

# Supplementary Section for Generative AI floods and dilutes the market for books

**AI Disclosures :** In accordance with responsible AI-disclosure authors of the paper acknowledge that the tables, mathematical Equations and some boiler template text was generated with AI-assistance in the Supplementary Section. Authors of the paper verified every detail for consistency and correctness.

## S1. Key quantities and figure sources

Throughout this Supplementary Information we use *title* for the ASIN-level catalog unit we analyze and *book* for the reader-facing work. Our analysis focuses on titles released from 1 January 2023 through 31 March 2026, with daily outcomes observed from 1 January 2023 through 29 June 2026. The analysis universe contains 14,419 titles in the eight focal genre clusters with non-missing full-text detected-AI scores. Revenue is an observed gross consumer-revenue proxy,  $R_{it} = Q_{it}P_{it}$ , where  $Q_{it}$  is observed unit movement and  $P_{it}$  is observed or imputed consumer price. We measure Kindle Unlimited (KU) as title availability. This is defined from current and historical KU participation flags. The panel does not separate KU page reads from a-la-carte purchases. We sort each title into one of three bands by the share of its text detected as AI: no AI text ( $S_i = 0$ ), light AI text ( $0 < S_i \leq 25\%$ ), and substantial AI text ( $S_i > 25\%$ ).

We use the 2026 Q1 endpoint for statements about releases from January 2023 through March 2026. The 2026 Q2 endpoint is a partial outcome quarter: the release universe is already closed, but sales and rank outcomes continue through 29 June 2026.

**Table 1:** Key quantities behind the main figures. Percentages may not sum to exactly 100 because of rounding.

| Quantity                                          | Value                                                                                                                                            |
|---------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------|
| Release universe after genre and score filters    | 14,419 titles                                                                                                                                    |
| No AI text                                        | 9,068 titles (62.9%)                                                                                                                             |
| Light AI text                                     | 2,471 titles (17.1%)                                                                                                                             |
| Substantial AI text                               | 2,880 titles (20.0%)                                                                                                                             |
| Titles observed for the full 90-day launch window | 14,419 titles                                                                                                                                    |
| Titles with positive launch-window sales          | 14,385 titles                                                                                                                                    |
| Launch-window sales                               | 57.0 million units                                                                                                                               |
| Launch-window revenue                             | \$280.2 million                                                                                                                                  |
| Figure 1C, no AI text share                       | 62.9% of titles, 71.7% of sales, 72.5% of revenue                                                                                                |
| Figure 1C, light AI text share                    | 17.1% of titles, 16.2% of sales, 16.2% of revenue                                                                                                |
| Figure 1C, substantial AI text share              | 20.0% of titles, 12.1% of sales, 11.3% of revenue                                                                                                |
| Figure 2A, 2026 Q1 AI text sales share            | 36.6% any AI text; 17.5% substantial AI text                                                                                                     |
| Figure 2B, 2026 Q1 AI text Top-25 slot share      | 38.5% any AI text; 18.7% substantial AI text                                                                                                     |
| Figure 3A, 2026 Q1 growth relative to 2023 Q1     | Released titles 38.35×; selling titles 19.21×; sales 7.30×; revenue 8.91×                                                                        |
| Figure 4A exposure gradient                       | No-AI text Top-25 share falls from 87.8% in the lowest-exposure bin to 62.8% in the highest                                                      |
| Figure 4B Kindle Unlimited heterogeneity          | No-AI text lead over substantial-AI text books is 8.5 points smaller for sales share and 8.4 points smaller for revenue share in KU-heavy genres |
| Figure 4C author-event population                 | 824 observed byline identities                                                                                                                   |
| Figure 4C ramp-up count                           | 287 of 385 bylines with substantial-AI text output raise their monthly output after adoption (74.5%)                                             |
| Figure 4E top author-revenue identities           | Top 15 identities earn \$8.8M of \$25.5M in substantial-AI text revenue (34.7%)                                                                  |
| Figure 4F top title-revenue pool                  | 1,753 substantial-AI text titles; top 15 earn \$4.8M of \$32.4M (14.7%)                                                                          |

**Table 2:** Analysis sample behind each figure.

| Figure or panel | Analysis sample                                                                   | Size                                          |
|-----------------|-----------------------------------------------------------------------------------|-----------------------------------------------|
| Figure 1A       | New releases in the eight genre clusters, by release quarter and AI text band     | 14,419 titles                                 |
| Figure 1B       | Launch-window sales tiers, ranked within genre-quarter cells and pooled           | 14,419 titles; Top 5% tier has 773 titles     |
| Figure 1C       | Title, sales, and revenue shares over the 90-day launch window                    | 14,419 titles                                 |
| Figure 2A       | Quarterly sales share by AI text band                                             | 14,419 titles                                 |
| Figure 2B–C     | Genre-week Top-25 rank slots, aggregated to quarters                              | 36,255 weekly slot records                    |
| Figure 2D       | Quarter-to-quarter churn in genre-level Top-25 sets                               | 104 genre-quarter transitions                 |
| Figure 3        | Quarterly accounting and launch-window revenue comparisons                        | 14,419 titles                                 |
| Figure 4A       | Genre-month exposure cells                                                        | 336 cells                                     |
| Figure 4B       | Kindle Unlimited light-versus-heavy contrast, after dropping the middle-KU genres | 252 cells across 6 genres                     |
| Figure 4C       | Author event-time panel                                                           | 824 bylines                                   |
| Figure 4D       | Top ramp-up authors                                                               | 385-byline pool; top 15 shown                 |
| Figure 4E       | Authors with a complete 12-month revenue window                                   | 466-byline pool; top 15 shown                 |
| Figure 4F       | Books with substantial AI text and a complete 12-month window                     | 1,753-title pool; top 15 shown                |
| Figure 5        | Rare-expression overlap sample                                                    | 200 books per group in the Top-200 comparison |

## S2. Notation, definitions, and analysis map

This section fixes the notation used throughout the SI. Later sections use these definitions without redefining them.

