

# **The First Annual Reports of the LLM Era**

*Language Change in 150 SEC 10-K Filings, 2019–2024*

**Richard Foster-Fletcher**

MKAI

Correspondence: [research@mkai.org](mailto:research@mkai.org)

March 2026

*Working paper*

## Abstract

We analysed 150 SEC 10-K filings across 50 of the largest US public companies, comparing FY2019, FY2022, and FY2024. Six dimensions of language change were scored on a 0 to 2 scale. Underlying extracts, quotations, word counts, section attributions, and absence claims were mechanically verified against source filings. Scores were then assigned using the stated rubric.

Across the five editorial dimensions, excluding the structurally affected new-section dimension, the annual rate of prose drift was 24.5 per cent higher in the later interval (2022 to 2024) than in the earlier interval (2019 to 2022). On a per-company basis, the average annual drift rate rose from 0.71 points to 0.88.

The composition of drift also changed, though not as a simple switch from one dominant form to another. In the earlier interval, sentence inflation was the dominant form, accounting for 29.2 per cent of editorial drift. In the later interval, no single form dominated. Hedge proliferation and sentence inflation were joint largest at 25.0 per cent each, with framing drift close behind. The later period is therefore better described as more diffuse than as dominated by any one pattern.

This research is descriptive. It documents observable changes in filing language and their timing. It does not attribute cause.

## What prose drift looks like

FY2024 is plausibly the first filing cycle in this dataset drafted largely after enterprise-grade large language model tools became widely accessible. We wanted to know whether the language in those filings changed, whether the change was visible in the text itself, and whether it differed from the changes already under way before that point.

We compared 150 SEC 10-K filings across 50 of the largest US public companies, using FY2019, FY2022, and FY2024 as observation points. For each company, we scored six dimensions of language change on a 0 to 2 scale: whether testable identity claims were softened, whether hedge words increased, whether generic framing became more separated from concrete examples, whether named specifics were replaced by categories, whether comparable passages grew longer without saying more, and whether newly added sections were more generic than legacy prose. A score of 0 means the language held steady or became more specific. A score of 2 means it became markedly less specific, more hedged, or more generic. The maximum total is 12.

Scores across the 50 companies range from 1 to 10. The mean is 4.8. The median is 5. The more interesting finding is not drift alone. It is that the pace of drift rose in the later interval, while the composition of drift became less concentrated in any one form.

This study takes a deliberately narrow measurement approach. Rather than applying automated text analysis across thousands of filings, the dataset is limited to 150 filings across 50 companies and relies on structured human scoring applied to

mechanically verified text extracts. Each passage comparison is tied to specific sections of the underlying filings, and every quotation, word count, section attribution, and absence claim was verified against the source documents before scoring decisions were made. The resulting dataset is therefore smaller than typical large-scale disclosure studies, but the measurement process is correspondingly more transparent. The objective is not statistical generalisation but disciplined observation of how filing language changed across the first reporting cycle after enterprise-grade LLM writing tools became widely accessible.

## **The interval question**

Because the dataset contains three filing years rather than two, the analysis can be split into two periods. Period A runs from 2019 to 2022. Period B runs from 2022 to 2024.

The periods are different lengths, so raw totals are less informative than annual rates. On the five editorial dimensions, Period A contains 106 points of drift and Period B contains 88. Annualised, that is 35.3 aggregate points per year in Period A and 44.0 in Period B. On a per-company basis, the average annual drift rate rises from 0.71 to 0.88 (total period points divided by years, averaged across 50 companies). That is a 24.5 per cent increase in the later interval.

That increase does not mean every company drifted later. Seventeen companies show more editorial drift in Period B, twenty-five show more in Period A, and eight are evenly split. Companies filing under identical rules, in the same years, still diverge sharply in timing. That weakens any simple environmental explanation. It points instead to internal editorial processes, internal controls, and internal tolerance for language change.

## **The composition question**

The later interval is not merely a faster version of the earlier one. It is also less concentrated.

