

Silent Failure Modes in the Automated Extraction of Merger Terms from Primary Filings, and a Fail-Closed Gate That Detects Them

Maksim Aleksandrovich Andreev, Joseph Howden
Quark Quantitative Research

Abstract

Event-driven research on corporate control transactions depends on reading consideration terms out of primary filings. We report five failure modes of that operation, observed in an instrumented production system whose extracted terms were displayed on a decision surface, and each documented against the filing and the commit that repaired it. The modes are the selection of the wrong document within the correct submission, the resolution of a structural placeholder to a real and unrelated listed company, the reading of the wrong clause within the correct sentence, the inversion of acquirer and target, and the emission of a sized instruction by a record already marked defective. Their common property is that none produces an error, a gap, or a diagnostic: each returns a value of the expected type and plausible magnitude, and one of them wrote absence of evidence into the record as a clean read with the human-review flag set false. The five are not independent degrees of severity of one defect. A single card carried four of them in series, and the resulting displayed price of \$107.19 against a \$4.60 quote was recomputed on every build from a live quote for an unrelated ticker, so it was not stable between two builds of the same day. We then describe four detection practices assembled into a fail-closed gate: corroboration of a resolved counterparty against the document its terms were read from, role direction read from the definitional recital rather than from name matching, clause-typed provenance on every extracted number, and certification of the extractor by blind reads against a bar frozen in advance. On the population where they can be measured, the practices do not nest: an audit of 22 cached merger filings returned five role inversions and one unresolvable role, and role inversion is invisible to name corroboration by construction, because the counterparty is genuine and, in one instance, occurs 35 times in the text. The individual techniques are drawn from the anaphora-resolution and record-linkage literatures and are not new. What is reported here is their assembly into a gate over primary filings that refuses rather than annotates, together with the blind-read record of the extractor that gate protects: a pre-registered bar of 13 of 15 filing families, failed at 2 of 15 on first reading and at 7 of 15 on second, and passed at 14 of 15 on the third after four rounds of outcome-blind repair.

1. Introduction

A merger-arbitrage position is a claim on a number printed in a filing. The consideration per share, the exchange ratio, the outside date and the identity of the parties determine the payoff, and each is stated in prose in a document the Commission publishes at no charge. Automating their retrieval is therefore the operation on which filing-based event research rests, and its quality bounds every quantity computed downstream, because almost all of them are ratios against an extracted price.

The literature on textual analysis in accounting and finance treats extraction quality as a measurable property and has documented at length that naive lexical methods transfer poorly to filing prose

(Loughran and McDonald, 2011, 2016). The engineering literature makes an adjacent observation about systems rather than about texts: a statistical component embedded in a larger pipeline degrades in ways that produce output rather than exceptions, and the surrounding system has no means of noticing (Sculley *et al.*, 2015). This paper reports what that combination looks like in a running system, using as evidence the instrumented failures of one such system over a period in which its extracted terms were displayed to an operator and acted upon.

The five failure modes catalogued below share a property that determines how they must be handled. None of them raises an exception, returns a null, or writes a diagnostic. Each returns a value of the expected type whose magnitude is plausible for the field it occupies: a price in dollars, a ratio near unity, a

company name with a corporate suffix. One of them, the document-selection failure of §4.1, wrote the absence of any consideration language into the record while setting the human-review flag to false, which is to say that it recorded absence of evidence as a clean read. A defect with that signature cannot be found by monitoring for errors, and it cannot be found by a plausibility check on the output field alone, because plausibility is exactly what it preserves.

Contribution. A taxonomy of five silent failure modes in merger-terms extraction, each documented against the specific filing that exhibited it and the change that closed it, with byte and character counts, occurrence counts and blind-read scores quoted from the programme’s own instrument record. A worked cascade in which four of the five compose on a single record. Four detection practices stated as reusable procedure, together with the measured observation that they do not nest, so coverage is a property of the assembly and not of any one check. A blind-read certification record covering three readings and four repair rounds against a bar frozen before the first reading.

Three restrictions belong with the reader before any number. The evidence is an instrumented case series from one system, not a sample from a population of extractors, so nothing here supports a rate at which any of these modes occurs in extraction generally. Several counts below are census counts over a small live board or a small document cache and carry no interval. The blind-read scores of §7 are the only figures in the paper measured against a bar frozen in advance of the measurement, and they certify the tender-offer tier of one extractor on fifteen filing families; they do not certify the merger tier, whose validation took a different and weaker form described in the same section.

2. Relation to the availability boundary, and to prior work

A companion paper in this series catalogues the boundary at which free filing data ceases to support a research question: populations too small to score, prices unavailable for delisted securities, control pools whose composition differs from the treated arm,

coverage differentials between outcome arms. The units described there stopped. Each was sealed by a validity gate before its statistics were read, and the verdicts recorded are measurement-limited rather than null.

The subject of this paper is the complementary case, and the distinction is worth stating precisely because the two are easy to conflate under the single heading of data quality. Here the document was fetched, the text was in hand, the extractor ran to completion, and the value it returned was wrong, without halting the pipeline at any point. The failures reached a decision surface, were ranked, were sized, and in one instance were acted on. Availability limits what can be attempted; extraction correctness determines whether what was attempted means anything, and the two are detected by entirely different machinery. A coverage gate counts how many episodes yielded a value and cannot tell whether the values are right. The gate described in §6 asks whether a value is corroborated by the document it claims to come from and cannot tell whether coverage is adequate.

