Precrime: Structural-Metadata Screening of Pre-Financial Securities Issuers on EDGAR

Mark Phillips

June 2026

Abstract

This paper publishes the design and methodology of *Precrime*, a software system that screens securities issuers from raw EDGAR structural metadata — before any analyzable financial statement exists. Securities fraud is its most consequential application, but the screen's scope is broader: the same structural signals serve as predictors of corporate manipulation, of abnormal price trajectories (the engineered distribution and subsequent collapse), and of *entity duration* — how long a shell or microcap issuer is likely to persist before going dark or being absorbed. The method is a feedback loop running in reverse of the academic literature: rather than fitting a classifier on filed financials, we begin from the *outcome* — the corpus of published DOJ and SEC enforcement pleadings — and work backward to the structural fingerprints those matters left in the filing record, encoding them as weighted heuristics that then run forward as a predictor over live filing feeds. Postcrime labels train precrime flags. The labeled, enriched learning corpus is published openly so that third parties can inspect, reproduce, and extend the pipeline. The heuristic catalog is built from the full postcrime corpus and grounded most deeply in a primary-source anchor, *Exemplar 1* — an enterprise running schemes in parallel across more than 650 entities and CIKs — drawn from sealed federal pleadings that afford an unprecedented, paragraph-level view of a sophisticated, sustained manipulation operation that drew no regulatory or criminal intervention over decades, whose principal used defamation litigation to deter its public documentation. Exemplar 1 doubles as a litmus test of the screen's efficacy and a demonstration of why the decisive check requires the government's subpoena power over the shareholder register rather than public disclosures alone. Against a labeled corpus of 37,619 DOJ-SDNY and SEC litigation-release documents, three hand-traced cases show a mean lead time of 6.6 years from first structural signal to enforcement. We survey four mature detection families, argue Precrime occupies an empty pre-financial position complementary to all of them, and observe that comparable postcrime-to-precrime analysis is plausibly already practiced privately, as proprietary trade secrets, by quantitative hedge funds. Beyond a handful of retrospective cases, the paper outlines the program that scales the screen across the full EDGAR universe — seeding from every corpus entity, expanding multi-hop to a monitored watchlist of probable targets re-evaluated on an 18-month cadence — and serves as an outline of research-and-development activities open to collaborators and contributors. All quantitative figures here are illustrative projections on synthetic data, not measured results.

Contents

1. Introduction

The dominant paradigm in quantitative securities-fraud detection assumes a filed, analyzable financial record. Beneish's M-score reads eight accounting ratios off audited statements; Dechow's F-score is fit on the SEC's Accounting and Auditing Enforcement Releases (AAERs); Bao et al. (2020) train a RUSBoost ensemble on 28 raw financial variables. The NLP family — Loughran-McDonald sentiment, FinBERT — needs a narrative-rich 10-K to read tone and obfuscation from. The graph family needs enough entity history to build a network. All three are, structurally, *post-financial*: they fire once a company has produced the document they consume.

A consequential class of issuer never produces that document. Microcap and shell-company vehicles operate through thin exemption filings — Form D notices, S-8 employee-plan registrations, Regulation A offering circulars, Form 15 going-dark certifications — and through dormancy, name recycling, and shared filing infrastructure. By the time a ratio model or a tone classifier has anything to read, the engineered event — the distribution and dump — is already at or past completion. The firms that most need an early warning are precisely the ones the existing toolkit cannot see.

This paper publishes *Precrime*, a software system built for that blind spot, and the methodology behind it. Securities fraud is the most consequential and best-labeled outcome, and we describe the system primarily against it, but its outputs generalize: the structural score is equally a predictor of corporate manipulation, of the timing and likelihood of an engineered distribution, and of issuer duration. The approach inverts the usual modeling direction. Instead of fitting a model on financials and hoping it generalizes to enforcement outcomes, Precrime starts from the outcomes — the public record of DOJ and SEC pleadings — and reconstructs the structural signatures those adjudicated matters left in EDGAR years earlier. Those signatures become weighted heuristics that run forward as a predictor on live filing feeds. The loop is closed: each newly published enforcement action is a fresh label that refines the heuristics, so the postcrime corpus continuously trains the precrime flag.

Two design commitments follow. First, Precrime is *recall-first*: at the top of the funnel the cost of missing a pre-operational shell exceeds the cost of an extra flag, so the system accepts a high false-positive rate, absorbed by a tiered human-review cadence (§6). Second, the heuristics are not abstract — they are reverse-engineered from documented conduct. The anchoring case study is *Exemplar 1* (§5), an operation whose mechanics are laid out at paragraph granularity in sealed federal pleadings, providing the kind of ground truth that adjudicated-but-thinly-described enforcement summaries rarely supply.

We do not claim Precrime beats the post-financial models on their own ground. We claim they share an upstream blind spot, that Precrime fills it, and that the natural research move is composition, not competition. We also note the obvious: postcrime-to-precrime structural analysis of this kind is plausibly already conducted, privately and without publication, by quantitative hedge funds that hold it as a proprietary trade secret. Publishing the method openly — as a companion to the Precrime documentation — is the point of this paper. The paper also serves as an outline of the research-and-development activities the system requires (§10) — an open invitation to individuals and organizations to collaborate or contribute resources.

Contributions: (1) the postcrime-to-precrime methodology and feedback loop (§3); (2) the open, enriched learning corpus and its per-dimension optimization (§4); (3) a primary-source case study, Exemplar 1, mapping documented conduct to specific heuristics (§5); (4) a description of the Precrime scoring model, catalog, tiering, and action ladder (§6); (5) a survey of the four detection families

and an explicit complementarity analysis (§7–§8); (6) an enforcement application — an automated share-register audit that turns a flag into dated, attributable proof (§9); (7) a population-scale watchlist program that runs the screen across all of EDGAR with prospective 18-month monitoring (§10); and (8) a proposed evaluation protocol centering lead time (§11). The quantitative projections in §12 are synthetic and labeled as such throughout. A companion README distributed with this paper records collaboration and benchmarking plans.