### S2.1 Units and market outcomes

The market unit is the ASIN-level title, denoted  $i$ . A book is the reader-facing work attached to that title. Day is denoted  $t$ , release cohort or calendar period by  $q$  or  $c$ , and genre by  $g$ . Observed unit movement is  $Q_{it}$  and consumer price is  $P_{it}$ . Revenue is the observed gross consumer-revenue proxy

$$R_{it} = Q_{it}P_{it}. \quad (1)$$

This is gross consumer spending before platform commissions and other deductions and does not mean author royalties or net author earnings.

### S2.2 Detected-AI bands

Each title has a full-text detected-AI score  $S_i$  on a 0–100 percentage scale. The three bands are: no detected AI text,  $S_i = 0$ ; light detected AI text,  $0 < S_i \leq 25\%$ ; and substantial detected AI text,  $S_i > 25\%$ . These labels refer to detector output after the preprocessing described in S3.5. We do not make claims about authorship in the legal sense given how a detector observes the product and not the process of how the text was produced.

### S2.3 Observation windows

The 90-day launch window contains observed pre-sales from at most 365 days before on-sale, bounded below by 1 January 2023, + days 0–90 after on-sale. A title is a launch-window selling title if it has positive observed unit movement in this window. Rank-slot analyses use genre-week or genre-month cells. Within a cell, titles are ordered by median valid Amazon sales rank. The author-event clock is measured around the first month in which an observed byline identity releases an  $S > 25\%$  title; month 0 is excluded, the post-period is months +1 through +12, and the pre-period uses all observed months before month 0. The title-own 12-month clock runs from a title’s on-sale date through the day before its 12-month anniversary. The rare-expression analysis uses top-revenue books within its comparison groups.

**Table 3:** Observation windows and where we use them.

| Window                          | Definition                                                                                                                           | Used for                               |
|---------------------------------|--------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------|
| 90-day launch                   | Pre-sales from up to 365 days before the on-sale date, starting no earlier than 1 January 2023, plus days 0 through 90 after on-sale | Figures 1 and 3; launch-window revenue |
| Genre-week and genre-month rank | Median observed sales rank within each genre-week or genre-month cell, ordered within genre                                          | Figures 2 and 4A–B                     |
| Author event time               | The months before an author’s first substantial-AI text release and months +1 to +12 after it, excluding the release month           | Figures 4C–E                           |
| Book’s first 12 months          | The day of sale through the day before the 12-month anniversary, for books with complete follow-up over that period                  | Figure 4F                              |
| Rare-expression Top- $K$        | The top-revenue books in the mostly-AI and no-AI text comparison groups                                                              | Figure 5                               |

### S2.4 Genre-month exposure panel

The genre-month exposure panel is the shared panel behind Figures 4A–B and the dilution heterogeneity checks. For each genre-month cell it records the i) share of released titles with substantial AI text ii) the no-AI text and substantial-AI text shares of sales, revenue, and Top-25 rank slots iii) Kindle Unlimited availability iv) and the month, genre, and

exposure bin. Kindle Unlimited availability is a title-level measure averaged up to the genre-month cell. The panel does not track page reads or separate Kindle Unlimited borrows from ordinary purchases.

## S2.5 Figure-to-analysis map

The SI is organized around the four claims in the paper rather than around figure order. Supplementary Table 4 maps each claim to its main-text panels and supporting SI analyses.

**Table 4:** Map from paper claims to figures and supporting SI analyses.

| Claim                                          | Main-text panels   | Main estimand                                                                                                                                                             | SI analyses                                                                                                                                                                                     |
|------------------------------------------------|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Market participation and commercial conversion | Figures 1 and 2    | Whether books with detected AI enter the catalog, convert into sales/revenue, and occupy top rank slots                                                                   | Title/sales/revenue shares, launch-sales tier representation, rank-slot construction, churn, heatmap endpoint summary, threshold and Top- $N$ robustness                                        |
| Market dilution and concentration of losses    | Figures 3 and 4A–B | Whether the expanding catalog lowers returns and rank share for books with no detected AI, and whether losses are larger where AI exposure and KU availability are higher | Quarterly accounting, annual cohort means, shift-share decomposition, price-quantity split, detector-drift bound, genre compression, exposure gradient, KU heterogeneity, few-cluster inference |
| Producer scaling and concentration             | Figures 4C–F       | Whether AI adoption is associated with higher producer output and concentrated revenue at the byline and title levels                                                     | Author-event design, population ladder, output scaling, Gini and inverse-HHI concentration, top earners and top titles                                                                          |
| Rare-expression overlap with existing books    | Figure 5           | Whether successful substantial-AI books overlap more with rare expressions from existing books than no-detected-AI books do                                               | Query design, self-exclusion, same-author/series handling, Top- $K$ coverage, revenue-gradient regression, award-reference comparison                                                           |

## S3. Data, sample, and measurement construction

This section gives the measurement and details about sample-construction details that support all four claims. Claim-specific specifications and robustness checks appear with the claims they defend.

### S3.1 Market panel and how we count the sample

The market panel links each title’s metadata to its daily Amazon outcomes. At the title level we use the ASIN, title and author metadata, on-sale date, raw subject and genre labels, current and historical Kindle Unlimited availability, and the share of the full text detected as AI. At the daily level we use unit sales, consumer price, and Amazon sales rank.

We start from books whose full text we obtained, scored for AI, matched to its Amazon record, and placed in one of the eight genre clusters. Counts from the earlier acquisition steps live in a separate acquisition log. The figures in this SI begin from the cleaned set of books described here. Supplementary Table 5 shows how the sample narrows from that point to each analysis group.

**Table 5:** How the sample narrows to each analysis group.

| Stage                                                                            | Titles or units | Note                                                                                  |
|----------------------------------------------------------------------------------|-----------------|---------------------------------------------------------------------------------------|
| Released 1 Jan 2023 to 31 Mar 2026, eight genres, with an AI score               | 14,419          | Main set of releases                                                                  |
| Observed for the full 90-day launch window                                       | 14,419          | Complete through day 90, with no missing launch-window sales or price                 |
| Positive launch-window sales                                                     | 14,385          | Denominator for revenue per selling title                                             |
| Ever appears in the weekly Top-25 rank slots                                     | 4,365           | Rank-slot analyses                                                                    |
| Genre-month exposure cells                                                       | 336             | Figures 4A and 4B                                                                     |
| Authors in the event-time panel                                                  | 824             | At least one substantial-AI text release and at least three months observed afterward |
| Authors with substantial-AI text output afterward                                | 385             | Figure 4D                                                                             |
| Authors with a complete 12-month revenue window                                  | 466             | Figure 4E                                                                             |
| Books with substantial AI text, a complete 12-month window, and positive revenue | 1,753           | Figure 4F                                                                             |
| Rare-expression Top-200 comparison                                               | 600             | 200 mostly-AI, 200 no-AI text self-published genre books, 200 award books             |

### S3.2 Getting the full texts and matching them to Amazon records

We obtained full texts three ways: authors sent us digital files, we borrowed ebooks from online libraries, and we bought ebooks from Amazon. We matched each text to its Amazon record by ASIN when the file carried one, and by title, author, edition, and release date when it did not. Because the ASIN is our market unit, we resolve or drop edition-level ambiguities before we assign a book its AI score.