None of the individual detection techniques assembled in §6 is novel, and the paper does not present them as such. Resolving an expression such as “Parent” or “the Company” to the entity it denotes is the anaphora-resolution problem, treated computationally since Hobbs (1978), given a widely implemented syntactic and salience-based algorithm by Lappin and Leass (1994), and placed on a discourse-theoretic footing by the centering framework of Grosz, Joshi and Weinstein (1995); the supervised coreference literature that followed is surveyed by Ng (2010). Deciding whether a name in a document denotes a particular registrant is the entity-resolution or record linkage problem, formalised by Fellegi and Sunter (1969), surveyed for databases by Elmagarmid, Ipeirotis and Verykios (2007), and extended to the relational setting by Bhattacharya and Getoor (2007). Merger agreements are in one respect an unusually tractable instance of the first problem, because the document supplies its own antecedents: a recital names each party once and then defines a term for it, so the anaphors that follow are resolvable from an explicit definition list rather than from salience heuristics.

They are an unusually adversarial instance of the second, because the placeholders are designed to look like company names and, as §4.2 shows, one of them resolves cleanly to a real listed issuer.

What is reported as contribution is the assembly: these techniques composed into a gate that sits over primary filings, that refuses rather than annotates, that is certified by blind reading against a bar frozen in advance, and whose refusals are enforced at the point where a number would otherwise become an instruction. The failure taxonomy that motivates the assembly, and the observation that its components do not subsume one another, are the empirical part.

3. The system under study and the observation regime

The system is a daily research desk that reads corporate-event filings from the Commission’s public endpoints, extracts transaction terms, computes a spread and a probability of completion for each open transaction, and publishes a ranked board of cards to an operator. Two extractor tiers are relevant. A tender tier reads Schedule TO complexes and the corresponding target responses, emitting an offer price, an expiration, a minimum-tender condition and the bidder identity. A merger tier reads proxy statements and their supplements, emitting a consideration type, a cash amount per share or an exchange ratio, a special-meeting date, an outside date and a termination fee. A third layer, added later, reads the parties themselves out of the merger agreement’s recital. Every emitted field carries the verbatim sentence it was read from and the accession that supplied it.

The observation regime matters for how the evidence below should be weighed. These are not failures found by seeding synthetic errors into a test corpus. They were found in production, by four distinct routes, and the route is recorded with each. Two were found by an operator reading a card and disagreeing with it. One was found by an alert engine comparing a displayed

cash face against the evidence recorded for it. One was found by a coherence checker that audits the published board against invariants. One was found only by writing a purpose-built read-only sweep and running it over every card on the live board. The last of these is the important case, because it establishes that the first four routes were insufficient: the defect it found had been on the board for five days, under a header that named it a defect, while the card continued to issue instructions.

Terminology used throughout. A *card* is one published record for one transaction, carrying extracted terms, a computed probability, a rank and, when cleared, a position size and an entry instruction. A *gate* is a predicate that refuses to publish a value rather than flagging it. A *refusal* is machine-readable, carries the name of the gate that issued it, and is displayed. A *blind read* is a comparison of extractor output against an independent reader who did not see that output, scored by a comparator that fails closed on disagreement.

4. A taxonomy of silent extraction failures

4.1 Selection of the wrong document within the correct submission

An accession on the Commission’s archive is a directory of files rather than a single document. A definitive or preliminary merger proxy arrives alongside a filing-fee exhibit, a set of viewer renderings produced by the inline XBRL toolchain, an index page, and assorted graphics. The reader in this system resolved an accession to a document by walking the directory listing in the order the archive returns it, skipping the two index files, and taking the first file with an `.htm` extension.

That rule returns the wrong file whenever the fee exhibit or a viewer rendering happens to precede the proxy in the listing, which is common. On three transactions held by the operator it returned the following, with sizes as the archive itself reports them, captured on 17 August 2026.

TRANSACTION	DOCUMENT SELECTED	BYTES	PRIMARY DOCUMENT	BYTES
Arcosa	filing-fee exhibit	17,774	preliminary merger proxy	1,983,695
Caesars Entertainment	filing-fee exhibit	12,232	preliminary merger proxy	1,593,403
LiveRamp	XBRL rendering, R1.htm	10,831	preliminary merger proxy	4,269,105

The three selected files flatten to between 2.7 thousand and 17.8 thousand characters of text; the corresponding primary documents flatten to between 799 thousand and 1,315 thousand. None of the three selected files contains merger-consideration language, because none of them is about the consideration. The extractor searched them, found nothing, recorded nothing, and set the human-review flag to false, since no inconsistency had been encountered and no parse had failed. All three transactions are genuine cash mergers, and the operator subsequently verified each price by hand against the primary document. The record therefore asserted the absence of consideration evidence for three transactions whose consideration is stated plainly in a document the system had already fetched.

Detection. A search that returns nothing, over a corpus that was never the right corpus, is indistinguishable from a search that returns nothing over the right corpus. The extractor cannot tell them apart, and the field it writes is identical in both cases. The failure is detectable only by a predicate on the document rather than on the extraction, which is why the correction below is an adequacy floor and not a better search.