2. The Pre-Financial Gap

Consider the lifecycle of a microcap shell. It is incorporated, registers an exempt offering via Form D, perhaps files an S-1 or S-8 with skeletal disclosures, and then goes quiet — sometimes filing a Form 15-12G to suspend reporting obligations. Months or years later it is acquired in a reverse merger, control changes via an 8-K Item 5.02 event, promotional volume spikes, and the stock is distributed and dumped. The enforcement action, if any, follows years after that.

The post-financial models can only engage in the final phase, when financials and narrative finally exist — and by then the harm is largely done. The pre-financial phase emits a different kind of evidence: not numbers but *structure*. The same registered agent files for a cluster of otherwise-unrelated issuers; the same opinion letter and Regulation S issuance language recurs verbatim across shells; a name is recycled from a previously dissolved entity; a filer goes dark on a predictable cadence. These are signals in metadata and boilerplate, not in audited ratios or management tone. The same signals also forecast the issuer's path — whether a manipulative distribution is likely and how long the entity will persist. Precrime is built to read them.

3. From Postcrime Pleadings to a Precrime Predictor

The Precrime methodology is a closed feedback loop with four stages.

(1) Corpus of outcomes. The system maintains a labeled corpus of 37,619 DOJ Southern District of New York and SEC litigation-release documents (`corpus.db`), each annotated with its enforcement date. This is the asset the academic literature lacks: a labeled, lead-time-annotated bridge between structural EDGAR signals and adjudicated outcomes.

(2) Backward extraction. For each adjudicated matter, the system traces the issuer's EDGAR history *prior* to enforcement and identifies the structural and document-level fingerprints present at the pre-financial stage — shared filing agents, recycled offering boilerplate, reverse-merger form sequences, name recycling, going-dark cadences. These are the artifacts that were observable, in the public filing record, years before any model fit on financials could have engaged.

(3) Forward prediction. The extracted fingerprints are encoded as weighted heuristics (§6) that run forward over live EDGAR feeds, scoring every issuer as it files. The score is read three ways: as an enforcement-risk flag, as a likelihood-and-timing estimate for an engineered distribution, and as a predictor of entity duration. Issuers that clear the gate enter a tiered review queue.

(4) Confirmation and re-labeling. Flags are escalated through an action ladder to confirmation; confirmed outcomes — and every newly published DOJ or SEC pleading — re-enter the corpus as fresh labels, refining the weights and surfacing new fingerprints. Postcrime data thus continuously trains the precrime predictor.

The inversion is the methodological core: the literature asks "given these financials, is this firm misstating?"; Precrime asks "given the public record of which issuers *were* prosecuted, what did their filings look like before anyone noticed, and which live issuers look like that now?" The first question needs financials the target may never file; the second needs only metadata every issuer emits from its first filing.

4. The Learning Corpus and Enrichment Pipeline

The corpus that closes the feedback loop of §3 is not a static file; it is an open, browsable learning corpus, published with full implementation details so that third parties can inspect, reproduce, and extend the pipeline.[1] Every enforcement document is *enriched* before it trains the predictor: named entities (issuers, individuals, registered agents, counsel), filing identifiers (CIK, accession), dates, and cross-references to related matters are extracted and linked, turning an unstructured pleading into a structured, queryable record. An enriched indictment — the kind of document from which the structural fingerprints of §3 are read — illustrates the output of this step.[2]

Enrichment is paired with continuous learning optimization. The system exposes a statistics dashboard in which each chart is a *dimension* of the corpus — lead-time distribution, signal co-occurrence, agent-overlap density, form-sequence frequency, and so on — and each dimension is an independent optimization target: the heuristic weights (§6.2) are tuned against the labeled outcomes sliced along that dimension, so the predictor is re-fit per dimension as the corpus grows rather than fit once against a single opaque loss.[3] Postcrime data and precrime weights are connected through an explicit, inspectable set of per-dimension objectives.

The dimensions above are a starting set, not a ceiling. The largest gains come from data the current catalog only approximates. Priority integrations include: explicit entity graphs with learned edge weights over individuals and entities; litigation records and their *distance to indictment* (how many filings, entities, or years separate a given issuer from a charged matter); CourtListener dockets; SEC and DOJ indictments and litigation releases as first-class nodes; named witnesses; multi-hop traversal through shared CIKs and accessions to reach participants no single filing names; and the recurring participation *patterns* those hops expose. Precrime's `filer_agent_overlap` and `name_recycling` heuristics (§6.2) are single-hop shadows of this graph; the graph extension noted in §7.3 is where the recall gains lie.

Exemplar 1 illustrates why multi-hop linkage matters (sealed pleadings, §5; specifics deferred to future publication). In that matter a small set of recurring principals — a controlling securities attorney, shared registered agents, and repeat signatories — reappears across a chain of nominally unrelated vehicles. Following shared agents, signatories, and accession chains outward by additional hops surfaces further common participants that single-issuer screening never connects. A graph that weights these recurring participants by their proximity to charged matters converts a scatter of individually-unremarkable shells into a single, high-score network, which is precisely the signal a pre-financial screen should escalate.

[1] The open learning corpus is published at postcrime.atsignhandle.xyz, with implementation details at postcrime.atsignhandle.xyz/docs.

[2] Enriched indictment example, corpus browser: document 36089.

[3] Per-dimension learning-optimization dashboard, corpus browser: statistics view.

5. Exemplar 1: A Primary-Source Case Study

Heuristics reverse-engineered from enforcement summaries inherit the summaries' coarseness: a litigation release states that a defendant "used shell companies" without enumerating the filing-level mechanics. The Precrime catalog is therefore built from the full postcrime corpus — thousands of enforcement matters, together with the structural intuitions and extractions distilled from them (§3–§4) — and grounded most deeply in a single primary-source anchor documented at paragraph granularity in a set of sealed federal pleadings, referred to here as *Exemplar 1*.[4] Exemplar 1 is not one case but an enterprise: across hundreds of paragraphs and accompanying tables, the pleadings set out a sustained, sophisticated manipulation operation running schemes in parallel across more than 650 entities and CIKs, built on shell formation, reverse mergers, and recycled offering infrastructure. It served as a litmus test of precrime's efficacy — an investigation into sophisticated bad actors who evaded detection for decades, precisely the population the post-financial toolkit never engages.