### S3.3 Panel construction and genre clustering

We build two linked tables. The title table holds one row per title: when a metadata field takes more than one value for a title, we keep the most common non-missing value. The title-day table holds one row per title and calendar date, and we collapse any duplicate rows to a single value per day before building the figures. Within the launch window, every one of the 14,419 titles has complete sales and price data, with no missing days.

We group the raw genre and subject labels into eight commercial clusters, and we fix the clusters before computing any market outcome, so the grouping cannot depend on the results. Supplementary Table 6 gives the title counts by cluster, and Supplementary Table 7 shows which raw-label families map to each one.

### S3.4 AI text detection, score construction, and preprocessing

We score each book with Pangram 3.3, chapter by chapter, after stripping out non-narrative material. The stripping rule removes tables of contents, acknowledgments, dedications, copyright and ISBN pages, author biographies, newsletter and social-media pages, “also by this author” lists, advertisements, and preview chapters wherever we can identify them. Pangram labels each stretch of text it analyzes, and we treat a stretch as AI when its label is anything other than “Human Written.” Pangram’s own validation of version 3.3 reports a false-positive rate of 0.04% and a false-negative rate of 0.14%, weighted by sample size, on a held-out set of 566,745 texts.

**Table 6:** Final title counts by genre cluster.

| Genre cluster                           | Titles | Share (%) |
|-----------------------------------------|--------|-----------|
| General Romance                         | 4,165  | 28.9      |
| Speculative / Adventure Romance         | 2,225  | 15.4      |
| Fantasy / Supernatural / Horror         | 2,190  | 15.2      |
| General / Contemporary Fiction          | 1,647  | 11.4      |
| Mystery / Thriller / Crime              | 1,538  | 10.7      |
| Crime / Suspense Romance                | 1,376  | 9.5       |
| Sports Romance                          | 836    | 5.8       |
| Science Fiction / Adventure / Dystopian | 442    | 3.1       |

**Table 7:** Genre-cluster crosswalk used in the frozen analysis.

| Cluster                                 | Raw-label families                                                                                                                                            |
|-----------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------|
| General Romance                         | Contemporary, small-town, billionaire, comedy, holiday, historical, gothic, military, western romance, erotica, and residual romance labels                   |
| Speculative / Adventure Romance         | Fantasy romance, paranormal romance, science-fiction romance, and action/adventure romance                                                                    |
| Crime / Suspense Romance                | Romantic suspense and mafia romance                                                                                                                           |
| Sports Romance                          | Sports romance and hockey romance                                                                                                                             |
| Fantasy / Supernatural / Horror         | Fantasy, dark fantasy, epic fantasy, LitRPG/progression fantasy, supernatural fiction, paranormal fantasy, urban fantasy, and horror                          |
| Mystery / Thriller / Crime              | Crime fiction, mystery/suspense, thriller, psychological thriller, and cozy mystery                                                                           |
| Science Fiction / Adventure / Dystopian | Science fiction, adventure fiction, dystopian, and post-apocalyptic fiction                                                                                   |
| General / Contemporary Fiction          | Contemporary fiction, women’s fiction, LGBTQ+ fiction, historical fiction, literary fiction, young adult fiction, humorous fiction, and miscellaneous fiction |

Let  $W_i$  be the set of analyzed text windows for book  $i$ , and let  $\ell(w)$  denote the Pangram label assigned to window  $w$ . We define the book’s primary AI-detection score as the percentage of analyzed windows that Pangram labels as anything other than “Human Written”:

$$P_i = 100 \times \frac{|\{w \in W_i : \ell(w) \neq \text{“Human Written”}\}|}{|W_i|}. \quad (2)$$

As an alternative, we define a length-weighted score that weights each window by its token count:

$$P_i^{LW} = 100 \times \frac{\sum_{w \in W_i} \text{tokens}_w \mathbf{1}\{\ell(w) \neq \text{“Human Written”}\}}{\sum_{w \in W_i} \text{tokens}_w}. \quad (3)$$

Thus,  $P_i$  measures the percentage of analyzed windows flagged as AI, whereas  $P_i^{LW}$  measures the percentage of analyzed token mass contained in flagged windows. Both scores are computed after the preprocessing described above.

### S3.5 Revenue, price, and Kindle Unlimited

We define revenue in S2.1. When a launch-window sale has a missing or zero-or-negative price, we fill it with the title’s median valid price over the same window; after this step, no launch-window rows need to be dropped for a missing quantity or a bad price. We measure Kindle Unlimited as whether a title is available on the service, not as how much of it readers actually read. The panel does not tell us whether a given unit is a Kindle Unlimited borrow, a page-read allocation, or an ordinary purchase, so in S5.7 we report the Kindle Unlimited result using unit-sales shares alongside revenue shares.

## S4. Claim 1: AI text books spread widely but sell below their share

The first claim is that books with AI text are not stuck at the bottom of the catalog. They appear across genres, take a smaller share of sales and revenue than of titles, and turn up more and more often in the top rank slots. Figure 1 looks at how these books enter the catalog and how they sell in their launch window; Figure 2 looks at their share of sales, their share of the Top-25 rank slots, and how the top ranks turn over.

## S4.1 Title, sales, and revenue shares

A book’s share of the catalog need not match its share of sales, so we first ask whether AI text books turn their presence in the catalog into a matching share of sales and revenue.