The repair is a selector with four ordered rules. Files are rejected by kind before any size is consulted, so that fee exhibits, viewer renderings, stylesheet stubs, index pages, graphics and linkbases are excluded whatever they weigh. Among survivors, the archive's own declaration of the primary document governs, subject to the same kind rejection, because that declaration is itself junk for some form types. Size acts only as a tiebreak among survivors the archive did not label, and never as the primary rule, since size alone selects a 225,249-byte viewer rendering over a short but authoritative current report. A survivor below a minimum size is not treated as a filing, and when nothing qualifies the selector refuses; the caller then reads the full submission text, which contains every

document in the accession and so cannot be the wrong one. A refusal degrades to more text rather than to less.

The same defect was subsequently promoted from a repair into a blocking gate in a registered study of realised consideration, where the threshold was derived from measurement rather than chosen. The three documented failure documents flatten to between 2.7 thousand and 17,774 characters. A cache of 120 genuine merger proxies has a minimum flattened length of 119,539 characters, a tenth percentile of 184,213 and a median of 288,627. A floor of 50,000 characters sits a factor of 2.8 above the largest known bad document and a factor of 2.4 below the smallest known good one, and rejects none of the 120. In that study a document below the floor is recorded as a miss with a stated reason and never as a clean read.

4.2 Resolution of a structural placeholder to a real company

Transaction documents are populated with entities that exist only to carry a role. A newly incorporated holding company is referred to throughout as New PubCo; an acquisition vehicle appears as Merger Sub; the acquiring entity is Parent and the acquired entity the Company. These are not names in the referential sense, and the document says so: the filing at issue introduces New PubCo as a newly formed company.

The system resolved a counterparty name to a ticker by lowercasing it, taking the first word, and returning the shortest registrant title beginning with that word. Applied to New PubCo this takes the first word "new" and returns News Corp, whose ticker is NWSA. The resolution was reproduced exactly against the live registrant map of 10,387 entries. The published card then named News Corp as the counterparty to a transaction in which News Corp has no part whatsoever.

Two properties of this failure govern how it must be detected. The first is that the output is well formed: a real company, a real ticker, a live quote, a plausible corporate name. The second is that the resolver did not require a placeholder to misbehave. Applied to “Vertex Pharmaceuticals Incorporated”, a complete, unambiguous and correct name, the same shortest-title rule returned VERX, a tax-software company. The shortest-title rule was a tiebreak promoted into a matcher, and it was deleted rather than tuned, replaced by four ordered tiers in which a unique match is required and ambiguity returns nothing. Multiple share classes filed under one registrant title are treated as one entity rather than as competing candidates, an adjustment made after the first build of the replacement refused a correct resolution over that ambiguity.

4.3 Selection of the wrong clause within the correct sentence

Consideration language in a merger proxy is enumerated within a single sentence. Clause (i) states what each share of common stock converts into; clause (ii) states what each share of a preferred series converts into; further clauses cover options, restricted units and warrants. The clauses are separated by semicolons, not by full stops.

The exchange-ratio extractor anchored on the phrase “right to receive” and scanned forward through a character class admitting any character except a full stop, to a budget of 160 characters. It therefore crossed the semicolon that ends the common-stock clause and came to rest inside the preferred clause, returning 3.676, the ratio at which a Series C preferred share converts into common. The operative common-stock ratio in the same sentence is one for one. Both numbers are ratios, both are near-plausible for the field, and the extracted one is in the document, quotable, and attributable to the correct accession and the correct sentence. It answers a different question from the one asked.

A second site in the same file, the extractor for the acquirer name, carried the identical exposure at a 200-character budget, and no prior review had identified it. The repair at both sites extends the terminating

character class to stop at semicolons and colons as well as full stops. The more consequential change is the one described in §6.3: an extracted number now carries the clause it was read from as a typed field, so that a ratio governing preferred conversion is distinguishable from a ratio governing merger consideration without a reader having to reconstruct the sentence.

4.4 Inversion of acquirer and target

The fourth mode differs from the first three in that no value is malformed and no lookup misfires. The extractor identifies two companies, both real, both named in the filing, and assigns them to the wrong roles. A card built on an acquirer is then priced, ranked and instructed as though it were a target, which is the opposite side of the transaction.

The instance that made this visible published a buy instruction at a limit of \$135.24 with a payoff quoted as 0.3050 times the price of the counterparty’s stock. The issuer is the acquirer; the counterparty is the target; the number quoted as the payoff is what the issuer pays, not what it receives. The defect is not detectable by any corroboration of the counterparty against the document, because the counterparty is genuine and occurs 35 times in the text. Name-based checks are structurally blind here, and this is the observation that forces §8’s conclusion about non-nesting.

Direction is recoverable from the document, but only from its definitional apparatus rather than from its narrative. A merger agreement’s recital names the acquiring party once and thereafter refers to it by a defined term, so a probe that searches for company names in sentences finds the acquirer at most once and often not at all, while the anaphor that carries the role appears throughout. The repair reads the filing’s own defined terms first, resolves anaphors through them, cross-checks the result against a sentence-level tier so that a contradiction fails closed, and refuses when an anaphor is never defined. Consulting definitions before sentence probes is load-bearing rather than stylistic: on one live proxy, an intra-group sentence about the issuer’s own corporate tree caused the

sentence-first path to report a role conflict on a transaction the operator held and whose direction was correct.

