Does Generative AI Make Startup Ideas Alike? Evidence from Product Hunt

Zhaoqi (ZQ) Cheng * Kaige Gao †

June 2026

Abstract

Founders increasingly use generative AI to develop new products. Because a market's stock of distinct ideas shapes what it can build next, we ask whether generative AI broadens or narrows the ideas that founders bring to it. We study 291,314 product launches posted to Product Hunt from January 2018 to April 2026. After ChatGPT's release, launches become more concentrated around common category concepts, and the share of distinctive launches falls by more than a third, from 25% to 16%. Using an outcome-blind LLM ideation-exposure index, we find that the post-ChatGPT increase in similarity is larger in categories where general-purpose LLMs can more effectively assist product ideation. First-time teams introduce more distinctive ideas than experienced teams, but similarity rises within both groups, and holding founder composition fixed leaves the exposure gradient essentially unchanged. These findings are consistent with independent founders drawing on a common input and converging toward common category concepts. Generative AI may thus expand individual entrepreneurs' ability to develop products while narrowing the range of ideas that entrepreneurial entry supplies to the market.

Keywords: Generative AI; idea diversity; entrepreneurial innovation; Product Hunt.

1 Introduction

The innovation process is cumulative and recombinant: a more varied idea pool expands the set of possibilities for future discovery (Romer, 1990; Weitzman, 1998). Entrepreneurship is an important mechanism through which novel combinations are developed and brought into economic use (Shane & Venkataraman, 2000). Startup launches make this process publicly visible by turning entrepreneurial ideas into market-facing product concepts. Collectively, these launches form an idea pool from which competitors and future founders can learn, imitate, and recombine. What matters for future innovation, therefore, is not only how many ventures enter the market, but also how differentiated the ideas they introduce are.

*Worcester Polytechnic Institute. zcheng3@wpi.edu

†Binghamton University, SUNY. kgao13@binghamton.edu

Generative AI has diffused rapidly across knowledge-work tasks and now allows founders to brainstorm, draft, and refine product concepts at low cost (Bick et al., 2024). Yet its implications for idea diversity in real-world markets remain unclear. On the one hand, lower ideation and development costs could enable more people to pursue a broader range of ideas, allowing previously unrealized concepts to reach the market (Kim et al., 2026). On the other hand, founders consulting a small set of general-purpose models—possibly through similar prompts—may be guided toward similar solutions. Previous experimental research finds that AI-assisted ideas cluster more tightly than human-generated ones (Doshi & Hauser, 2024; Girotra et al., 2023; Si et al., 2024). Experiments examine ideas generated within bounded tasks and controlled settings; markets reveal which ideas founders ultimately select, revise, develop, and bring to the public. Founders can reject or substantially transform AI suggestions, while competition creates strong incentives to differentiate from existing products. These market processes may reinforce, attenuate, or even reverse the convergence observed during idea generation. Therefore, whether generative AI homogenizes the ideas that reach the market is an open empirical question.

To examine which of these forces dominates in a real-world entrepreneurial market, we study 291,314 product launches posted between January 2018 and April 2026 across all major Product Hunt categories. We measure how closely launches cluster in idea space by calculating their average conceptual similarity within each category and month. To keep writing style from driving the measure, we paraphrase each launch description before embedding it. A simple comparison before and after the November 30, 2022 release of ChatGPT could also capture changes in platform entry, post-pandemic demand, platform composition, and broader technological trends. We therefore follow the task-exposure logic of Eloundou et al. (2024) and construct an ex-ante LLM exposure index that measures how relevant LLM assistance is to developing product ideas in each category. We then test whether the post-ChatGPT increase in similarity is larger in categories with greater ex-ante exposure, while calendar-month fixed effects absorb changes shared across categories.

We find three patterns. First, product ideas become more concentrated after ChatGPT. Across the fourteen categories with sufficient pre-period observations, the average post-ChatGPT excess in within-category similarity is 0.027, about 15% of the pre-period mean, and twelve categories show a positive increase. The increase persists through April 2026. The distribution also shifts toward common category concepts: the share of distinctive launches falls by more than one-third, from 25.0% before ChatGPT to 16.0% afterward.