**Table 8:** Share of titles, sales, and revenue by AI text band over the launch window.

| AI text band        | Titles | Title share (%) | Sales share (%) | Revenue share (%) |
|---------------------|--------|-----------------|-----------------|-------------------|
| No AI text          | 9,068  | 62.9            | 71.7            | 72.5              |
| Light AI text       | 2,471  | 17.1            | 16.2            | 16.2              |
| Substantial AI text | 2,880  | 20.0            | 12.1            | 11.3              |

Books with substantial AI text make up 20.0% of launch-eligible titles but only 12.1% of launch-window sales and 11.3% of revenue. Books with no AI text make up 62.9% of titles and a larger 72.5% of revenue.

## S4.2 AI text across the sales tiers

We rank books by launch-window sales within each genre-quarter cell, then pool the top fractions across cells, so that a few large genres or busy release quarters cannot drive the comparison at the top of the market.

**Table 9:** AI text makeup of the launch-sales tiers. We form tiers within genre-quarter cells and then pool them.

| Tier       | Titles | No AI text (%) | Light AI text (%) | Substantial AI text (%) | Any AI text (%) |
|------------|--------|----------------|-------------------|-------------------------|-----------------|
| All titles | 14,419 | 62.9           | 17.1              | 20.0                    | 37.1            |
| Top 50%    | 7,235  | 65.9           | 17.6              | 16.4                    | 34.1            |
| Top 25%    | 3,643  | 68.7           | 17.2              | 14.1                    | 31.3            |
| Top 10%    | 1,490  | 71.0           | 17.7              | 11.3                    | 29.0            |
| Top 5%     | 773    | 72.7           | 16.8              | 10.5                    | 27.3            |
| Top 1%     | 196    | 71.4           | 17.9              | 10.7                    | 28.6            |

The substantial-AI text share falls from 20.0% of all launch-eligible titles to 10.5% of the Top 5% and 10.7% of the Top 1%. Apart from the smallest tier, where  $n = 196$  and the share ticks up slightly, the pattern is a steady decline toward the top of the market.

## S4.3 Sales ranks, top slots, and turnover

For each genre-week cell, we take each title’s median valid Amazon sales rank and order titles from best to worst rank. The Top-25 set is the first  $\min(25, n_{gw})$  titles in genre  $g$  and week  $w$ ; the Top-10 check uses the same rule with ten. We then aggregate the makeup of these slots to calendar quarters by AI text band.

**Table 10:** Share of sales and of Top-25 slots by AI text band, at selected quarters. The full quarterly series appears in Figure 2. The 2026 Q2 row is a partial quarter ending 29 June 2026.

| Quarter | Sales share (%) |       |             | Top-25 slot share (%) |       |             |
|---------|-----------------|-------|-------------|-----------------------|-------|-------------|
|         | No AI           | Light | Substantial | No AI                 | Light | Substantial |
| 2023 Q1 | 95.6            | 4.2   | 0.2         | 90.7                  | 8.1   | 1.2         |
| 2026 Q1 | 63.4            | 19.1  | 17.5        | 61.5                  | 19.7  | 18.7        |
| 2026 Q2 | 60.3            | 19.4  | 20.3        | 59.6                  | 17.1  | 23.3        |

Let  $T_{gq}^{25}$  be the Top-25 set in genre  $g$  and quarter  $q$ . Retention measures how many of last quarter’s no-AI text bestsellers are still in the Top-25 this quarter,

$$\text{Retention}_{gq} = \frac{|\{i \in T_{g,q-1}^{25} : i \in \text{NoAI}\} \cap T_{gq}^{25}|}{|\{i \in T_{g,q-1}^{25} : i \in \text{NoAI}\}|}. \quad (4)$$

The entrant share measures how many of this quarter’s new arrivals carry substantial AI text,

$$\text{EntrantShare}_{gq} = \frac{|\{i \in T_{gq}^{25} \setminus T_{g,q-1}^{25} : i \in S > 25\%\}|}{|T_{gq}^{25} \setminus T_{g,q-1}^{25}|}. \quad (5)$$

Figure 2D plots the unweighted genre means of these two quantities by quarter.

**Table 11:** Turnover in the Top-25, at selected quarters. Values are unweighted genre means.

| Quarter | No-AI text retention (%) | Substantial-AI text entrant share (%) |
|---------|--------------------------|---------------------------------------|
| 2023 Q2 | 44.1                     | 3.3                                   |
| 2024 Q2 | 28.2                     | 6.8                                   |
| 2026 Q1 | 52.5                     | 30.7                                  |
| 2026 Q2 | 62.5                     | 31.7                                  |

#### S4.4 Substantial AI text in the top slots, by genre

The full Figure 2C heatmap covers 112 genre-quarter cells, which we release as machine-readable data. Supplementary Table 12 shows a compact version at the endpoints; the full series across all quarters is in Figure 2C. A cell is blank only when a genre-quarter has no valid Top-25 observations after rank filtering, and in this paper every genre has valid observations in every outcome quarter. Because the 2026 Q2 cells cover only a partial quarter, we read any large jumps in that column as noisy boundary values rather than as trends on their own.

**Table 12:** Substantial-AI text share of Top-25 slots by genre, at selected quarters. Values are percentages, and the full quarterly series is in Figure 2C. The 2026 Q2 column covers a partial quarter ending 29 June 2026, so its values are more volatile than those of complete quarters.

| Genre                                   | 2023 Q1 | 2026 Q1 | 2026 Q2 |
|-----------------------------------------|---------|---------|---------|
| General Romance                         | 0.3     | 15.7    | 17.8    |
| Speculative / Adventure Romance         | 0.0     | 20.3    | 14.2    |
| Fantasy / Supernatural / Horror         | 1.2     | 5.8     | 21.8    |
| General / Contemporary Fiction          | 0.0     | 21.8    | 17.5    |
| Mystery / Thriller / Crime              | 3.4     | 33.5    | 33.2    |
| Crime / Suspense Romance                | 0.0     | 24.3    | 25.2    |
| Sports Romance                          | 6.6     | 12.6    | 17.5    |
| Science Fiction / Adventure / Dystopian | 0.6     | 15.7    | 39.4    |

We use the Top-25 in the main rank-slot figure because it captures the visible top of the market while leaving enough books in each genre-quarter cell to compare. Narrowing to ranks 1–10 gives the same picture: books with AI text move into the top ranks over time, while books with no AI text still hold most of the slots. The Top-50 version, built from the full rank file, is in the machine-readable rank data.