Two properties make Exemplar 1 uniquely valuable as design substrate. First, it describes an operation that drew *no regulatory or criminal intervention* over an operating life spanning decades — a matter that the post-financial detection apparatus, by construction, never engaged. It is precisely the false negative the existing toolkit produces, and therefore the ideal template for a system meant to catch such cases earlier. Second, the documentation exists at all only because of an unusual primary source. The enterprise's principal — a securities attorney and sole controlling shareholder — pursued defamation litigation against those who would publish analyses of the operation, a pattern of suppression that helps explain why detailed accounts of active microcap manipulation are so rare in the open literature.[5] The sealed pleadings, brought by a relator, are one of the few paragraph-level descriptions of such an operation in existence. Exemplar 1 also marks where the decisive test lies: not in the public disclosures, which were self-certified and clean on their face, but in the shareholder register the government alone can subpoena and weigh against those disclosures (§9).

The pleadings' mechanics map directly onto the catalog: the documented reuse of filing agents and counsel across nominally unrelated issuers informs `filer_agent_overlap` and `opinion_letter`; the reverse-merger structures inform `reverse_merger_chain`; the recycling of dissolved entity names informs `name_recycling`; the Regulation S and Rule 144 mechanics inform `reg_s_issuance` and `form_144_outlier`. The case is not a statistical sample — it is the worked example from which the abstract heuristics draw their concrete shape, complementing the breadth of the 37,619-document corpus with depth on a single, fully-traced operation.

[4] The underlying pleadings for Exemplar 1 are sealed. Full documentation of the exemplar — parties, entities, docket, and transaction specifics — will be published when the court unseals the pleadings. References here are to the sealed factual narrative as the design substrate for the heuristic catalog, not as adjudicated findings; the allegations are unproven.

[5] The principal's pursuit of defamation litigation to deter publication of analyses of the operation is itself among the conduct described in the sealed pleadings. The methodological point is that such suppression is why primary-source documentation of active manipulation is scarce, which is what makes the pleadings' filing-level detail valuable for heuristic design.

6. The Precrime System

6.1 Scoring model

Each issuer e receives an additive score over a catalog of binary heuristics:

$$S(e) = \sum_{h} w_h \, \mathbf{1}[h \text{ fires on } e],$$

where each heuristic h carries an integer weight $w_h \in \{2, 3\}$. An issuer is gated for review only when it both trips at least two distinct signals and clears a score threshold:

$$\text{gate}(e) = \big(n_{\text{sig}}(e) \geq 2\big) \;\wedge\; \big(S(e) \geq \tau\big), \qquad \tau = 3.$$

The two-signal floor suppresses singletons — a lone weak match is treated as noise — and the threshold τ sets the minimum corroborated weight required to enter the review queue. Both are deliberately conservative on precision and permissive on recall. The same score $S(e)$ is read not only as an enforcement-risk flag but as a monotone predictor of manipulation likelihood and of shortened entity duration.

6.2 Heuristic catalog

The catalog has 16 heuristics spanning two classes. **Structural** heuristics read filing infrastructure and entity history: `filer_agent_overlap` (shared registered filing agent across nominally unrelated issuers, $w = 3$), `reverse_merger_chain` (form-type sequence consistent with a reverse-merger vehicle, $w = 3$), `name_recycling` (entity name reused from a prior dissolved filer, $w = 3$), and `going_dark_blackout` (Form 15-12G suspension following a characteristic filing gap). **Body-text** heuristics fingerprint recycled offering boilerplate: `reg_s_issuance` (Regulation S issuance language), `promissory_clauses`, `opinion_letter` (recycled counsel opinion letters), and `form_144_outlier` (anomalous Rule 144 resale notices). The body-text heuristics detect the document-template reuse that characterizes shell factories; the structural heuristics detect the shared human and corporate infrastructure behind them.

6.3 Tiering and action ladder

Gated issuers are bucketed into three review tiers by score, each with its own cadence:

Tier	Score range	Review cadence
T1	$S \geq 10$	Daily
T2	$6 \leq S \leq 9$	Weekly
T3	$3 \leq S \leq 5$	Quarterly

The action ladder escalates evidence rather than asserting wrongdoing: **flag** (enter the queue), **label** (attach corroborating signals and corpus matches), **confirm**. Confirmation requires an independent corroborating event — an 8-K Item 5.02 change-of-control coinciding with abnormal trading volume — converting a structural suspicion into an actionable, dated lead.

6.4 Lead-time evidence

Three hand-traced cases from the corpus illustrate the lead time from first structural signal to SEC complaint:

Case	Signals fired	Lead time to complaint
A	17 of 38	6.3 years
B	11 of 38	7.1 years
C	13 of 38	6.3 years
Mean		**6.6 years**

These three cases are not a validation sample — they are chosen retrospectively and are discussed as such in §13, and are scaled into a prospective, population-wide program in §10 — but they motivate the central claim: the structural evidence was present, in EDGAR metadata, years before the post-financial models could have engaged. Full system documentation is hosted at the project docs site.[6]

7. Related Work

We survey four detection families. The synthesis throughout is that each is mature on its own terrain and each presumes a post-financial target.

7.1 Financial-ratio machine learning

Beneish's (1999) M-score is an eight-variable probit on accruals and growth ratios with a manipulation threshold near -1.78; it is interpretable but degrades out of sample. Dechow, Ge, Larson and Sloan (2011) fit the F-score via logistic regression on AAER cases, where $F > 1$ indicates above-base-rate manipulation probability; the F-score also supplies labels and features for downstream ML. Bao, Ke, Li, Yu and Zhang (2020) move to a RUSBoost ensemble over 28 raw financial variables, drawn from 413 AAER cases spanning 964 fraudulent firm-years out of 146,045 total — a base rate near 0.7%. They introduced AUC and NDCG@k as the appropriate metrics for this extreme class-imbalance regime, reporting a test AUC near 0.725 (public code: JarFraud/FraudDetection). Walker (2021) and related critiques caution that such models are fragile to look-ahead bias and out-of-sample drift. The Cecchini et al. (2010) line and subsequent cost analyses illustrate the operational burden: roughly 47 true matters per year against a model surfacing on the order of 1,891 predictions implies an alert ratio near 40:1 — which is precisely why tiered human review is necessary.