Three further defects in name-probing were exposed by real filings during that work, each closed with a failing control. A comma-delimited list of parties bound a defined term to the preceding party as well as the intended one. A defined term was bound to a jurisdiction rather than to an entity, giving Cayman Islands and State of Israel as counterparty names on two separate filings. A shared industry word placed a party on a filing in which two energy companies both matched it. None of the three produces a malformed value.

Measured incidence. A role audit run over the cache of merger filings on 17 August 2026 read 22 cards. Five resolved as role-inverted, on transactions in which the card's own issuer is the acquiring party, published under the symbols AWK, RCLB, EQR, PASG and PUSA. One further card resolved as unverified, the filing placing its issuer on both sides without its defined terms settling which is the deal role. The remaining 16 resolved correctly, target-side. Two of the five inversions had already been refused on other grounds and three had been refused by a name gate that returns before role is examined, so their inversion had never been reached, recorded or displayed. This is a census over one cache on one day and supports no rate.

4.5 Refusal without consequence

The fifth mode concerns what follows a correctly diagnosed extraction failure rather than the extraction itself: the failure of that diagnosis to prevent anything downstream of it, which is the mode that determines whether the other four cost money.

The card discussed in §5 was marked with a distress flag, a review hold and a header reading DATA DEFECT, DO NOT ENTER. It simultaneously published a position size of \$596 and 129 shares, a rank score of 552.57 placing it first on a board of 43

cards, and a plan step reading ENTER: buy at limit \$4.61. It did so for five days, and the operator acted on it.

The mechanism is a deny-list. Three earlier rounds of work on the same file had addressed this class of defect by enumerating markers or by reordering operations. One wrote a row-level gate correctly and applied it to the whole-book plan alone; another repaired the card report for held cards alone; a third cleared two of nine instruction fields. Each closure was correct for the cases it enumerated and left every unenumerated route open, which is the defining property of a deny-list: it fails open for each card kind and each marker introduced after it was written.

The replacement inverts the default. A card must acquire an affirmative clearance token before any producer emits a size, a rank, an entry price or an entry step. Absence of a token is the resting state, so a newly introduced card kind carries none until one is minted for it, a newly introduced flag denies by default until reviewed, and an exception raised while evaluating clearance yields none. The predicate is enforced at the callee rather than at each call site, so that routes nobody enumerated are covered, and a serialisation assertion raises if a refused card reaches output carrying an instruction. Refused cards receive a zero-valued refusal receipt rather than a null, since a null already means that the sizer never ran, and they are ranked at the stand-aside sentinel so that they sort down the board's own ordering rather than riding a high certainty floor to the top.

5. Composition: one card, four modes, five gates

The modes of §4 are presented separately because they have different detectors, not because they occur separately. One card carried four of them in series, and its arithmetic is worth following because each step is individually defensible and the terminus is wrong by a factor of twenty-three.