#### S4.5 Does the 25% cutoff matter?

We call a book substantial AI text when more than 25% of its text is flagged, and the result should not hinge on that exact line. Supplementary Table 13 moves the cutoff up and down and reports the share of titles, sales, and revenue that fall in the substantial band at each setting. The picture holds throughout: a higher cutoff puts fewer books in the substantial band, but at every setting the band takes a smaller share of sales and revenue than of titles.

### S5. Claim 2: The flood diluted returns for books with no AI text

The second claim is that the surge of AI text books pulled down returns and rank positions for books with no AI text, rather than simply adding low-revenue AI text books to the denominator. We support this claim three ways: a shift-share decomposition, a price-quantity split, and a detector-drift bound. We then show that the losses run deeper where substantial AI text is more common and where more titles are available on Kindle Unlimited.



**Table 13:** The substantial band at different cutoffs, over the launch window.

| Cutoff     | Titles | Title share (%) | Sales share (%) | Revenue share (%) |
|------------|--------|-----------------|-----------------|-------------------|
| $S > 10\%$ | 3,586  | 24.9            | 16.1            | 15.2              |
| $S > 20\%$ | 3,066  | 21.3            | 13.1            | 12.2              |
| $S > 25\%$ | 2,880  | 20.0            | 12.1            | 11.3              |
| $S > 30\%$ | 2,741  | 19.0            | 11.0            | 10.2              |
| $S > 40\%$ | 2,440  | 16.9            | 9.4             | 8.6               |
| $S > 50\%$ | 2,168  | 15.0            | 7.9             | 7.1               |

### S5.1 How we track the market quarter by quarter

For each quarter  $q$ , we count the cumulative released titles, the titles with any observed sales, total sales, and total revenue, and we set each series to 1 in 2023 Q1 so later quarters read as multiples of that starting point. Releases stop in 2026 Q1, so the released-title line is flat in 2026 Q2 even though sales and rank keep coming in through 29 June 2026.

Revenue per selling title in a calendar quarter is

$$\text{RevPerSellingTitle}_q = \frac{\sum_i \sum_{t \in q} R_{it}}{|\{i : \sum_{t \in q} Q_{it} > 0\}|}. \quad (6)$$

For the launch-window comparisons, we compute the numerator and denominator over the launch window from S2.3.

**Table 14:** Growth of each series relative to 2023 Q1.

| Series          | 2026 Q1 | 2026 Q2 |
|-----------------|---------|---------|
| Released titles | 38.35×  | 38.35×  |
| Selling titles  | 19.21×  | 19.14×  |
| Sales           | 7.30×   | 8.12×   |
| Revenue         | 8.91×   | 9.84×   |

**Table 15:** Launch-window revenue per selling title, by release year. The 2026 row is a partial year because releases stop on 31 March 2026. These yearly means show the 2024 bump for books with no AI text that appears in the detector-drift bound, and they make clear that the decomposition below rests on a net decline from 2023 to 2025 rather than a fall in every single year.

| Release year | All books | No AI text | Light AI text | Substantial AI text |
|--------------|-----------|------------|---------------|---------------------|
| 2023         | \$20,134  | \$23,877   | \$16,911      | \$5,776             |
| 2024         | \$22,076  | \$24,316   | \$20,182      | \$11,135            |
| 2025         | \$17,312  | \$19,739   | \$19,218      | \$11,576            |
| 2026 partial | \$13,176  | \$15,524   | \$11,590      | \$11,487            |

### S5.2 Splitting the decline into composition and within-band change

The main worry about a mix shift is that revenue per book could fall on its own if the catalog simply fills with many low-revenue AI-heavy books. To separate that composition effect from real change within each band, write launch-window revenue per selling title in cohort  $t$  as

$$\bar{R}_t = \sum_b s_{bt} \mu_{bt}, \quad (7)$$

where  $s_{bt}$  is the share of selling titles in band  $b$  and  $\mu_{bt}$  is revenue per selling title within that band. We split the change from 2023 to 2025 into three parts,

$$\Delta \bar{R} = \underbrace{\sum_b (s_{b,2025} - s_{b,2023}) \mu_{b,2023}}_{\text{composition}} + \underbrace{\sum_b s_{b,2023} (\mu_{b,2025} - \mu_{b,2023})}_{\text{within-band}} + \underbrace{\sum_b (s_{b,2025} - s_{b,2023}) (\mu_{b,2025} - \mu_{b,2023})}_{\text{interaction}}. \quad (8)$$

**Table 16:** Splitting the 2023-to-2025 change in launch-window revenue per selling title. Dollar values are per selling title.

| Band                | $s_{2023}$ (%) | $\mu_{2023}$ | $s_{2025}$ (%) | $\mu_{2025}$ | Composition | Within-band | Interaction |
|---------------------|----------------|--------------|----------------|--------------|-------------|-------------|-------------|
| No AI text          | 66.3           | \$23,877     | 53.4           | \$19,739     | \$-3,078    | \$-2,742    | \$533       |
| Light AI text       | 21.2           | \$16,911     | 18.0           | \$19,218     | \$-538      | \$490       | \$-73       |
| Substantial AI text | 12.5           | \$5,776      | 28.6           | \$11,576     | \$928       | \$725       | \$932       |

Overall, launch-window revenue per selling title falls by \$2,822, from \$20,134 in 2023 to \$17,312 in 2025. The year-by-year path includes a 2024 bump, and that bump shows up in all three bands, not just the no-AI text band (Supplementary Table 15), so the claim is a net decline from 2023 to 2025, not a fall in every year. Because the 2024 rise is shared across bands, it points to a common demand or supply movement rather than something specific to books with no AI text. Most important, the within-band term for no AI text is negative: revenue per selling title for these books falls from \$23,877 to \$19,739, a 17.3% drop in levels (the same change reads as  $-19.0$  log points; Supplementary Table 17 and Supplementary Table 21 line up the two). This within-band fall for no-AI text books is the heart of the result, the decline is not just the catalog filling with cheaper AI books.