Second, the concentration is strongest where LLM assistance is most relevant. A category one standard deviation higher in ex-ante LLM exposure exhibits a 0.018 larger post-ChatGPT increase in within-category similarity, about 10% of the pre-period mean. The distinctive tail follows the same ordering: a one-standard-deviation increase in exposure is associated with a 5.3-percentage-point larger decline in the distinctive share. After ChatGPT, distinctive launches account for 9.1% of launches in the four highest-exposure categories, compared with 19.2% in the eight low-exposure categories.

Third, the changing founder mix does not account for this convergence. First-time teams introduce more distinctive ideas, and their share of launches rises from 46% in 2022 to 71% by 2026, so the compositional shift toward first-time teams pushes aggregate similarity downward. Yet similarity rises within both first-time and experienced teams, the exposure gradient is statistically indistinguishable across the two groups, and holding founder composition at its pre-period level leaves the aggregate gradient essentially unchanged. The convergence therefore extends across founder experience.

One interpretation is that founders increasingly draw on a common, publicly available knowledge source. Such a source introduces a correlated input into otherwise decentralized entrepreneurial search and pulls ideas toward common category concepts. Founders have competitive incentives to differentiate their products, yet the exposure-ordered convergence appears across both first-time and experienced teams. This pattern is consistent with a common input shaping the variety of ideas that entrepreneurial entry brings to the market.

This paper makes three contributions. First, it extends research on generative AI and idea production from bounded experiments to a real product market. Using LLM-assisted text analysis, we distinguish product concepts from writing style across more than 291,000 launches and show that the individual-level gains documented in prior research can coexist with a narrower market-level idea pool. Second, it complements research on entrepreneurial entry by looking beyond the volume of entry to the variety of ideas that entry supplies, and by showing that this variety can narrow even as entry grows (Botelho et al., 2021; Guzman & Stern, 2020; Kerr et al., 2014). Third, it connects entrepreneurial differentiation with the common-information literature by documenting an exposure-ordered rise in similarity across both first-time and experienced teams, even though founders have incentives to distinguish their offerings (Kleinberg & Raghavan, 2021; Morris & Shin, 2002).

2 Literature Review

2.1 Generative AI and Its Impact on Knowledge Work

The first stream of literature this paper builds on asks how generative AI reshapes knowledge work. Generative AI, as a general-purpose technology (Bresnahan & Trajtenberg, 1995; A. Goldfarb et al., 2023), exerts a common input shock felt unevenly across the economy, and empirical studies confirm that this shock reached the workforce rapidly and broadly after the wide adoption of large language models (Bick et al., 2024; Liang et al., 2025). One line of work measures where this input lands, building ex-ante, capability-linked exposure scores at the task and occupation level in the tradition of Autor et al. (2003), with canonical implementations scoring occupations by exposure to LLM capabilities (Eloundou et al., 2024) and to AI more broadly (Felten et al., 2021).

A parallel line of work measures what this input does to worker performance. Generative AI raises

output most for less-experienced workers, compressing the performance distribution (Brynjolfsson et al., 2025; Cui et al., 2026; Dell'Acqua et al., 2023; Noy & Zhang, 2023), though gains can instead accrue to the most skilled when the technology complements expertise (Choudhury et al., 2020; Otis et al., 2023; Toner-Rodgers, 2024).

The evidence most relevant to this paper concerns what this common input does to the ideas themselves. In controlled ideation and creative tasks, AI assistance produces ideas that are individually more novel or higher-rated but collectively more alike (Anderson et al., 2024; Doshi & Hauser, 2024; Girotra et al., 2023; Meincke et al., 2025; Si et al., 2024). The same pattern holds when the object is text rather than concepts: writing with language models narrows the content diversity of what people produce and pulls style toward a common register (Agarwal et al., 2025; Padmakumar & He, 2023; Zhang et al., 2025). Dynamic accounts show that exposure to AI-generated ideas shapes the diversity of the human ideas that follow (Ashkinaze et al., 2025).