Period A is clearly dominated by sentence inflation. Period B is not dominated by a single dimension. Hedge proliferation and sentence inflation are tied, and framing drift sits close behind. The defensible shift is therefore not from inflation to hedging. It is from concentration to dispersion.

That distinction matters. Earlier drift can be characterised relatively simply: filings got longer in comparable passages, and that pattern outweighed the others. Later drift is harder to compress into one label. It contains more hedging, more framing drift, and more named-specifics loss than before, but without any one of those forms establishing clear dominance.

This makes the later interval harder to monitor with blunt instruments. Word-count tracking will still catch some of it, but not all of it. A filing can now change materially in tone or informational sharpness without announcing the shift through simple length growth alone.

**Table 1. Dimension profile by period, D1 to D5 only**

| Dimension                     | Period A share | Period B share | Period A rate (points/year ) | Period B rate (points/year ) | Change |
|-------------------------------|----------------|----------------|------------------------------|------------------------------|--------|
| D1 Distinctive claims         | 23.6%          | 15.9%          | 8.3                          | 7.0                          | -16%   |
| D2 Hedge proliferation        | 21.7%          | 25.0%          | 7.7                          | 11.0                         | +43%   |
| D3 Framing/example separation | 17.9%          | 22.7%          | 6.3                          | 10.0                         | +58%   |
| D4 Named specifics loss       | 7.5%           | 11.4%          | 2.7                          | 5.0                          | +88%   |
| D5 Sentence inflation         | 29.2%          | 25.0%          | 10.3                         | 11.0                         | +6%    |

***Rates are reported as aggregate points per year, not per-company rates.***

---

## The sector split

The aggregate masks a more uneven sector pattern. Financial services shows clear early concentration, with the annual editorial drift rate 31 per cent lower in the later interval than in the earlier one. Technology shows strong late acceleration, with the annual rate 86 per cent higher in Period B. Industrials show the sharpest late acceleration, with a 175 per cent increase. Healthcare also trends later, with a 34 per cent increase. Energy is broadly stable, with a 6 per cent decline rather than a meaningful late-period rise.

The later-period increase is therefore real in the aggregate, but it is not uniform across sectors. These sector-level rates are based on subgroups of 6 to 12 companies and should be read as indicative rather than conclusive.

## The timeline problem

The temporal frame matters, but it should not be stated more strongly than the evidence allows. ChatGPT launched on 30 November 2022, GPT-4 followed on 14 March 2023, Claude 2 in July 2023, ChatGPT Enterprise on 28 August 2023, and

Microsoft 365 Copilot became generally available for enterprise customers on 1 November 2023<sup>1</sup>. Taken together, these launches place the practical spread of enterprise-grade LLM writing tools in mid to late 2023.

For most companies in the sample, FY2022 filings were drafted in late 2022 and filed in early 2023. For most, FY2024 filings were drafted in late 2024 and filed in early 2025. In practical terms, FY2022 serves as the last broadly pre-enterprise-LLM baseline in this dataset, while FY2024 is the first filing cycle likely to have been drafted after these tools became widely accessible.

That does not prove LLM use. It does establish a timing window in which such use became more plausible. The later interval therefore captures the first filing cycle in which LLM-assisted prose could realistically have entered the drafting process at meaningful scale. Even on the narrower editorial dimensions alone, the annual drift rate is already about a quarter higher in that later interval.

### **What this does not prove**

This research documents observable patterns in filing language. It does not prove that any company used a large language model to draft its filing. Corporate prose changes for many reasons: legal review, regulatory pressure, personnel turnover, house-style drift, simplification, standardisation, and internal governance choices.

The treatment of Dimension 6 is particularly important. SEC cybersecurity disclosure rules mechanically concentrate new Item 1C sections in the later period<sup>7</sup>. For that reason, D6 is not suitable for the main editorial interval claim. It is reported separately as sensitivity.