### S5.3 Is the decline about price or about units?

Revenue per selling title can fall for two reasons: each selling title sells fewer units, or each unit brings in less money. The two mean different things for dilution, so we separate them. For any cohort or quarter,

$$\frac{R}{N_{sell}} = \frac{Q}{N_{sell}} \times \frac{R}{Q}, \quad (9)$$

and taking logs turns the product into a sum,

$$\Delta \log \left( \frac{R}{N_{sell}} \right) = \Delta \log \left( \frac{Q}{N_{sell}} \right) + \Delta \log \left( \frac{R}{Q} \right). \quad (10)$$

**Table 17:** Splitting the change in revenue per selling title into a units part and a price part, 2023 versus 2025. Log changes are multiplied by 100, so they are not level percentage changes; Supplementary Table 21 gives the matching level percentages for selected rows.

| Group      | $R/N$ , 2023 | $R/N$ , 2025 | $\Delta \log R/N$ | $Q/N$ , 2023 | $Q/N$ , 2025 | $\Delta \log Q/N$ | $R/Q$ , 2023 | $R/Q$ , 2025 | $\Delta \log R/Q$ |
|------------|--------------|--------------|-------------------|--------------|--------------|-------------------|--------------|--------------|-------------------|
| All books  | \$20,134     | \$17,312     | -15.1             | 4,506        | 3,423        | -27.5             | \$4.47       | \$5.06       | 12.4              |
| No AI text | \$23,877     | \$19,739     | -19.0             | 5,282        | 3,844        | -31.8             | \$4.52       | \$5.14       | 12.8              |

For books with no AI text, revenue per selling title falls even though revenue per unit rises. So the decline comes from each selling title moving fewer units, only partly offset by a higher price per unit. It is a demand story, not a price war.

### S5.4 Could undetected AI books explain the decline?

The within-band fall for books with no AI text could be challenged one way: if later AI-assisted books are harder to catch, some of them slip into the no-AI text group and drag its average down. We bound how much of this would have to be happening. The rows below are release years, with 2026 a partial year because releases stop on 31 March 2026. Let  $\mu_{0c}^{obs}$  be the observed no-AI text revenue per selling title in release year  $c$ ,  $\mu_{Hc}^{true}$  the true human mean,  $\mu_c^{missedAI}$

the mean revenue of the AI books the detector misses, and  $\pi_c$  the share of the observed no-AI text group that is really missed AI. Then

$$\mu_{0c}^{obs} = (1 - \pi_c)\mu_{Hc}^{true} + \pi_c\mu_c^{missedAI}. \quad (11)$$

If there were no real dilution, the true human mean would hold at its 2023 level  $\mu_H^{base}$ . The share of missed AI needed to produce the observed decline is then

$$\pi_c^{req} = \frac{\mu_H^{base} - \mu_{0c}^{obs}}{\mu_H^{base} - \mu_c^{missedAI}}. \quad (12)$$

**Table 18:** Share of undetected AI that would be needed to explain the no-AI text decline if there were no real dilution. Rows are release years; 2026 is a partial year ending 31 March 2026. The baseline is the 2023 no-AI text launch-window revenue per selling title, \$23,877. The “substantial-AI text mean” column assumes missed AI books earn as much as the substantial band; the “zero-revenue” column assumes they earn nothing, which gives the smallest share the bound allows. Negative values mean no missed AI is needed, because that year’s observed no-AI text mean is above the baseline.

| Year | No-AI text mean | Substantial-AI text mean | Needed share, substantial mean (%) | Needed share, zero revenue (%) |
|------|-----------------|--------------------------|------------------------------------|--------------------------------|
| 2024 | \$24,316        | \$11,135                 | -3.4                               | -1.8                           |
| 2025 | \$19,739        | \$11,576                 | 33.6                               | 17.3                           |
| 2026 | \$15,524        | \$11,487                 | 67.4                               | 35.0                           |

To wipe out the 2025 decline, 33.6% of the no-AI text group would have to be undetected AI if those missed books earned as much as the substantial band, or 17.3% even under the extreme assumption that they earned nothing. Both shares are far above the detector’s false-negative rate, and a miss rate that high would be visible in the score distribution unless the misses were both severe and heavily concentrated in the later years. So undetected AI does not account for the decline.

## S5.5 The decline shows up genre by genre

If this is real dilution and not just a market-wide slump, we should see the decline across most genres, but not in the corner of the market where AI barely showed up.

**Table 19:** Launch-window revenue per selling title by genre, 2023 versus 2025.

| Genre                                   | All books |          |            | No AI text |          |            |
|-----------------------------------------|-----------|----------|------------|------------|----------|------------|
|                                         | 2023      | 2025     | Change (%) | 2023       | 2025     | Change (%) |
| General Romance                         | \$20,546  | \$18,235 | -11.2      | \$24,400   | \$22,407 | -8.2       |
| Speculative / Adventure Romance         | \$15,999  | \$11,920 | -25.5      | \$18,694   | \$11,970 | -36.0      |
| Fantasy / Supernatural / Horror         | \$13,258  | \$16,019 | +20.8      | \$14,128   | \$19,098 | +35.2      |
| General / Contemporary Fiction          | \$28,239  | \$21,123 | -25.2      | \$35,349   | \$22,052 | -37.6      |
| Mystery / Thriller / Crime              | \$15,265  | \$11,180 | -26.8      | \$20,079   | \$12,638 | -37.1      |
| Crime / Suspense Romance                | \$29,360  | \$21,982 | -25.1      | \$35,514   | \$22,989 | -35.3      |
| Sports Romance                          | \$45,983  | \$23,169 | -49.6      | \$33,618   | \$27,081 | -19.4      |
| Science Fiction / Adventure / Dystopian | \$17,226  | \$17,934 | +4.1       | \$20,699   | \$19,826 | -4.2       |

Revenue per selling title falls in six of the eight genres for all books, and in seven of the eight for books with no AI text. The one exception among no-AI text books is Fantasy / Supernatural / Horror, which rose 35.2%, and that is exactly the genre AI reached last and least. The decline is broad, and it fades where AI stayed away, which is what real dilution should look like.

## S5.6 The more AI in a genre, the fewer top slots for books with no AI text

For Figure 4A we sort the genre-month cells by how much of the catalog is substantial AI text, and in each group we plot the average share of Top-25 slots held by books with no AI text. Bars are  $\bar{y} \pm 1.96 SE$ .