Nearly all of this evidence, however, comes from controlled experiments: it documents what AI assistance does to ideas within a lab task or a single writing session. The studies that move beyond the lab stay within a single platform: adopters of text-to-image models on an art-sharing platform produce more work of declining novelty over time (Zhou & Lee, 2024), and generative AI reduces participation and contribution volume in online knowledge communities (Burtch et al., 2024); both document platform behavior rather than the ideas that enter a market. Whether the homogenization observed in individual sessions accumulates in the population of ideas actually brought to market has not been tested at market scale; this paper provides that evidence.

2.2 Innovation and Combinatorial Creativity

The second stream of literature conceptualizes new ideas as novel combinations of existing ones, and innovation as a recombinant search over the space of ideas (Boden, 1998; Weitzman, 1998). Under this view, growth depends on the variety and dispersion of ideas rather than their number alone (Akcigit & Kerr, 2018): the stock of distinct seed ideas bounds the potential for future discovery (Arthur, 2011; Henderson & Clark, 1990; Romer, 1990). The value of recombination is concentrated in the tails: atypical and novel combinations generate the highest-impact science and the most consequential inventions, even as they carry higher outcome variance (Arts & Veugelers, 2015; Fleming, 2001; Kaplan & Vakili, 2015; Uzzi et al., 2013). Novel combinations are also systematically undervalued when they first appear (Wang et al., 2017); the distinctive tail is therefore fragile, and any force that thins it compounds an existing bias against novelty. Our findings bear directly on this margin: after ChatGPT, the distinctive tail of the idea distribution contracts. Dispersion and atypicality are statements about distances between ideas: to say that a combination is atypical, or that the tail is thinning, one needs a metric over the space in which ideas reside (Stirling, 2007).

Operationalizing these quantities requires locating each launch as a point in that space. The conceptual footing comes from the characteristics view of products and the spatial-differentiation tradition (Hotelling, 1929; Lancaster, 1966); text-based methods now make the positioning measurable at

scale (Guzman & Li, 2023; Guzman & Stern, 2020; Hoberg & Phillips, 2016). Our launch-text similarity measure builds on this tradition, with a paraphrase step that keeps shared wording from driving measured similarity.

2.3 Entrepreneurial Strategy and Common Information

The third stream joins two established literatures: entrepreneurial strategy, which studies how entrants choose and defend a market position (Gans et al., 2019), and the economics of common information, which studies how beliefs and choices converge when agents draw on shared signals (Morris & Shin, 2002). In the entrepreneurship literature, entry operates as a portfolio of diverse, experimental bets (Botelho et al., 2021; Kerr et al., 2014), placed by founders acting on idiosyncratic private theories of where value lies (Felin & Zenger, 2017); what the market learns from entry depends on the bets differing. The reward structure sustains the dispersion: optimal-distinctiveness research shows that entrants gain from differentiation even as category norms pressure them to conform (Deephouse, 1999; Taeuscher et al., 2021; Zhao et al., 2017), and the most distinctive, hard-to-categorize entrants form a recognizable tail (Pontikes & Barnett, 2017). In this literature, dispersion is part of what entry supplies.