That distinction matters numerically. On D1 to D5, the annualised increase is 24.5 per cent. When D6 is added back, Period A rises from 106 to 107 points and Period B rises from 88 to 134. Annualised, that becomes 35.7 points per year versus 67.0, or an 87.9 per cent increase. That arithmetic is real, but it is driven heavily by a structurally induced dimension and should not lead the argument.

The temporal coincidence between the later interval and the spread of enterprise LLM tools is part of the research design. It is not, by itself, a causal finding.

### **The gap in the middle**

AI now appears widely in corporate disclosure as a business topic, a governance topic, and a risk topic. The Conference Board reported in October 2025 that 72 per cent of the S&P 500 disclosed at least one material AI risk, up from 12 per cent in 2023<sup>2</sup>. EY's Center for Board Matters reported that 48 per cent of Fortune 100 companies cited AI risk as part of risk-oversight disclosures and 36 per cent treated AI as a stand-alone risk factor<sup>3</sup>.

What companies do not disclose is whether LLM-generated prose shaped the annual filing itself. There is no rule that squarely requires them to. Current disclosure frameworks remain principles-based. The unresolved question is whether LLM-assisted

drafting, if used, crosses a materiality threshold or creates an internal-controls issue that investors should understand.

That gap may matter more over time. The SEC Investor Advisory Committee considered a recommendation in December 2025 regarding disclosure of AI's impact on issuer operations<sup>4</sup>. In the EU, Article 50 of the AI Act applies from 2 August 2026 and includes transparency obligations for certain AI-generated text published to inform the public on matters of public interest, while preserving an editorial-control carve-out<sup>5</sup>. In the UK, the FRC's revised Corporate Governance Code makes Provision 29 effective for financial years beginning on or after 1 January 2026<sup>6</sup>. A company that says nothing about LLM use preserves flexibility in the short term. It also leaves itself with less evidence of editorial control if the question later becomes regulatory, litigious, or reputational.

### **The companies that did not drift**

Not every company in the dataset followed the aggregate pattern. NVIDIA scored 1 out of 12 and became more specific over time, adding product names, export thresholds, and country groupings that did not appear in earlier filings. Morgan Stanley's identity language remains close to verbatim across all three years despite major corporate change. ExxonMobil added named technologies to prose that was already operationally concrete.

Seven companies scored 2 or below. That matters because it shows that disclosure quality can hold steady or even improve under the same broad market conditions and the same tool availability that affected every other company in the sample.

Variance is not a side note here. It is part of the finding. Companies facing the same filing framework still produce very different textual outcomes.

### **Where this leads**

This analysis covers US-listed companies and SEC filings. The institutional environment for UK annual reports is different in structure, but not exempt from the same broad pressures: faster drafting tools, heavier disclosure loads, and rising expectations that published language remains traceable to identifiable human control.

The measurement window here is still short. FY2024 is best treated as the first observable filing cycle after enterprise-grade LLM writing tools became widely accessible. Even in that early window, editorial drift on D1 to D5 rises by about a quarter on an annualised basis. What changes more noticeably than the headline rate is the composition. Earlier drift is led by one pattern. Later drift is more spread across several.

That leaves governance with a more difficult problem than length alone. A filing can still pass conventional review while becoming less sharp in several smaller ways at once. The point of this analysis is not to claim a settled cause. It is to show that the language a board approves may now change faster, and in a less concentrated pattern, than many current review processes are built to detect.



## Appendix: Scoring methodology

**Sample.** The dataset comprises 50 companies selected by market capitalisation across five sectors: Financial Services (12), Technology (12), Healthcare (12), Energy (8), and Industrials (6). The sample was fixed before filing analysis began. Companies were not selected because they showed drift.

**Filing years.** FY2019 serves as the pre-pandemic baseline. FY2022 serves as the broadly pre-enterprise-LLM midpoint. FY2024 serves as the first later-period filing cycle after enterprise-grade LLM writing tools became widely accessible. For each company, comparable passages were extracted from Items 1, 1A, and 7 of the SEC 10-K.