Gap. Every model in this family needs filed, audited financials. Precrime's cohort — Form-D-only issuers, dormant shells, name-recyclers — files no analyzable financials and is therefore invisible to ratio ML. Precrime is the complementary recall-first structural screen upstream of it.

7.2 NLP and EDGAR text analysis

Loughran and McDonald (2011, 2014, 2016) built the finance-tuned sentiment dictionary and readability/obfuscation measures that anchor textual fraud research; their Software Repository for Accounting and Finance (SRAF) is the standard resource. Purda and Skillicorn (2015) showed a data-driven bag-of-words deception classifier outperforming fixed dictionaries. Loughran, McDonald

[6] Precrime system overview, *Precrime Project Documentation* (2026), https://sec.atsignhandle.xyz/docs/precrime-overview.html.

and Yun (2009) — "A Wolf in Sheep's Clothing" — established that elevated ethics-and-reassurance language can itself be a manipulation signal, a precedent for treating reassuring boilerplate as positive evidence. FinBERT (Huang, Wang and Yang, 2023) brought transformer language modeling to financial text, and subsequent work (Management Science, 2022) confirmed ML approaches generally outperform dictionary methods. Larcker and Zakolyukina (2012) extended deception detection to earnings-call transcripts.

Gap. The entire NLP stack assumes a narrative-rich 10-K. Precrime targets thin filings (S-1, S-8, Reg-A, Form D) through structural document fingerprints rather than tone. **Extension.** FinBERT or Purda-Skillicorn tone could serve as a T3-to-T2 tie-breaker applied *after* a structural flag, where a narrative document does exist.

7.3 Alternative data, graphs, and legal signals

A substantial line of work discovers shell networks via shared registered agents, addresses, and auditors, surfacing dense subgraphs; cycle detection identifies reverse-merger chains, and PageRank and community detection (commonly on Neo4j, drawing on AML practice) rank suspicious clusters. Precrime's `filer_agent_overlap` and `name_recycling` heuristics are single-hop versions of these constructions; multi-hop graph expansion is the obvious recall-gain extension. The consensus in this family is a rules-plus-ML hybrid. Microcap information scarcity — the exact regime where ratio and tone models starve — is where graph and alternative-data signals add the most, and it is Precrime's cohort. The institutional consumers already exist: the SEC Market Abuse Unit and Microcap Fraud Task Force. Whistleblower submissions under Section 21F are a label augmenter. Adjacent quantitative work includes Mai et al. (2019) on deep-learning bankruptcy prediction and La Morgia et al. (2021) on crypto pump-and-dump ML.

Gap. Academic graph work rarely binds to a labeled EDGAR enforcement corpus with *measured lead time* — exactly the asset Precrime contributes (§4).

7.4 Trading-anomaly detection and the confirmation layer

The state of the art in trade-level surveillance is deep unsupervised and graph modeling on limit-order-book data: LSTM autoencoders, VAEs and GANs for anomaly scoring, contrastive methods (arXiv 2508.17086), social-fusion models (AIMM, arXiv 2512.16103), and human-in-the-loop triage. Reported deployments cut false-positive rates from 51.4% to 11.9%. These systems fire near the dump, when manipulation is contemporaneously visible in order flow. Precrime fires years earlier on structure. The two are not rivals but stages: a structural pre-financial screen feeding a trade-level confirmation layer.

8. Positioning and Complementarity

Precrime is not a better fraud classifier; it is a screen positioned upstream of every family in §7. The distinction is the input it consumes and the moment it fires.

The unifying claim: ratio ML, NLP, and graph models share a *post-financial* assumption, and trading-anomaly models a *contemporaneous* one. Precrime occupies the empty pre-financial position. Its recall-first design is the correct posture for that slot — at the top of the funnel the cost of a

Family	Required input	Fires when	Precrime relation
Ratio ML (§7.1)	Audited financials	Statements filed	Upstream; surfaces firms before any financials exist
NLP (§7.2)	Narrative 10-K	Annual report filed	Upstream; tone usable as a tie-breaker after a flag
Graph / alt-data (§7.3)	Entity / network history	Network observable	Overlapping; Precrime = single-hop, graph = multi-hop extension
Trading anomaly (§7.4)	Order-book / trade data	Near the dump	Downstream confirmation layer Precrime feeds

Table 1: Precrime's position relative to the four detection families.

missed pre-operational shell exceeds the cost of an extra flag, and the tiered review cadence (§6.3) absorbs the resulting alert volume.

It is worth stating plainly that the methodology is not exotic. Reverse-engineering structural fingerprints from public enforcement records and running them forward as a screen is the kind of analysis private quantitative hedge funds plausibly already perform — but hold as proprietary trade secrets, with no incentive to publish. The contribution here is not novelty of intuition but open publication of a working method, its learning corpus, and its grounding case study, so that regulators, researchers, and third-party implementers can reproduce and extend it rather than reinvent it behind closed doors.

9. Enforcement Application: Automated Share-Register Audit

The natural consumer of a recall-first pre-financial screen is not a trading desk but a resource-constrained enforcement agency. The SEC's enforcement capacity is small relative to the universe of microcap and shell issuers it must police; the binding constraint is analyst attention, not legal authority. A screen that narrows tens of thousands of issuers to a ranked review queue is precisely the leverage such an agency lacks — and, unlike a private fund, an agency can confirm a flag with records it has the power to compel and that no outside analyst can obtain.

That confirming record is the share register. The structural metadata that drives Precrime ends at the filing boundary; the decisive corroboration lives one layer deeper, in the certificate-deposit and beneficial-ownership records held by transfer agents and the depository system. An automated audit mechanism with API fulfillment from transfer agents (for example, Pacific Stock Transfer), from the depository nominee CEDE & Co. and the Depository Trust Company that hold shares in street name, and from issuer shareholder registries, lets the agency pull the deposit history that a Precrime flag points it to. The screen says *look here*; the register answers *here are the certificates from the predecessor entity, deposited on this date, by this controlling party.* A predecessor's certificates surfacing on a specific date ahead of a distribution is the corroborating evidence a metadata screen cannot itself produce but can precisely target — the difference between a statistical suspicion and a dated, attributable act.