Common information explains why convergence occurs anyway. In sequential social learning, agents learn from observing earlier movers (Banerjee, 1992; Bikhchandani et al., 1992, 2024; Welch, 1992). Launches in the same category-month appear within a short window rather than a long observed sequence, so imitation of contemporaneous peers is a less natural driver of within-cell similarity than a common input, though the monthly design cannot rule it out. The branch relevant here builds on common signals: when independent agents combine private information with a signal held in common, their posteriors, and with them their choices, move toward a common point. This logic motivates the non-strategic signal-aggregation model in Appendix B (Angeletos & Pavan, 2007; Morris & Shin, 2002). The aggregate cost is the loss of the independence on which crowd wisdom relies (Da & Huang, 2020; Golub & Jackson, 2010; Hong & Page, 2004; Lorenz et al., 2011). Nor is the convergence self-correcting. Information-choice theory implies that agents who want to act differently should seek out *different* information (Hellwig & Veldkamp, 2009); general-purpose models may blunt that response simply by making one broadly common source far more convenient than a diversified set of inputs. At the technology level, this is the concern of algorithmic monoculture: reliance on common AI systems produces correlated outcomes across their users (Bommasani et al., 2022; Kleinberg & Raghavan, 2021; Laufer et al., 2025).

How strongly a common signal moves the market also depends on who is drawing on it. The leveling documented in the knowledge-work literature concerns the quality and quantity of output; whether the same common signal also levels the distinctiveness of what entrants propose is the question our founder-experience analysis takes up. In our setting, Kim et al. (2026) document that generative AI fuels solo entry on Product Hunt, and entrepreneurial entry arrives in waves around salient new technologies (B. Goldfarb & Kirsch, 2020), so who brings ideas to market shifts at the

same time as what founders consult to generate them. These are the two margins our analysis distinguishes.

The intersection of these two literatures gives us both our mechanism and a demanding test of it. Founders have private incentives to resist convergence, yet convergence occurs. The stylized model in Appendix B illustrates the non-strategic common-signal case; a competitive urge to differentiate would push the other way, so the observed rise in similarity is consistent with the common-signal channel outweighing the differentiation motive in this market. The convergence also extends across founder experience: experienced teams shift almost as much as newcomers. Where the monoculture literature documents correlated outputs among users of common AI systems, we trace the narrowing one step downstream, to the composition of ideas that entry supplies a real product market.

3 Data Context

We use Product Hunt as the empirical context for our research questions. Product Hunt (https://www.producthunt.com/) is a public launch platform where entrepreneurs introduce new products and users can upvote them. Each launch page records the product name, a short tagline, a longer description, the launch date, maker accounts, and category labels.

Product Hunt fits our design for three reasons. First, each launch contains a short public pitch. This gives us comparable text for measuring how products are positioned (Guzman & Li, 2023). Second, launches are time-stamped and linked to maker accounts. This lets us study changes around the November 30, 2022 release of ChatGPT and separate first-time teams from experienced teams. Third, the platform spans product categories that differ in how susceptible their ideas are to generative AI, from artificial intelligence and marketing to games.

We collected launches through Product Hunt's official GraphQL API, paginating every category-quarter to exhaustion. Our holdings match the platform's own reported total in every category-quarter. Our analysis uses all published launches: 291,314 product launches in 16 categories[1] between January 2018 and April 2026, ending with the last fully collected month before the data pull. Table 1 reports summary statistics. Entry increases sharply after ChatGPT, especially in artificial intelligence, where launches rise from 6,611 before ChatGPT to 37,447 after.

4 Method

We ask whether generative AI changes the composition of the ideas that reach the market: whether within-category similarity rises after ChatGPT, whether the rise concentrates in categories whose ideas are more susceptible to generative AI, whether it thins the distinctive tail rather than shifting

[1] These 16 categories cover all the major-level project categories curated by Product Hunt, except LLMs and AI Agents, which did not exist on the platform pre-ChatGPT. See https://www.producthunt.com/categories.

Table 1. Summary statistics

	Value
Sample	
Product launches in analysis window	291,314
Categories	16
Date range	Jan 2018 – Apr 2026
Pre-ChatGPT (before Nov 30, 2022)	103,060
Post-ChatGPT	188,254
Launch attributes	
Mean makers per launch	1.27
Mean tagline + description length (characters)	274
Median tagline + description length	271
Mean upvotes per launch	37.1
Median upvotes per launch	6
Mean comments per launch	7.2
Median comments per launch	1
Analysis panel	
Panel cells (category × calendar-month, $n \geq 30$)	1,355
Mean launches per cell	213
Median launches per cell	105
Mean within-cell pairwise cosine, pre-ChatGPT	0.185
Mean within-cell pairwise cosine, post-ChatGPT	0.213