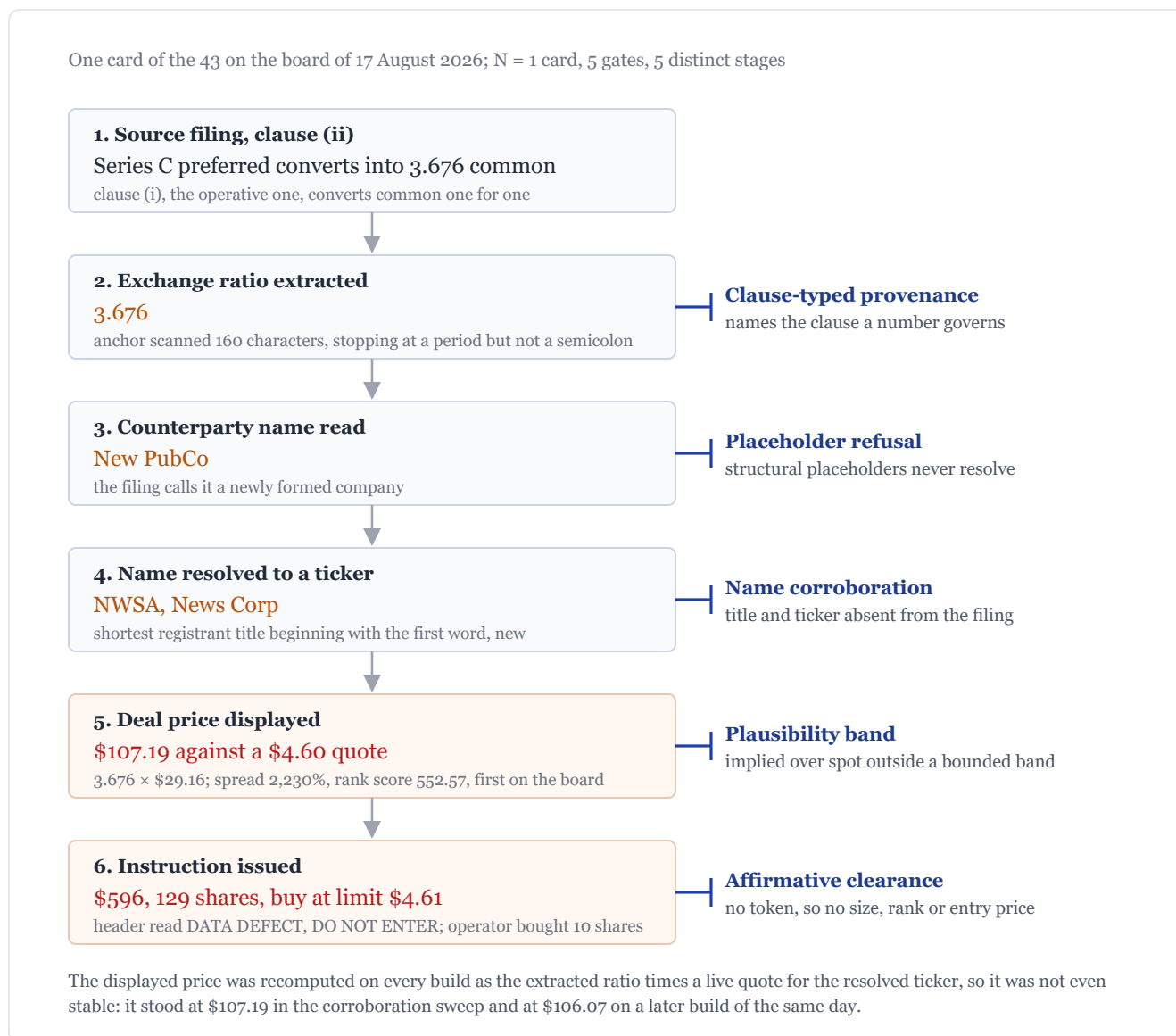


Figure 1. The cascade on a single board card, with the five gates that now refuse it. Reading down: the extractor crossed a semicolon out of the operative common-stock clause into a preferred-conversion clause and took 3.676 as the exchange ratio; it read the counterparty as the structural placeholder New PubCo; it resolved that placeholder to NWSA by taking the first word and returning the shortest registrant title beginning with it; it multiplied 3.676 by that company's \$29.16 quote and displayed \$107.19 as the deal price against a \$4.60 quote, a spread of 2,230% and a rank score of 552.57 that placed the card first on a board of 43; and it issued a sized instruction of \$596 and 129 shares at a limit of \$4.61 beneath its own header reading DATA DEFECT, DO NOT ENTER. The operator bought 10 shares. The conclusion the reader should draw is in the placement of the gates rather than in any one value: five gates refuse at five distinct stages, no gate acts on more than one, and no gate can restore a value corrupted upstream of it, so coverage is a property of the assembly. The instability of the displayed price is a separate diagnostic. Because it was recomputed on every build as the extracted ratio times a live quote for the resolved ticker, it stood at \$107.19 in the corroboration sweep and \$106.07 on a later build of the same day, which is a value no reading of the filing could produce and which no check on the filing would have caught. N = 1 card. Values are quoted from the corroboration sweep's own record and from the commit that repaired the cascade.

Each link of the cascade is wrong in a way that is correct at its own level of description, and that property is what makes the sequence instructive. The ratio 3.676 is in the document. New PubCo is the name the document uses. News Corp is a company

whose title begins with the first word of that name. The product of a ratio and a share price is a per-share value. The spread between that value and a quote is a spread. No step commits an error of arithmetic or of type, and no step has access to the information that would show it to be wrong, because that information

lives at a different level: in the clause structure of the sentence for the first, in the document's own characterisation of the entity for the second, in the entire text for the third, in the distribution of spreads across the board for the fourth, and in the card's own flags for the fifth.

The decisive test for the third link costs one request. The strings "News Corp", "News Corporation" and "NWSA" occur zero times in the filing from which the terms were read. A counterparty whose own name never appears in its own transaction document was not read out of that document.

6. Detection practices

The four practices below are stated as procedure. Each is implementable against primary filings alone, each refuses rather than annotates, and each has a failing control in the accompanying test suite, meaning a fixture that reproduces the pre-repair behaviour and on which the check must fail.

6.1 Corroboration of a resolved entity against its own source

When an extractor resolves a name in a filing to an external identifier, require that the identifier's own official title and symbol occur in the source document. Zero occurrences refuses the resolution rather than annotating it. Two refinements were forced by practice. A single generic token cannot corroborate, so the ordinary English word "news" does not clear News Corp, and the rule is stated as a requirement on non-generic tokens rather than as a phrase list. Binary payloads must be stripped before counting, because encoded exhibit blobs generate spurious matches; in one run the audit instrument itself scored three standalone matches for a three-letter ticker out of uuencoded content.

The clause intended to remove those payloads is itself worth recording, because it illustrates the paper's theme applied to a detector. The pattern carried two literal backspace bytes around its terminator, so it compiled to an expression that matched nothing in real filing text and had been inert for its entire existence. What had been holding the line was an unrelated

backstop that treats any run of 45 or more non-space characters as non-prose, which happens to catch uuencoded body lines because they are 60 characters or more. The clause now has its own failing control, a block whose body lines are 21 characters and which the backstop provably cannot catch; replanting the two bytes turns three checks red and removing them turns them green. A check that reads as covering a case and covers nothing is the same defect as a summary computed after the information it summarises has been destroyed.

6.2 Role direction from the definitional recital

Do not infer transaction direction from which company names appear in the text. Read the recital's defined terms, resolve the role-bearing anaphors through them, and cross-check the result against a sentence-level reading so that a disagreement fails closed. Where an anaphor is never defined, report the role as unresolved and refuse on that basis rather than treating an unresolvable role as a passed test. The distinction is operative: in the system studied here, an unresolved role was initially a hold, which left the card on the board with its numbers intact, and was converted to a named refusal that denies clearance.

