we built to make sure that when we're looking for this anomalous behavior, it's checking against the facts. The second piece of this is a timeseries. We want to know when these events or anomalous behaviors are taking place. The combination of the policy graph and the timeseries analysis is applied against a Large Language Model, which produces outputs that are user-friendly and actionable for investigators to take and look into further.”
Milliman, Inc.	“Our proposal bridges the gap between unsupervised machine learning and something that is actually usable by program integrity investigators. We developed a provider risk score based on this algorithm we developed and packaged that up into a user-friendly dashboard where investigators can dive into the providers that exhibit the greatest risk and see what metrics are really driving that provider's risk.”

Finalist Team	Solution Synopsis
MindPetal	“We started by first looking to say, are these particular providers or suppliers anomalous or not. The next thing we looked at is potential relationships between these entities. So, looking at providers in the context of the hospices they work with or the DME suppliers they may work with, and do we see certain relationships that are also interesting. Do they share beneficiaries? Are there referral patterns that may be different than their peers? We looked at those to say, are we seeing certain clusters of groups that are sort of standing out in those communities when compared to the whole picture. The last thing we did, which we think is the most interesting, leaned into some of these cognitive models that are now available. We took some of our network analysis data, deidentified it, and fed it into an agentic model to say, 'We are a CMS claims investigator and are trying to look for anomalous information. Can you tell me what you would do, would you actually do it, and show us your work?' The findings from that work corroborated our earlier analysis as well as identified additional findings worth further investigation.”
Stanford, UCFC	“As part of the Chili Cook-off, we had access to the 5% Limited Data Set, specifically around the Carrier files, physician Part B claims, as well as Durable Medical Equipment and Hospice. We wanted to take it beyond outliers, thinking about how we could focus on a discrete area where we have clinical expertise and marrying that with the big picture - looking at the patterns, the knowledge graph of diagnosis and procedures, that matrix that is being provided - but also looking at it with a clinical lens and focusing on understanding what expected practice patterns and behaviors should be and being able to see what truly deviates from norms and is something we can validate clinically. We focused on ophthalmology primarily, using the Carrier file, feeling that that was going to be a particular challenge, but something that we could really demonstrate that we have that clinical expertise and it would be scalable. It's a specialty that has very high volume, as well as some very high-cost procedures and care that can be provided within it, but something that is truly defined by just the sheer volume and scale and, as such, can be particularly challenging to identify fraudulent practices.”
Visual Connections	“We started with the fraud analyst themselves. Instead of thinking about complex data models, we thought about what the fraud analyst is looking for and what their day-to-day lives are. It shouldn't be about what dataset to use and what type of query to run. We focused on the UI, first looking to build a user experience for a fraud analyst. Then, we started putting the math behind what we call a fraud library, where we built different types of fraud cases specific to the DME, Hospice, and Part B Carrier data, looking at over 100 different use cases. We essentially built a platform where all an analyst has to do is point to the data, and now the analyst has a user experience analyzing the data without having to worry about the complexity behind the models.”

3.3 Challenge Winner: Milliman

As the winner of the Crushing Fraud Chili Cook-Off Competition, Milliman, Inc. demonstrated a comprehensive and innovative approach to applying explainable AI/ML techniques for Medicare fraud detection. Their submission stood out among a competitive pool of finalists, effectively addressing the challenge objectives while maintaining a strong emphasis on interpretability and scalability.

The CMS Chili Cook-Off team highly valued Milliman's rigorous and transparent actuarial-based approach to FWA detection. Milliman's feature engineering method was informed by strong domain expertise, while at the same time, remaining flexible enough to be applied to novel areas to uncover subtle patterns of behavior indicative of potential FWA. The evaluation panel was impressed by Milliman's composite risk scoring approach that was weighted and scaled by financial value to produce risk tiers for triage and prioritization purposes. Overall, CMS found Milliman's final submission presented the most sophisticated and interpretable framework for detecting novel patterns in a highly complex data environment.

The following executive summary, provided by Milliman, details their methodology and key findings.

"To support the Centers for Medicare and Medicaid Services (CMS) in its mission to combat systemic fraud, waste, and abuse (FWA), we present a solution designed to enhance current program integrity efforts. By redefining FWA detection through the lens of actuarial risk management, our framework offers the statistical rigor and transparent evidence needed to augment CMS's defensive capabilities.

In our analysis of the 2022-2024 Medicare 5% Limited Data Set (LDS), our framework identified $870 million in payments directed to the top 0.5% of anomalous providers based on the model-calculated risk scores. Scaled to the 100% Medicare FFS population, this suggests $17.4 billion in potential financial exposure.[1] Within this $17.4 billion potential financial exposure, our analysis highlights Tier 1 (Critical Risk) providers representing only 0.16% of providers yet accounting for 21.5% of all paid claims in the sample (scaled, this represents $14.9 billion in potential financial exposure). The model's precision is reinforced by its independent identification of providers under Department of Justice (DOJ) investigation, demonstrating its ability to align with real-world enforcement using claims data alone.