Exemplar 1 supplies the worked example (sealed pleadings, §5; specifics deferred to future publication). Ahead of a successor entity's distribution and dump, a large block of certificates was deposited from a predecessor entity. The controlling principal — a securities attorney and the only shareholder positioned to authorize that deposit — sits at the single point of control. The mechanism turns on self-certification: entity and accounting reporting is self-certified, as are the counsel opinion letters that authorize removal of the Rule 144 restrictive legend. The underlying

move is a stock swap converting restricted predecessor bearer certificates into successor shares held in street name through the depository nominee, laundering restricted predecessor paper into freely-trading stock. Precrime flags this pattern from its origin — traceable back through the chain of predecessor entities — years before the dump, on structural metadata alone; the share-register audit then supplies the certificate-deposit proof that closes it.

The same template recurs. A cell-network analysis across similarly-operating practitioners — the multi-hop entity-graph integration of §4 — reveals a common practice well predating cryptocurrency: a small, recurring set of agents, counsel, and controlling parties wiring predecessor certificates into street-name stock through self-certified legend removals. The ICO analogy is exact. Where a programmer mints a token, here a securities attorney gates a CUSIP: the opinion letter is the minting function and the restrictive-legend removal is the unlock. Reading the network of which attorneys gate which CUSIPs is the same graph problem as tracing token deployers across contracts, and it is where the recall gains of §4's multi-hop integration and the graph extension of §7.3 are realized.

10. Scaling to a Population Watchlist

The three hand-traced cases of §6.4 are a proof of concept, not the program. They show that the structural evidence was present years early in individual matters; they do not yet show that the screen can be run forward, at the scale of the entire market, to surface targets *before* enforcement. That is the central research-and-development activity this paper outlines, and the work to which outside contributors are invited.

The program seeds from the corpus. Every CIK, entity, and individual named in the 37,619-document postcrime corpus (§4) is taken as a starting node and run against the entire SEC EDGAR universe. From each seed, the multi-hop entity graph (§4, §7.3) is expanded along shared registered agents, filing accessions, signatories, and counsel to reach individuals and entities that no single enforcement document names but that sit one or more hops from a charged matter. The result is a ranked, deduplicated **watchlist**: a current, countable population of probable entities and individuals — far larger than the enforced set — each carrying a Precrime score and tier (§6.3) but no enforcement record yet.

The watchlist is then monitored prospectively. Each member is re-evaluated on an 18-month cadence by default; as its score or tier rises, or a confirming event fires (§9), the interval shortens. New filings are re-scored as they arrive, multi-hop neighbors are refreshed, and any member that subsequently draws an enforcement action becomes a *prospective* lead-time data point — converting the retrospective n=3 evidence of §6.4 into a falsifiable, population-scale experiment whose hit rate and lead-time distribution can be measured rather than asserted. The monitoring window is the unit of the experiment: a shorter interval raises sensitivity and reviewer cost; a longer one conserves attention. Calibrating that trade-off against realized outcomes is itself a research output.

This is where the paper functions as an outline of R&D activities for collaborators. The program decomposes into independently-fundable workstreams — corpus enrichment and entity resolution, multi-hop graph construction and edge-weight learning, the live EDGAR ingestion and scoring pipeline, the monitoring-and-re-evaluation scheduler, and the share-register confirmation interface of §9 — each of which an individual developer or an organization can adopt, extend, or resource. The companion README records specific institutional collaborators and the first joint benchmark;

this section records the activities themselves, open to anyone seeking to contribute compute, labeled data, domain expertise, or individual development effort.

11. Proposed Evaluation Protocol

A credible evaluation must measure Precrime on the dimension that distinguishes it. We propose a joint benchmark against the standard baselines on a shared, labeled microcap/shell holdout drawn from the enforcement corpus, reporting four metrics:

1. **AUC** — ranking quality under extreme class imbalance, following Bao et al. (2020).
2. **NDCG@k** — top-of-list precision, the operationally relevant quantity given finite reviewer capacity and the ~40:1 alert ratio.
3. **Lead time** — median years from first model signal to SEC complaint. This is the metric the existing literature omits and the one on which a pre-financial screen must be judged. A model with lower AUC but multi-year lead time is operationally superior for prevention.
4. **Price-based outcome** — enforcement is rare, slow, and bounded by agency resources, so it under-labels true positives. The flagged issuer's subsequent market pricing — abnormal returns and the post-distribution collapse of the target stock — is a faster-resolving, higher-coverage proxy outcome that closes in months rather than years. It is the near-term evaluation target; enforcement is the long-run one.

Baselines: RUSBoost (Bao et al.) and FinBERT (Huang et al.) on the subset of the holdout where their required inputs exist, with Precrime evaluated on the full cohort including the pre-financial firms the baselines cannot score. The asymmetry of coverage is itself a finding. The illustrative figures in §12 anticipate the shape of such a run; they are not its results. The companion README records the specific baseline implementations and the collaborators whose labeled data such a benchmark requires.

12. Illustrative Projections

The figures below are **illustrative projections on synthetic data, not measured results.** They visualize the hypotheses the protocol in §11 would test and are included to make the expected evaluation shape concrete.

13. Limitations

This is a position-and-survey paper publishing a system in development, and its claims are bounded accordingly.

- **n = 3 traced cases.** The 6.6-year mean lead time rests on three hand-traced cases, not a sample. It is motivating evidence, not measurement; §10 outlines the population-scale program that would replace it with a measured, prospective distribution.
- **Prospective watchlist unproven.** The §10 monitoring program is described, not yet run; its hit rate and lead-time distribution remain hypotheses until the 18-month windows close.
- **Single anchoring case study.** Exemplar 1 supplies depth, not breadth; its underlying pleadings are sealed and its allegations unproven, and the specifics are deferred to future