The share of top slots held by books with no AI text falls from 87.8% in the least-exposed group to 62.8% in the most-exposed one.

**Table 20:** Top-25 share for books with no AI text, by how much AI text is in the catalog.

| Group | Cells | Mean substantial-AI text share | Mean no-AI text Top-25 share | 95% interval |
|-------|-------|--------------------------------|------------------------------|--------------|
| 1     | 38    | 0.4                            | 87.8                         | 85.3–90.2    |
| 2     | 38    | 5.5                            | 81.8                         | 78.7–84.9    |
| 3     | 38    | 8.3                            | 78.9                         | 76.6–81.2    |
| 4     | 37    | 10.3                           | 79.5                         | 76.9–82.0    |
| 5     | 37    | 12.6                           | 72.5                         | 70.3–74.6    |
| 6     | 37    | 15.4                           | 69.9                         | 66.9–72.9    |
| 7     | 37    | 17.5                           | 67.6                         | 64.0–71.2    |
| 8     | 37    | 19.7                           | 66.7                         | 62.8–70.6    |
| 9     | 37    | 24.7                           | 62.8                         | 58.6–67.0    |

### S5.7 The 2023-to-2025 decline under alternative measures

We check the 2023-to-2025 decline under several alternative measures (Supplementary Table 21). The pattern holds throughout: units sold per selling title fall for all books and for books with no AI text alone, and the decline stays visible whether we use the mean, the median, or a winsorized mean.

**Table 21:** The 2023-to-2025 decline under other measures. The change column is a level percentage change, not a log change; for instance, the fall in units per selling title for books with no AI text is –27.2% here and –31.8 log points in Supplementary Table 17.

| Measure                                        | Group      | 2023     | 2025     | Change (%) |
|------------------------------------------------|------------|----------|----------|------------|
| Revenue per released title                     | All books  | \$20,100 | \$17,244 | -14.2      |
| Revenue per released title                     | No AI text | \$23,836 | \$19,655 | -17.5      |
| Mean revenue per selling title                 | All books  | \$20,134 | \$17,312 | -14.0      |
| Mean revenue per selling title                 | No AI text | \$23,877 | \$19,739 | -17.3      |
| Median revenue per selling title               | All books  | \$5,664  | \$5,235  | -7.6       |
| Median revenue per selling title               | No AI text | \$6,640  | \$6,334  | -4.6       |
| 1/99 winsorized mean revenue per selling title | All books  | \$18,426 | \$16,371 | -11.2      |
| 1/99 winsorized mean revenue per selling title | No AI text | \$22,288 | \$18,653 | -16.3      |
| Unit sales per selling title                   | All books  | 4,506    | 3,423    | -24.0      |
| Unit sales per selling title                   | No AI text | 5,282    | 3,844    | -27.2      |

## S6. Claim 3: A broad base of authors scale-up, and a few reach large scale

The third claim is that adopting AI lets a wide range of authors put out more books, while a prolific few reach very large output and revenue. We track authors by their byline or pen name rather than by legal identity, so when one person writes under several pen names our concentration figures understate how concentrated production really is.

### S6.1 When an author adopts AI, and who enters the analysis

For each byline, the adoption month is the first month it releases a book with more than 25% AI text. We leave out that month itself when comparing before and after, because it holds the first substantial-AI text release and often a burst of others alongside it. The “after” period is the next twelve months, and a byline enters the analysis once we can observe at least three of them.

### S6.2 Do authors publish faster after adopting AI?

It depends which authors we mean. Across all 824 bylines, average monthly output edges up from 0.143 books before adoption to 0.190 after, but only 311 of the 824 (37.7%) actually publish faster than before. The clearer story is among the authors who keep using AI: of the 385 who release more substantial-AI text books afterward, 287 (74.5%) raise their monthly output. So the speed-up is real, but it is concentrated among the authors who stick with AI, not spread evenly across everyone who tries it once.

**Table 22:** How the author groups in Figure 4C–E are built. Counts are byline or pen-name identities, except the last row, which counts individual books.

| Group                               | Count | Who is in it                                                                                           |
|-------------------------------------|-------|--------------------------------------------------------------------------------------------------------|
| Authors in the analysis             | 824   | A first substantial-AI text release and at least three months observed afterward.                      |
| Authors we saw before adoption      | 374   | Published at least one book before their first substantial-AI text release.                            |
| Authors new at adoption             | 450   | First appear when they release their first substantial-AI text book.                                   |
| Full twelve-month follow-up         | 469   | Can be followed through the twelfth month after adoption.                                              |
| Kept publishing afterward           | 426   | Released at least one book in the twelve months after adoption.                                        |
| Kept publishing substantial AI text | 385   | Released at least one substantial-AI text book afterward; Figure 4D.                                   |
| And raised their output             | 287   | Of those, published faster after adoption than before.                                                 |
| Top-revenue author pool             | 466   | Full-follow-up authors with positive substantial-AI text revenue; Figure 4E.                           |
| Top-revenue book pool               | 1,753 | Individual substantial-AI text books with full twelve-month follow-up and positive revenue; Figure 4F. |

**Table 23:** Output before and after adoption, by group of authors. Rates are books per month.

| Group                               | Bylines | Mean rate before | Mean rate after | Median rate after | Publish faster (%) |
|-------------------------------------|---------|------------------|-----------------|-------------------|--------------------|
| All authors in the analysis         | 824     | 0.143            | 0.190           | 0.083             | 37.7               |
| Authors we saw before adoption      | 374     | 0.314            | 0.308           | 0.167             | 42.5               |
| Authors new at adoption             | 450     | 0.000            | 0.091           | 0.000             | 33.8               |
| Kept publishing substantial AI text | 385     | 0.250            | 0.384           | 0.250             | 74.5               |
| And raised their output             | 287     | 0.125            | 0.380           | 0.250             | 100.0              |

### S6.3 How concentrated is output and revenue, and who is at the top?

We measure concentration two ways: the Gini coefficient, which runs from 0 (everyone equal) toward 1 (all in one hand), and the inverse-Herfindahl effective count  $N_{eff} = 1/\sum_a p_a^2$ , which reads roughly as “how many equal-sized producers this would take.” Here  $p_a$  is the share of output or revenue held by author or book  $a$ .