The cost of the weaker form is measurable. Before the definitional tier was added, five cards could not be cleared at all, because the name-only probe drops every anaphoric capture, and the acquirer in a United States merger proxy is named once in the recital and referred to as Parent thereafter. Of 16 cards examined at that stage, 5 carried an acquirer named only by that anaphor. After the definitional tier, all five previously uncleared cards resolved, none of them inverted, including two the operator held.

6.3 Clause-typed provenance

Record, for every extracted number, the clause it was read from and what that clause governs, and publish the classification alongside the value. A ratio governing the conversion of a preferred series is then distinguishable from a ratio governing merger consideration without a reader reconstructing the sentence. Record also whether the anchor pattern

crossed a semicolon or a colon to reach the number, since that crossing is the mechanism by which a correct sentence yields a wrong clause.

An ordering constraint accompanies this practice and was discovered by failing to observe it. When gates return at the first failure, an extractor downstream of the failing gate never runs, so at the moment a record is refused there is nothing on it to withdraw and nothing to explain the refusal with. On the card of §5 this produced a correct refusal that could not say where the fabricated price had come from. The correction separates extraction from adjudication: every extractor runs first, each isolated so that one raising an exception cannot suppress the others, and each records its finding together with the clause or lookup that produced it; the gates then adjudicate over the completed record. A gate added later inherits the behaviour without per-gate bookkeeping, and a refusal can state what it is refusing.

6.4 Affirmative clearance over deny-listing

Where a refusal must prevent an action, express it as a token the record must earn rather than as a list of markers that block. The argument is the one given in §4.5 and is not specific to filings: a deny-list fails open for every case introduced after it is written, which in a system that gains card kinds and flags over time is a guarantee of eventual failure rather than a risk of one. Enforce the predicate at the callee so that unenumerated call routes are covered, and assert at serialisation so that a refused record carrying an instruction raises rather than publishing.

7. Certification by blind read against a bar frozen in advance

The practices of §6 detect specific defects. Certifying that an extractor is fit for use at all requires a different instrument, and the one used here is a blind read: a sample of filing families is read by an independent reader who does not see the extractor’s output, and a comparator that fails closed on disagreement scores agreement family by family. The bar was set at 13 of 15 families and frozen before the first read.

READ	EXTRACTOR VERSION	SCORE	OUTCOME	DOMINANT DEFECT CLASS
First	1.1.0	2 of 15	Fail; fills quarantined	Episode segmentation, 7 of 15 families; 12 of the 15 sampled were quarterly repurchase chains
Second	1.2.0	7 of 15	Fail	Bidder anaphors, 7 of 8 failures
Third	1.2.1	14 of 15	Pass; quarantine lifted	One residual, an extension stated in the sentence after the quoted schedule

The first read is the informative one. It failed at 2 of 15, and the record states that the bar is unmet under any of the sensitivity readings considered, which placed the score at 5 or 11 of 15 under the most generous. The failures concentrated in a single population. Interval funds file quarterly repurchase offers under one file number indefinitely, at intervals of roughly 91 days, shorter than the 120-day fence used to separate one offer complex from another, so a multi-year chain was read as a single offer and the extracted expiration landed on a superseded quarter. Within the same population, net asset value dates were

read as expirations, and the sentence “the Fund commenced operations” was read as the commencement of the offer, wrong in 7 of 8 claims. One negotiated transaction was labelled unsolicited from fiduciary-out boilerplate, in which the word “unsolicited” modifies a hypothetical competing proposal rather than the offer in hand.

Two consequences of that read are worth separating. The first is what was repaired: episode segmentation on repeated base forms, exclusion of valuation dates from the expiration walk, deletion of the commencement-date field outright on the ground that

the anchor space is not recoverable in that population and no consumer required the field, validation of bidder names against sentence-spanning captures, and a requirement that “unsolicited” modify the offer itself. The second is what happened in the interim: extractor fills were restricted to third-party tender complexes, the unsolicited field was barred from fills entirely, and no extracted field was permitted to feed any statistic. A failed certification suspended use rather than annotating it, which is the same discipline as §6.4 applied at the level of the instrument.

One positive finding from the first read is recorded because it bounds the failure class. Across all 15 families, no price was hallucinated. Every incorrect value was a real number read from the document under a wrong reading of what the document was saying, which is the property the entire taxonomy of §4 shares and which determines that the detectors must be structural rather than statistical.

What this certification does not cover. The 14 of 15 is the tender tier. The merger tier was validated by a different and weaker procedure: an adversarial review in which twelve findings were closed in one round, further findings in a second, and three remaining items in a third, each with fixtures drawn from the named source filing. No numbered blind read of the merger tier against a frozen bar appears in the programme’s record, and none is claimed here. The merger tier is the tier that produced the failures of §§4.1, 4.3 and 4.4, which is consistent with, though not evidence for, the weaker validation.

8. Coverage, and why the practices do not nest

Each detection practice refuses a different mode, and the mapping is close to one to one. That is the operationally important property, because it means the assembly cannot be replaced by strengthening any one member.