Notes. Upvote counts are cumulative as of the data-collection date and enter only the high-vote subsample robustness check. Within-cell pairwise cosine similarity is our primary measure; its construction is described in Section 4.1.

all launches uniformly, and whether it reaches experienced as well as first-time founders. To answer these questions, we develop a text-based measure of similarity between ideas (Section 4.1) and an ex-ante design variable that captures each category's susceptibility to generative AI (Section 4.2), and we set out a design that separates AI from other post-2022 shocks (Section 4.3). Figure 1 previews the overall workflow, from raw launch records through the idea-space measures and design variables to the empirical tests that follow.

4.1 Within-Category Idea Similarity and Effective Variety

Our primary measure captures how similar product ideas are within a category-month. To compare ideas across launches, we place each launch as a point in a common idea space and read similarity off the distances between points. Because launches are described in text, we follow research that uses embeddings of text to measure how firms and products are positioned and how novel they are (Guzman & Li, 2023; Hoberg & Phillips, 2016; Kaplan & Vakili, 2015). An embedding model is a neural network that maps a passage of text to a point in this shared space, trained so that passages with similar meaning land near one another; the closeness of two launches is then the cosine of their vectors, near one when they express the same idea and near zero when unrelated.

Text, however, carries more than the idea it describes. It also carries the way that idea is presented (e.g., tone, buzzwords, phrasing), and an embedding of the raw text blends style with substance: different wording can push two launches with the same idea apart, while shared marketing language

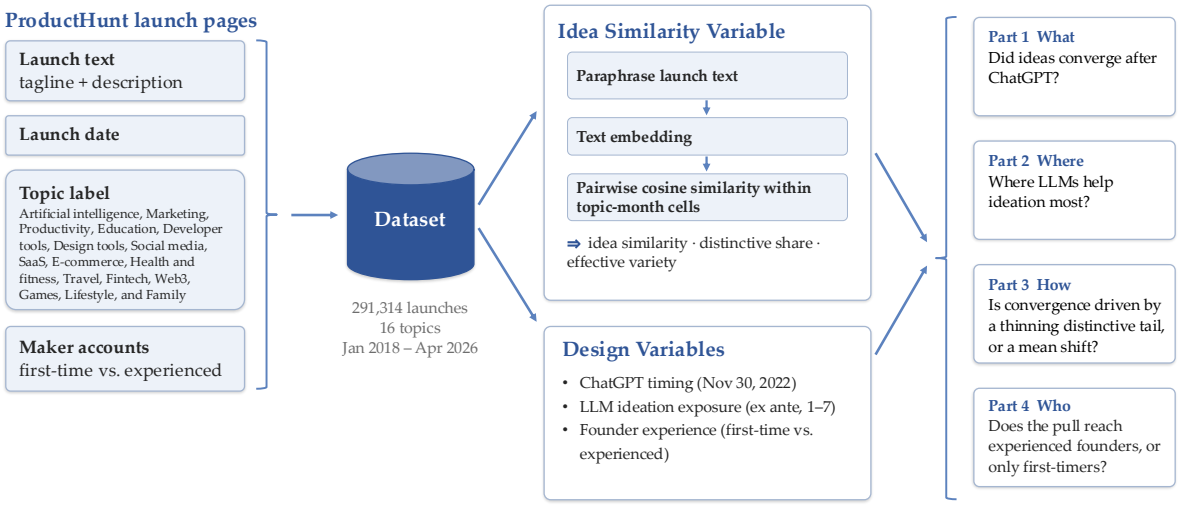


Figure 1. Overall research workflow. Product Hunt launch pages supply launch text, dates, category labels, and maker accounts for 291,314 launches in 16 categories from January 2018 to April 2026. We paraphrase each launch's text with an LLM, embed it, and compute its pairwise cosine similarities with peers launched in the same category and month; these similarities yield idea similarity, distinctive share, and effective variety. We then relate these outcomes to ChatGPT timing, each category's ex-ante LLM exposure, and founder experience to answer the four questions shown on the right.

can pull two different ideas together. This matters in our setting because generative AI may change how founders write as much as what they build, so a raw-text measure would confound the two. To isolate the idea, we first pass every launch through a single large language model that strips author-specific style while preserving the product it describes, and then embed the rewritten text, so that the remaining similarity reflects the idea rather than the prose. Figure 2 shows one launch pair in each direction.