**Passage selection rule.** Passage selection followed a deterministic rule within those sections. In Item 1, the extract began with the first substantive business-description paragraph containing a product, market, capability, or identity claim and ended when the paragraph sequence shifted from core business description into list structure, cross-reference, or a distinct subtopic. In Item 1A, the extract began with the first risk-factor passage containing operational, strategic, or competitive specificity relevant to the scored dimension and ended at the close of that immediately comparable passage. In Item 7, the extract began with the first MD&A passage discussing business performance, operations, or strategy with at least one named operational reference, and ended before the discussion moved into a different topic, segment, or accounting item. Where multiple candidate passages existed, priority was given to the earliest passage that was substantially comparable across all three years.

**Scoring.** Six dimensions of language change were scored on a 0 to 2 scale by the author, giving a maximum possible score of 12. A score of 0 means the language held steady or became more specific. A score of 2 means it became markedly less specific, more hedged, or more generic. Initial extraction and preliminary scoring were assisted by ChatGPT under structured prompts. Calibration review and verification support were assisted by Claude. All scoring judgments, methodological decisions, and editorial decisions were made by the author.

**Verification.** All extracts, quotations, word counts, section attributions, and absence claims were mechanically verified against source SEC filing extracts. Scores were then assigned by the author using the stated rubric. The interval analysis in this paper is based on a reconciled score matrix and a subsequent dimension-level period allocation conducted against the verified reports.

**Terminology.** The dimension labels use directional terms such as erosion, inflation, and proliferation because the framework measures distance from a 2019 baseline in a particular direction. A high score does not inherently mean poorer disclosure quality. It means greater change away from the baseline on that dimension.

**Limitations.** This is a descriptive study. The sample is modest by quantitative standards. The scoring involves structured human judgment; no formal inter-rater reliability test was conducted, and all final scoring decisions were made by a single author. Interval placement required reconciliation against verified reports and source

text, and some composition findings are sensitive to individual period-allocation decisions. The findings should therefore be read as disciplined observation rather than causal proof or statistical inference.

### Scoring rubric

| Dimension              | 0 (no drift)                                                | 1 (moderate)                                                  | 2 (significant)                                                   |
|------------------------|-------------------------------------------------------------|---------------------------------------------------------------|-------------------------------------------------------------------|
| D1 Distinctive claims  | Identity language unchanged or more specific                | Minor softening of falsifiable or comparative claims          | Significant removal or replacement of testable identity language  |
| D2 Hedge proliferation | No increase in qualifying language                          | 1–3 new hedges or moderate expansion of conditional phrasing  | 4+ new hedges or meaning-changing expansion of conditionality     |
| D3 Framing/example     | No separation of framing from evidence                      | Some introduction of generic framing before concrete examples | Clear pattern of abstract framing inserted ahead of specifics     |
| D4 Named specifics     | Named entities, statutes, and metrics retained or increased | Some named specifics replaced with generic categories         | Systematic replacement of named references with category language |
| D5 Sentence inflation  | Comparable passages stable or shorter                       | 10–20% longer without proportionate information gain          | 20%+ longer without proportionate information gain                |
| D6 New sections        | New sections match legacy quality or no new sections        | New sections somewhat more generic than legacy prose          | New sections markedly more generic or templated than legacy prose |

**Data availability.** The reconciled score matrix and company-level interval table are available from the author on request. The 50 verified company reports, each containing the Section 6.6 evidence tables used for scoring and interval allocation, are held by MKAI and may be released in a companion data supplement.

**Sample companies.** The 50 companies were selected from among the largest US public companies by market capitalisation, with representation across five sectors. All have December fiscal year-ends except where noted.

**Financial Services (12):** JPMorgan Chase (JPM), Bank of America (BAC), Goldman Sachs (GS), Citigroup (C), Wells Fargo (WFC), Morgan Stanley (MS), American Express (AXP), Capital One (COF), US Bancorp (USB), PNC Financial (PNC), BlackRock (BLK), Charles Schwab (SCHW).