FAILURE MODE	PRACTICE THAT REFUSES IT	PRACTICES STRUCTURALLY BLIND TO IT
Wrong document in the right submission	Document adequacy floor and kind rejection	All three others: the extraction is empty, so there is nothing to corroborate, no clause to type and no role to read
Placeholder resolved to a real company	Name corroboration; placeholder refusal	Clause typing; role direction; adequacy
Wrong clause in the right sentence	Clause-typed provenance	Name corroboration, which the correct counterparty passes; adequacy
Acquirer priced as target	Role direction from the recital	Name corroboration, by construction; clause typing; adequacy
Refusal that still instructs	Affirmative clearance	All extraction-side practices, which have already fired correctly

The role-inversion row is the one that settles the question, and it does so by construction rather than by observation. Name corroboration asks whether the resolved counterparty appears in the source document. In an inverted card the counterparty is the genuine other party to a genuine transaction and appears throughout the text, 35 times in the instance examined. The check passes, correctly, on a card whose direction is wrong. No tightening of the corroboration threshold reaches this case, because the quantity it measures is not related to the quantity that is wrong.

The first row makes the complementary point from the opposite direction. When the wrong document has been selected the extraction is empty, so corroboration has nothing to corroborate, clause typing has no clause, and role reading has no recital. A gate over extractions is silent on a defect in the choice of what to extract from, which is why that defect required a predicate on the document rather than on the output. A related asymmetry appeared when the programme audited its own board checker. Of 110 reported coherence violations, 57 were the checker failing correct behaviour rather than the board failing: one invariant derived a hard refusal by pattern-matching

prose that contained the clearance mechanism's own gated verb, and read 14 ordinary watch cards as hard refusals; another tested for a prefix that a subsequent change had legitimately moved; a third fired on cards that publish no probability at all. A further 9 were the gates working as designed, on cards carrying a real structural defect in their filing, every one refused, attributed to a named gate, ranked at the stand-aside sentinel and funding nothing. Those 9 were reclassified as notices and removed from the violation count, on the reasoning that a count which can only rise trains its reader to ignore it. The residual 53 are genuine desk defects. One sub-check within that suite was found to have had no coverage at all for its entire life, its pattern requiring quote delimiters that a refactor had removed from the text it scans, scoring one accidental match in its lifetime.

9. Limitations

This is a case series from a single system, and the failure modes are reported with the evidence that identified each one rather than with a frequency. Nothing here estimates how often any mode occurs in extraction from filings generally, and the counts that do appear are censuses over small populations on single days: 22 cached filings in the role audit, 43 cards on the board that carried the cascade, 15 filing families in each blind read, 120 proxies in the document-length reference set.

The blind-read scores are the only measurements in the paper taken against a bar frozen before the measurement, and their scope is stated in §7. Everything else is a diagnosis after the fact, with the attendant hazard that a diagnosis constructed after observing an outcome tends to fit it. The mitigation applied throughout the underlying work is that each repair ships with a failing control, a fixture reproducing the pre-repair behaviour on which the new check must fail, so that a repair whose mechanism is misdiagnosed is detectable as one whose control does not discriminate. The uuencode clause of §6.1 is the instance where that mitigation found a misdiagnosis, and it found it in a detector rather than in the system under test.

The gate described here has not been validated forward. It was constructed against the failures it was built to catch, which is the weakest possible evidential position for a detector, and the honest statement of its status is that it refuses the five documented modes and has not been shown to refuse a mode it was not designed for. The audit of §4.4 is the nearest thing to an out-of-sample reading, in that it applied the role layer to a cache of filings rather than to the cards that motivated it and found four inversions beyond the one already known; that is a small and favourable result on a population selected by the desk's own screen, and it should not be read as a validation.

Two further boundaries are inherited from the underlying work. All of this concerns terms as stated in filings, not consideration as received, and the difference is not negligible: two live transactions in the sample carry contractual adjustments running in opposite directions, a per-share ticking fee in one and a cash-shortfall reduction in the other, while both cards display a flat face. Whether the announced consideration is the payoff is a separate question, registered as a distinct study and not yet run. Second, a related registered study found that the Commission's machine-readable filing metadata names a counterparty only for two-sided filings, so that a cash one-step merger leaves the acquirer in prose; that study sealed at 33.7% coverage against a 40% floor, and its surviving 301 episodes were 89.7% stock deals, which is precisely the selection its coverage gate existed to prevent it from reporting. The role-extraction layer of §6.2 was built for the population that study could not reach, and its adequacy for research use, as opposed to display use, is not established.

10. Reproducibility

Every quantity in this paper is drawn from one of four sources, and the source is identified at the point of use in the generating module. Byte sizes for the document-selection failure are the archive's own, captured from the accession listings on 17 August 2026 and held as fixtures in the selector's test suite, so that the naive picker which caused the incident is retained as a

failing control and is shown to return those three files exactly. Character-length figures, the adequacy floor and the reference distribution over 120 proxies are quoted from the registration record of the consideration-realisation study, which froze them before any document was fetched. The cascade values are quoted from the read-only sweep written to diagnose it, whose module record states each link with its provenance. The blind-read scores, the gate inventory and the incidence counts are quoted from the commits that recorded them, each of which is dated and identifies the filings involved.

The programs are runnable offline against cached filings. Three test suites were executed in the working tree on 17 August 2026 while this paper was written: the desk builder's suite at 587 checks passing and none failing, the board coherence checker at 153 passing and none failing, and the corroboration sweep at 63 checks passing and none failing. The role audit reported in §4.4 was run in the same session over the sweep's own document cache and returned 22 cards read and 6 not resolving to a clean target-side role. The convention observed throughout is that a check ships with a fixture on which it fails, since a check that has never been observed to fail carries no information about the code it protects; the same convention identified the two inert checks reported in §§6.1 and 8.