To operationalize this, we use a locally served Qwen3-32B model to paraphrase each launch's concatenated tagline and description, under a prompt that preserves the product's meaning and main features while rewriting the prose at comparable length, and encode each paraphrase with all-MiniLM-L6-v2 (Reimers & Gurevych, 2019) as an L2-normalized 384-dimensional vector x_i. The similarity between launches i and j is the cosine $x_i^\top x_j \in [-1, 1]$. Within each (c, t) cell, where c indexes category and t indexes calendar month, the dependent variable $\text{sim}_{c,t}$ is the mean cosine over the $\binom{n_{c,t}}{2}$ distinct launch pairs; with unit-norm embeddings, a higher cell mean also means launches are more concentrated around the cell centroid. We retain category-months with at least 30 launches (roughly one launch per day over a calendar month), an activity floor that limits noise from sparse peer sets.[2] Several checks confirm that the paraphrase preserves the product concept:

[2]The primary mean-similarity result is not sensitive to this threshold. Re-estimating equation (1) with a minimum

each paraphrase stays close to its original text (mean cosine 0.90), near-duplicate launches remain near-duplicates, and blind ratings of pairwise closeness align with the cosine measure (Spearman's $\rho = 0.73$). Appendix A.2 reports the prompt and all three validation checks. We also reestimate the main specifications using embeddings of the original, un-paraphrased text (Section 5.5).[3]