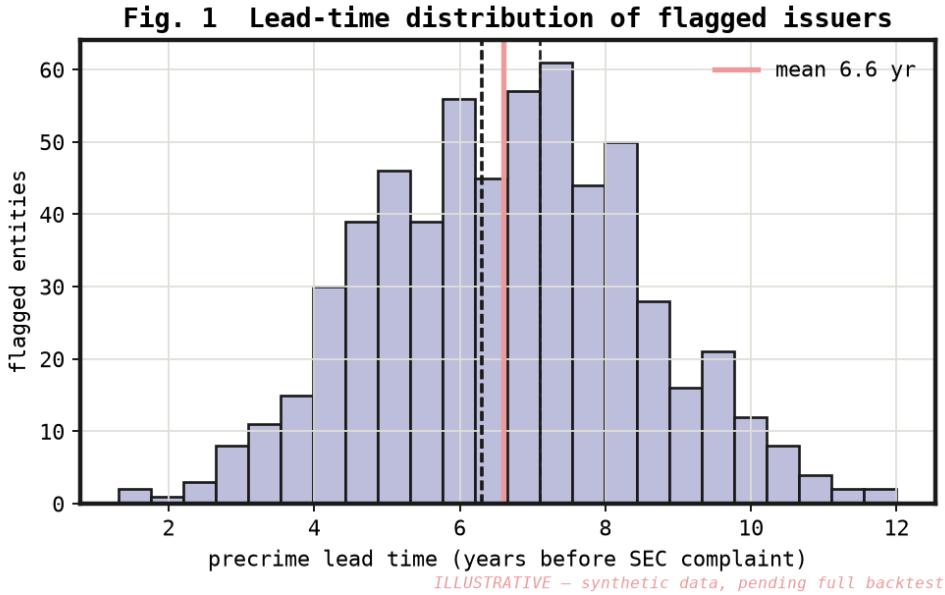

Figure 1: Projected lead-time distribution from first structural signal to enforcement. *Illustrative — synthetic data, not measured.*

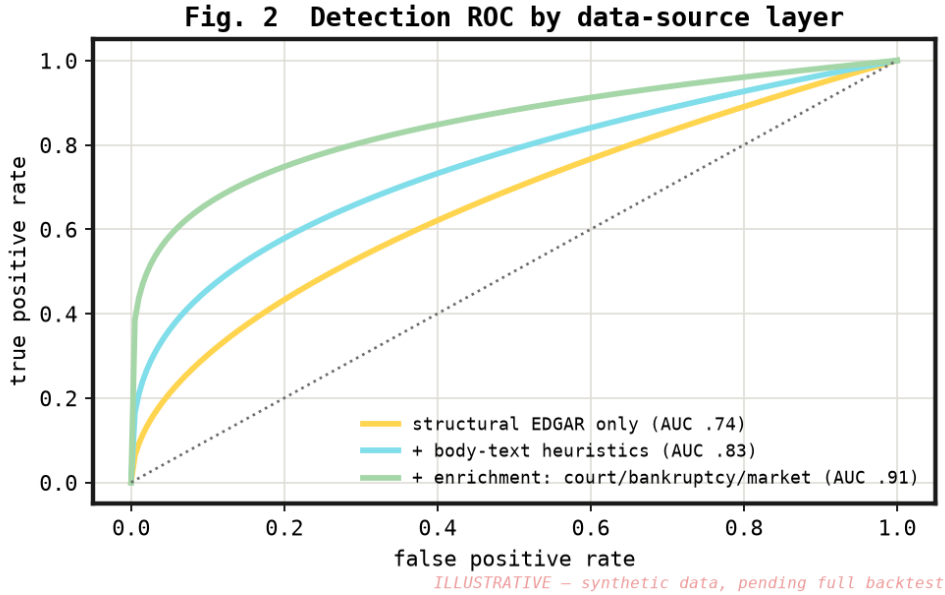

Figure 2: Projected ROC for the structural screen on a labeled holdout. *Illustrative — synthetic data, not measured.*

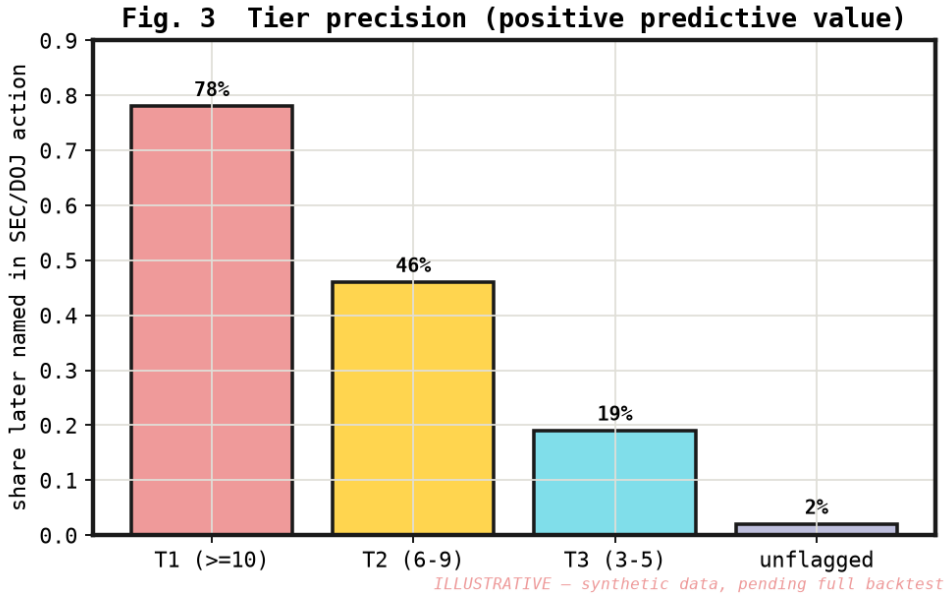

Figure 3: Projected precision by review tier (T1/T2/T3). *Illustrative — synthetic data, not measured.*