**Technology (12):** Microsoft (MSFT), Apple (AAPL), Alphabet (GOOGL), NVIDIA (NVDA), Tesla (TSLA), Meta Platforms (META), Amazon (AMZN), Oracle (ORCL, May FY), Salesforce (CRM, Jan FY), Intel (INTC), AMD (AMD), Cisco (CSCO, Jul FY).

**Healthcare (12):** UnitedHealth (UNH), Eli Lilly (LLY), McKesson (MCK, Mar FY), Bristol-Myers Squibb (BMY), Humana (HUM), Merck (MRK), AbbVie (ABBV), CVS Health (CVS), Cigna (CI), Elevance Health (ELV), Pfizer (PFE), Johnson & Johnson (JNJ).

**Energy (8):** Occidental Petroleum (OXY), Schlumberger (SLB), Marathon Petroleum (MPC), Chevron (CVX), ConocoPhillips (COP), Phillips 66 (PSX), ExxonMobil (XOM), Valero Energy (VLO).

**Industrials (6):** GE Aerospace (GE), Honeywell (HON), UPS (UPS), RTX Corporation (RTX), Caterpillar (CAT), Deere & Co (DE, Oct FY).

## Disclosures

**Conflict of interest.** The author is the founder of MKAI. A companion governance framework, the Management Information Integrity Addendum, is published separately by Reality & Reason, of which the author is also the founder. This research was conducted independently and was not funded by any third party.

**Use of AI.** This research used large language models in its production. ChatGPT assisted with initial extraction and preliminary scoring under structured prompts. Claude assisted with calibration review, verification support, interval reconciliation support, and drafting support. All analytical judgments, score decisions, methodological choices, and editorial direction were made by the author, who takes responsibility for the final content.

**Scope.** This paper documents observable patterns only. It does not attribute cause. It does not claim to prove that any company used a large language model to draft its annual filing.

## Endnotes and references

1. OpenAI, “Introducing ChatGPT Enterprise,” 28 August 2023; OpenAI, “GPT-4 Research,” 14 March 2023; Anthropic, “Claude 2,” 11 July 2023; Microsoft 365 Blog, “Announcing Copilot for Microsoft 365 general availability and Microsoft 365 Chat,” 21 September 2023, stating general availability for enterprise customers on 1 November 2023.
2. The Conference Board and ESGAUGE, “New Study: 7 in 10 Big US Companies Report AI Risks in Public Disclosures,” 6 October 2025. The study reports that 72 per cent of the S&P 500 disclosed at least one material AI risk in 2025, up from 12 per cent in 2023.
3. EY Center for Board Matters, “Cyber and AI oversight disclosures in 2025,” reporting that 48 per cent of Fortune 100 companies disclosed a focus on AI in the risk-oversight section of the proxy statement and 36 per cent included AI as a stand-alone risk factor.
4. SEC, Investor Advisory Committee meeting materials for 4 December 2025, including “Draft Recommendation Regarding the Disclosure of Artificial Intelligence’s Impact on Operations” and related meeting notices.
5. Regulation (EU) 2024/1689 (Artificial Intelligence Act), Article 50 and Article 113. Article 50 addresses transparency obligations for providers and deployers of certain AI systems, including AI-generated or manipulated text published to inform the

public on matters of public interest, subject to human review or editorial control. The Regulation applies from 2 August 2026, subject to specified exceptions.

6. Financial Reporting Council, “FRC Revises UK Corporate Governance Code,” 22 January 2024. The FRC states that the Board declaration expectation in revised Provision 29 comes into effect for financial years beginning on or after 1 January 2026.
7. SEC, Final Rule: Cybersecurity Risk Management, Strategy, Governance, and Incident Disclosure, 26 July 2023 (effective for annual reports for fiscal years ending on or after 15 December 2023). Requires registrants to disclose cybersecurity risk management, strategy, and governance in a new Item 1C of Form 10-K.