**Table 24:** Concentration of output and revenue after adoption. Revenue is gross consumer spending, not author earnings.

| Quantity                                           | Population  | Total       | Gini  | Effective count | Top-15 share (%) |
|----------------------------------------------------|-------------|-------------|-------|-----------------|------------------|
| Substantial-AI text books published after adoption | 385 bylines | 1,310 books | 0.476 | 165.4           | 20.2             |
| Same, among authors who raised their output        | 287 bylines | 1,022 books | 0.483 | 125.4           | 24.0             |
| Revenue from those books                           | 466 bylines | \$25.5M     | 0.737 | 71.2            | 34.7             |
| Revenue per book over its first 12 months          | 1,753 books | \$32.4M     | 0.730 | 297.3           | 14.7             |

Figure 4E ranks authors by the revenue their substantial-AI text books earn in the twelve months after their first such release, among authors we can follow for the full year. The top 15 authors earn \$8.83M of the \$25.49M in this pool, on 1.86M of its 5.65M sales, from 87 of its 1,216 books. That 1,216 is smaller than the 1,310 books in the concentration table above because this revenue pool keeps only authors with a full twelve months of follow-up and positive revenue, while the output count keeps every author with at least three months observed.

Figure 4F ranks individual substantial-AI text books by what each earns in its own first twelve months. The pool starts from the 2,880 substantial-AI text books in the data. Of these, 1,757 can be followed for a full twelve months by the 29 June 2026 cutoff, and 1,753 earned something over that window. The top 15 books earn \$4.78M of the pool’s \$32.44M, on 895,417 of its 7.32M sales. The single top book earns \$642,644 on 80,431 sales. Of the 1,753 books, 1,609 have a complete run of daily data over their twelve-month window and 144 have some days missing; all 15 of the top books have complete data.

## S7. Claim 4: Rare-expression overlap with existing books

The fourth claim is that successful substantial-AI books overlap more with rare expressions from existing books than comparable no-detected-AI books do. The measure captures textual overlap with rare expressions appearing in existing books; it does not by itself identify copying from any particular source.

## S7.1 Text preprocessing, tokenization, and query design

The rare-expression analysis uses the same stripped narrative text described in S3.4. We tokenize each book into word sequences and enumerate expressions of at least five words. An expression counts as rare existing-book language when it appears in at most five Google Books volumes and has count zero in infini-gram, a 4.7-trillion-token web snapshot. The first condition selects language concentrated in a small number of books; the second removes expressions that circulate widely on the web. For each book, rare-expression coverage is the share of tokens covered by at least one such expression. We do not count the tested book itself as an existing-book source for its own rare-expression coverage. The primary reported values in Figure 5 are rare-expression overlap with existing books after self-overlap controls.

## S7.2 AI threshold and top- $K$ comparisons

The rare-expression analysis compares books with no detected AI against books that are more than 25% detected AI.

**Table 25:** Rare-expression coverage among top-revenue self-published genre books. Values are mean token coverage percentages for rare expressions of at least five words. Differences are computed from unrounded means.

| Revenue tier | Substantial AI (%) | No detected AI (%) | Difference (points) | Test result         |
|--------------|--------------------|--------------------|---------------------|---------------------|
| Top 50       | 45.0               | 37.7               | 7.2                 | Welch $p < 10^{-8}$ |
| Top 100      | 43.2               | 37.6               | 5.5                 | Welch $p < 10^{-8}$ |
| Top 200      | 41.6               | 37.2               | 4.3                 | Welch $p < 10^{-8}$ |

## S7.3 Revenue-gradient regression and award-reference set

To test whether overlap rises with commercial success, we estimate on the top 200 books in each group ( $N = 400$ )

$$C_i = \alpha + \beta_1 AI_i + \beta_2 \log_{10}(Revenue_i) + \beta_3 AI_i \times \log_{10}(Revenue_i) + \epsilon_i, \quad (13)$$

where  $C_i$  is rare-expression coverage in percentage points and  $AI_i$  indicates substantial-AI books, so  $\beta_2$  is the revenue slope for no-AI text books and  $\beta_2 + \beta_3$  the slope for substantial-AI books. Coverage rises by  $\hat{\beta}_2 + \hat{\beta}_3 = 7.6$  percentage points per tenfold revenue increase for substantial-AI books ( $p = 1.4 \times 10^{-3}$ ), compared with  $\hat{\beta}_2 = 1.1$  points for no-AI text books ( $p = 0.67$ ); slope  $p$ -values are from per-group fits. The interaction, which tests the slope difference, gives  $\hat{\beta}_3 = 6.5$  points (95% CI  $[-0.4, 13.3]$ ,  $p = 0.063$ ); results are unchanged with heteroskedasticity-robust (HC3) standard errors ( $p = 0.067$ ).

**Table 26:** Rare-expression coverage in the three Figure 5C reference groups. Values are means for  $n = 200$  books per group.

| Group                                                 | Mean coverage (%) |
|-------------------------------------------------------|-------------------|
| Substantial-AI text (>25%) self-published genre books | 41.6              |
| No-AI text self-published genre books                 | 37.2              |
| Literary award-winning reference books                | 19.1              |

The award-reference set consists of books that won or were nominated for major literary prizes across literary, historical, and speculative fiction such as *The Road* by Cormac McCarthy, *Pachinko* by Min Jin Lee, *The Underground Railroad* by Colson Whitehead, *Americanah* by Chimamanda Ngozi Adichie etc. It is a reference group for high-status literary authorship, not a market-matched control for self-published genre fiction.

## S8. Reproducibility, data availability, and ethics

The raw Amazon sales/rank panel is proprietary, and we cannot redistribute it. The full book texts are copyrighted or author-provided, and we do not redistribute them. The reproducible release contains aggregate figure data, de-identified genre-month and cohort-level panels where permitted, derived score bands where permitted by the detector terms, and

code that recreates the figures from authorized inputs. Public tables anonymize byline and title identifiers in the producer and top-title analyses.

We use book texts only to compute detector scores and rare-expression coverage, and we do not redistribute underlying text passages. Author-provided files and purchased or borrowed books are stored in restricted project storage. Outputs report aggregate market quantities or anonymized identifiers unless a title is already discussed as part of a public figure-level result.