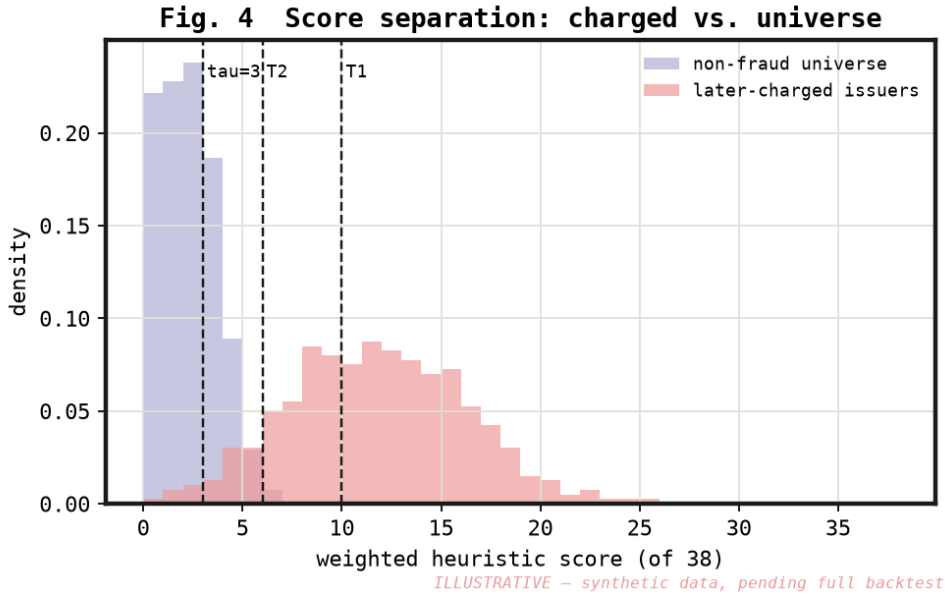

Figure 4: Projected score distribution across the issuer population. *Illustrative — synthetic data, not measured.*

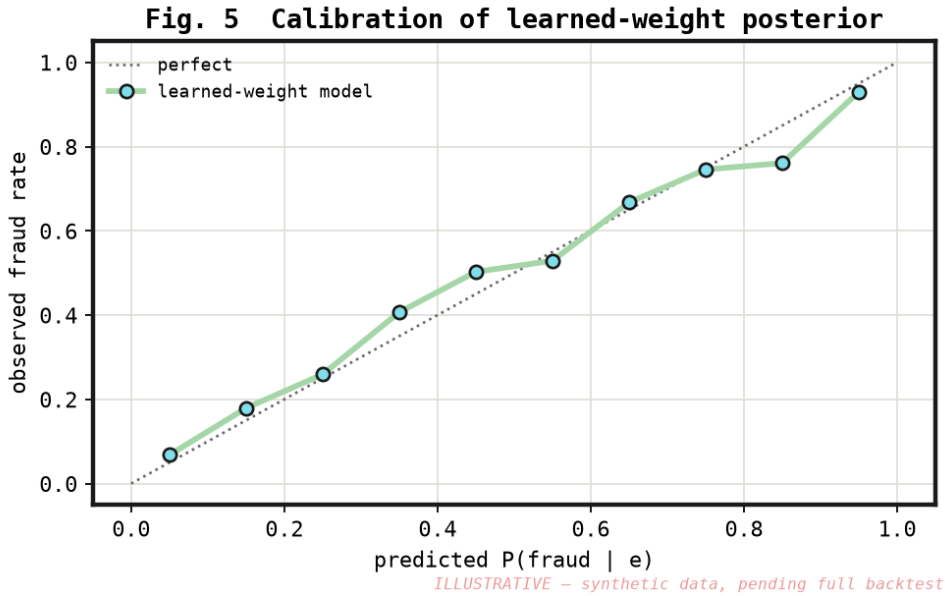

Figure 5: Projected calibration of gated scores against enforcement outcomes. *Illustrative — synthetic data, not measured.*

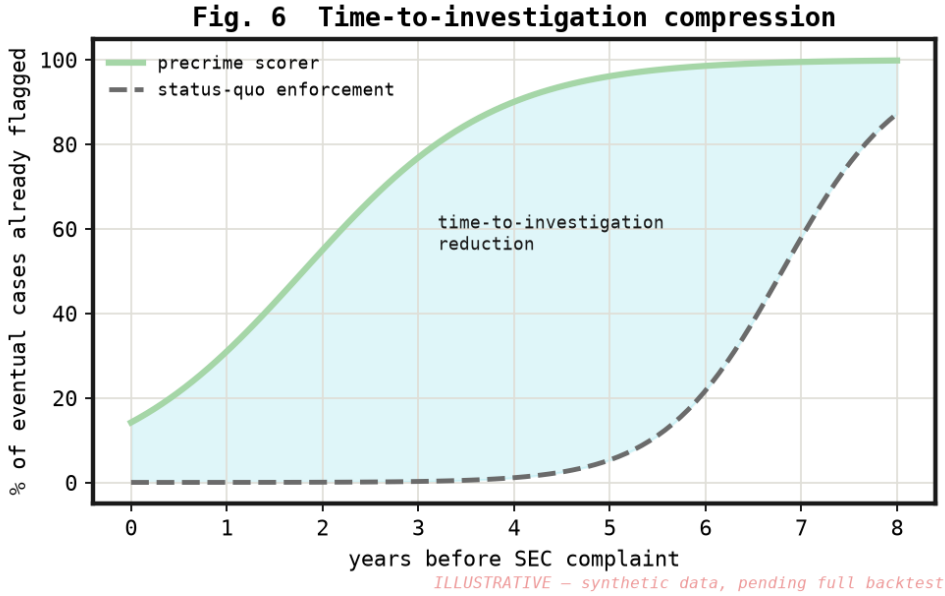

Figure 6: Projected time-to-investigation under tiered review. *Illustrative — synthetic data, not measured.*