We offer a distinct approach that complements traditional anomaly detection by accurately pricing financial risk. By grounding AI in proven actuarial principles, the model helps distinguish potential fraudulent behavior from legitimate clinical complexity, aiming to reduce false positives. Furthermore, our integration of advanced network analytics uncovers potential collusive activity, offering CMS tools to better identify coordinated fraud rings alongside individual actors.

To complement existing investigative workflows, our model produces a transparent and defensible Evidence Package for every provider. Typical "black box" AI models summarize risk in a single, non-decomposable score. In contrast, our "glass box" design provides empirical data explaining each component of the provider's risk score, giving investigators the "why" behind every flagged

[1] Calculated exposure is based on 2023 – 2024 claims data.

provider. This provides the clear, quantitative context required to streamline case reviews and support investigative decision-making.

Fraud, waste, and abuse remain persistent challenges that divert resources from patient care. CMS and taxpayers will benefit from our scalable solution that provides the clear, evidence-based indicators necessary to uncover broader fraudulent activity. By delivering transparent evidence, identifying potential network vulnerabilities, and providing statistically rigorous risk assessments, our solution aims to support CMS's evolution towards an increasingly data-driven defense, enabling timely, informed action to protect taxpayer resources and strengthen the Medicare program."

4 Chili-Cookoff Showcase & Networking Event

The CMS Crushing Fraud Chili Cook-Off Competition culminated in a Finalist Showcase and Networking Event held on December 9, 2025, at the MITRE campus in Woodlawn, Maryland. Finalist teams were invited to attend. While attendance was optional, representatives from all ten finalist teams were present. The event was intentionally structured as a high-value opportunity for finalists to engage directly with CMS leaders and share their innovations.

Each finalist team was assigned an exhibit booth to demonstrate and discuss its solution. This format enabled hands-on demonstrations, live Q&A, and direct conversations with CMS program, policy, and technical staff. In addition to the booth exhibits, finalists participated in one-on-one meetings with CMS leadership and technical experts. These sessions allowed CMS teams to gain deeper insight into the underlying analytics, data requirements, and implementation pathways for each solution.

Finalists heard remarks from CMS Administrator Dr. Mehmet Oz and Deputy Administrator & Chief Operating Officer Kim Brandt, as well as other senior leaders from CMS and the Center for Program Integrity (CPI). Dr. Oz congratulated the finalists and emphasized that their proposals represent the "best of the best" in the industry. He reinforced that the competition served as a mechanism for CMS to scan the market, understand cutting-edge capabilities, and build relationships with industry partners to crush FWA.

Kim Brandt shared the origin and vision of the Chili Cook-Off, explaining how early conversations with more than 100 external entities highlighted promising ideas that CMS needed a better way to evaluate. She noted that the event has already surfaced technologies and analytics that can enhance CMS' investigative operations and stop fraud closer to real time.

The showcase and networking event advanced CMS' understanding of available technologies, strengthened alignment around innovation priorities, and opened channels for ongoing collaboration with industry. For finalists, it provided direct access to decision-makers and increased visibility for their solutions.

Figure 4: Kim Brandt provides opening remarks

Figure 5: CMS Crushing Fraud Chili Cook-off Showcase and Networking Event

5 Closing Remarks

Each participant in the Crushing Fraud Chili Cook-Off supported CMS' fraud-fighting mission. This competition was designed as a hands-on learning opportunity to identify untapped technologies, surface innovations, and discover new approaches that will intensify our efforts to prevent fraud. The Chili Cook-Off created a space where technical experts, policy leaders, and front-line staff could come together, test ideas in a real-world context, and rapidly learn from one another. The resulting submissions were well-developed and played a role in helping CMS better understand the range of techniques currently available across the fraud prevention landscape. This public-private cross-collaboration is essential to ensuring that promising concepts do not remain theoretical but instead translate into practical tools and strategies that strengthen our programs and protect beneficiaries.

We are grateful to all the teams who joined our friendly competition and shared their expertise in pursuit of our shared goal, protecting American taxpayers by crushing fraud. Several submissions demonstrated potential for application to existing CMS work to incrementally improve our fraud fighting capabilities, with some approaches building upon techniques already in use and others introducing novel ideas that can be refined, further developed, and scaled over time. The outcomes of the competition also provide a foundation for informing future integration of analytic methods, including engaging CMS staff to assess which approaches best align with existing platforms and identifying pathways for continued exploration and collaboration. The insights and tools developed through this competition will guide us as we implement new capabilities and technologies to crush fraud and protect taxpayer dollars. We value your partnership and encourage everyone to stay tuned for future opportunities to partner with CMS.

Dr. Oz
X: @DrOzCMS
Instagram: @DrOzCMS
Facebook: CMS Administrator Dr. Oz
LinkedIn: Administrator Oz