The specific program, cache and commit identifiers behind each figure are recorded in the platform repository and in the generating module's own header, which names the file and line range or the commit for every number reproduced above.

11. Discussion

The recurring property of these failures is that they are correct at their own level of description and wrong at the level that matters. A regular expression that stops at full stops does stop at full stops. A resolver that returns the shortest registrant title beginning with a given word does return that title. A search that finds no consideration language in a filing-fee exhibit has searched the exhibit correctly. The error in each case is a mismatch between what the component was asked

and what the caller needed to know, and no amount of care applied within the component addresses it, because the component has no access to the information that would reveal the mismatch.

That diagnosis has a practical consequence for how such systems should be checked. Verification applied to the output field is ineffective against every mode described here, since each mode preserves type and plausibility; the failures were found only by predicates that reach outside the component, comparing an extraction against the document it claims to come from, a role against the definitions that establish it, a number against the clause that governs it, and a document against the population of documents of its kind. Each of these is a corroboration against an independent source rather than a consistency check within one.

The blind-read record supports the same reading from a different angle. Fifteen filing families, read by someone who did not see the extractor's answers, found in one pass a defect class affecting seven of them, within the twelve that were interval-fund repurchase chains, which no amount of internal testing had surfaced, because the extractor's treatment of quarterly repurchase chains was internally consistent and simply did not correspond to what those filings are. The cost of that instrument is one afternoon of reading; the cost of not having it was measured in this system as a set of extracted fields that had been feeding a live board.

What remains unestablished is the more interesting question. This paper reports that a fail-closed assembly refuses five documented failure modes and that its components do not subsume one another. It does not report the rate at which such modes arise, the fraction of an extraction corpus they affect, or whether the assembly refuses modes it was not built against. The registered consideration study of §9 will supply the first measurement bearing on any of these, since its document-adequacy gate is precisely the first failure mode promoted to a blocking condition and its coverage and symmetry gates will report how much of a 285-episode panel the extractor can actually read. Until such a measurement exists, the position defended here is narrow: these five modes are real,

they are silent, they compose, and the cheapest of the checks that refuses them is a single request asking whether a company named on a card appears in the document the card was built from.

References

- [1] I. Bhattacharya & L. Getoor. Collective entity resolution in relational data. *ACM Transactions on Knowledge Discovery from Data*, 1(1):5, 2007.
- [2] A.K. Elmagarmid, P.G. Ipeirotis & V.S. Verykios. Duplicate record detection: a survey. *IEEE Transactions on Knowledge and Data Engineering*, 19(1):1–16, 2007.
- [3] I.P. Fellegi & A.B. Sunter. A theory for record linkage. *Journal of the American Statistical Association*, 64(328):1183–1210, 1969.
- [4] B.J. Grosz, A.K. Joshi & S. Weinstein. Centering: a framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225, 1995.
- [5] C.R. Harvey, Y. Liu & H. Zhu. ...and the cross-section of expected returns. *Review of Financial Studies*, 29(1):5–68, 2016. The source of the programme’s convention that a bar is frozen before the measurement that tests it.
- [6] J.R. Hobbs. Resolving pronoun references. *Lingua*, 44(4):311–338, 1978.
- [7] S. Lappin & H.J. Leass. An algorithm for pronominal anaphora resolution. *Computational Linguistics*, 20(4):535–561, 1994.
- [8] T. Loughran & B. McDonald. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance*, 66(1):35–65, 2011.
- [9] T. Loughran & B. McDonald. Textual analysis in accounting and finance: a survey. *Journal of Accounting Research*, 54(4):1187–1230, 2016.
- [10] V. Ng. Supervised noun phrase coreference research: the first fifteen years. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1396–1411, 2010.
- [11] D. Sculley, G. Holt, D. Golovin, E. Davydov, T. Phillips, D. Ebner, V. Chaudhary, M. Young, J.-F. Crespo & D. Dennison. Hidden technical debt in machine learning systems. In *Advances in Neural Information Processing Systems* 28, 2015.
- [12] M.A. Andreev & J. Howden. The reachable boundary of free filing data for event-driven research: a catalogue of measurement limits with named defect classes. *Quark Quantitative Research Working Paper*, 2026. Companion paper; treats data availability, where this paper treats extraction correctness on data already held.
- [13] M.A. Andreev & J. Howden. A measured ladder of pre-announcement and in-process deal probabilities, reconstructed entirely from free EDGAR filings. *Quark Quantitative Research Working Paper*, 2026. Supplies the completion and stage estimates the cards discussed here display alongside their extracted terms.
- [14] Instrument record for the extractor, the corroboration sweep, the board coherence checker and the clearance gate, together with the dated commits that record each blind read, each gate and each incidence count quoted above, held in the platform repository.

© 2026 Maksim Aleksandrovich Andreev & Joseph Howden.
All rights reserved.

Quark Quantitative Research Platform — Working Paper
Generated 2026-08-17 21:46 UTC

This is a methods paper reporting instrumented failures of one system. It states no return, no edge and no forward-validated result. The counts reported are censuses over small populations on stated dates and carry no intervals; the blind-read scores are the only figures measured against a bar frozen in advance.

Correspondence: M.A. Andreev, Quark Quantitative Research.