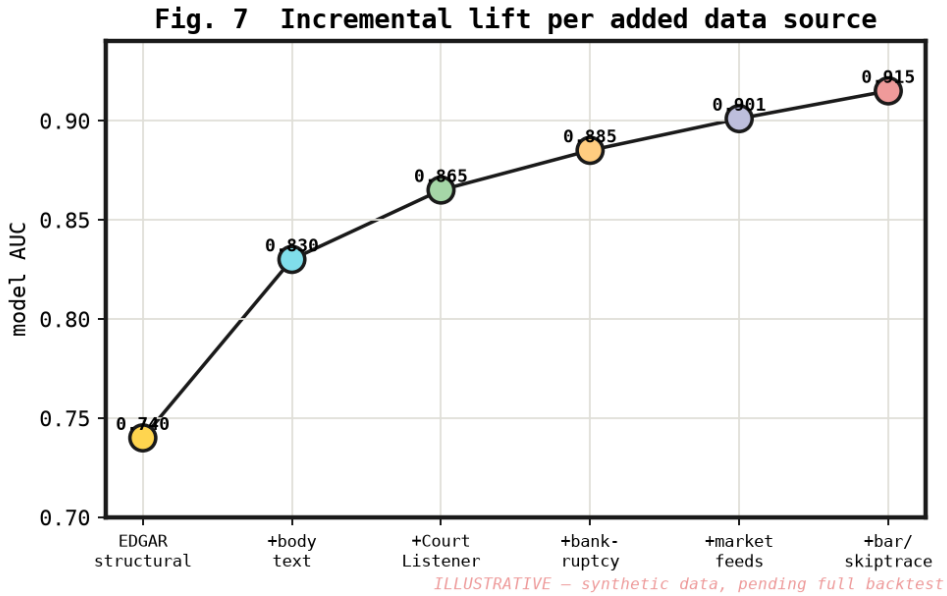

Figure 7: Projected recall ablation by signal source (structural vs. body-text). *Illustrative — synthetic data, not measured.*

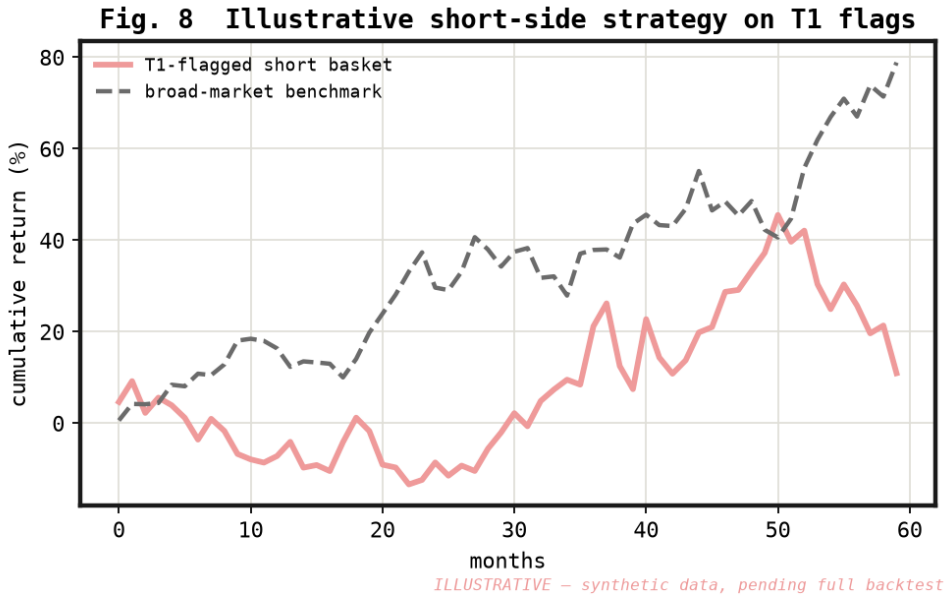

Figure 8: Projected short-side alpha from early structural flags. *Illustrative — synthetic data, not measured.*

publication. It grounds the heuristics' shape but cannot establish their population-level performance.

- **Synthetic figures.** Every figure in §12 is synthetic and labeled illustrative; none reflects a measured evaluation.
- **Selection bias.** The three traced cases were chosen after their enforcement outcomes were known. Retrospective selection inflates apparent performance and cannot estimate recall.
- **Uncalibrated threshold.** The gate $\tau = 3$ and the tier cutoffs are set by judgment, not calibrated against held-out outcomes.
- **False-positive cost.** A recall-first screen at a roughly 40:1 alert ratio imposes a real reviewer burden; the tiering mitigates but does not eliminate it, and the operating point is unvalidated.

The protocol in §11 is the path from these limitations to defensible empirical claims.

14. References

Bao, Y., Ke, B., Li, B., Yu, Y. J., & Zhang, J. (2020). Detecting accounting fraud in publicly traded U.S. firms using a machine learning approach. *Journal of Accounting Research*, 58(1), 199–235.

Beneish, M. D. (1999). The detection of earnings manipulation. *Financial Analysts Journal*, 55(5), 24–36.

Cecchini, M., Aytug, H., Koehler, G. J., & Pathak, P. (2010). Detecting management fraud in public companies. *Management Science*, 56(7), 1146–1160.

Dechow, P. M., Ge, W., Larson, C. R., & Sloan, R. G. (2011). Predicting material accounting misstatements. *Contemporary Accounting Research*, 28(1), 17–82.

Huang, A. H., Wang, H., & Yang, Y. (2023). FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research*, 40(2), 806–841.

La Morgia, M., Mei, A., Sassi, F., & Stefa, J. (2021). Pump and dumps in the cryptocurrency era: Real-time detection of cryptocurrency market manipulations. arXiv:2105.00733.

Larcker, D. F., & Zakolyukina, A. A. (2012). Detecting deceptive discussions in conference calls. *Journal of Accounting Research*, 50(2), 495–540.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance*, 66(1), 35–65.

Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4), 1187–1230.

Loughran, T., McDonald, B., & Yun, H. (2009). A wolf in sheep's clothing: The use of ethics-related terms in 10-K reports. *Journal of Business Ethics*, 89(S1), 39–49.

Mai, F., Tian, S., Lee, C., & Ma, L. (2019). Deep learning models for bankruptcy prediction using textual disclosures. *European Journal of Operational Research*, 274(2), 743–758.

Purda, L., & Skillicorn, D. (2015). Accounting variables, deception, and a bag of words: Assessing the tools of fraud detection. *Contemporary Accounting Research*, 32(3), 1193–1223.

Tuyet-Mai, T., et al. (2024). Deep learning for market manipulation detection. *Journal of Finance and Data Science*.

Walker, S. (2021). The fragility of accounting-fraud prediction models. *Econ Journal Watch*.

Exemplar 1 — sealed federal pleadings; parties, docket, and transaction specifics to be published upon unsealing.

Contrastive anomaly detection on limit-order-book data (2025). arXiv:2508.17086.

AIMM: Social-fusion market-manipulation detection (2025). arXiv:2512.16103.

U.S. Securities and Exchange Commission, Market Abuse Unit and Microcap Fraud Task Force — enforcement program references.