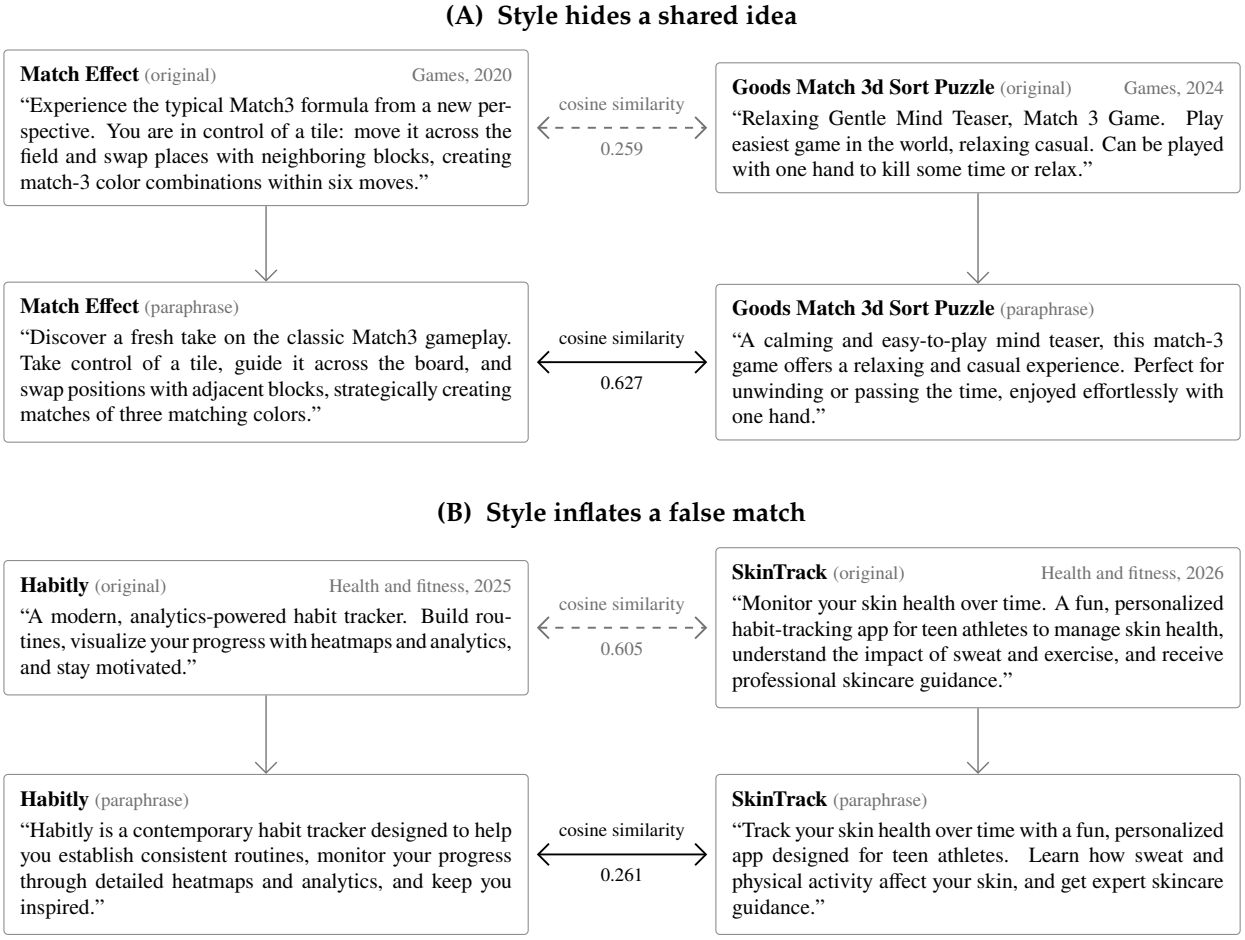


Figure 2. What paraphrasing changes. Each panel shows two launches, their original texts and paraphrases, and the pair's cosine similarity computed from embeddings of each version. Dashed arrows compare the original texts; solid arrows compare the paraphrases. Original texts are lightly edited for length.

Auxiliary Measures. Two auxiliary measures corroborate the primary mean-cosine result by summarizing different features of the same within-category geometry.

Distinctive Share. This measure tracks the low-similarity tail. We give each launch its average cosine to the other launches in its category-month, and call a launch *distinctive* when that value falls in the bottom quartile of its own category's pre-ChatGPT launches; a cell's distinctive share is the

of 20 launches per cell (1,410 cells) gives $\beta = +0.0223$ (cluster-robust SE 0.0052, wild-cluster bootstrap $p = 0.013$); a minimum of 50 launches (1,205 cells) gives $\beta = +0.0202$ (SE 0.0047, bootstrap $p = 0.029$).

[3]The exposure gradient in Section 5.2 also holds with BGE-large-en-v1.5, a structurally different 1024-dimensional sentence encoder (Section 5.5).

fraction of its launches below that cutoff.[4] By construction about one launch in four is distinctive in each category before ChatGPT, so post-ChatGPT movement in the share is what carries the information.

Effective Idea Variety. We use the order-1 Vendi score (Friedman & Dieng, 2022), a metric commonly used in machine learning that counts the number of distinct elements in a collection while accounting for the pairwise similarities among them. Intuitively, the Vendi score characterizes how many mutually dissimilar launches would carry the same variety as a cohort of k randomly selected launches from the category-month.[5] Following the rarefaction principle for comparing diversity across samples of unequal size (Gotelli & Colwell, 2001), we remove the 151 of 1,506 category-month cells that average fewer than one launch per day, and average the score over random cohorts of size $k = 30$; this keeps the one-to-thirty scale constant as launch volume changes. We report the construction details in Online Appendix OA2.1 and, in Appendix A.3, the estimated results using this auxiliary measure as a validity check on our main results.

4.2 LLM Exposure Index

This index captures how susceptible a category's ideas are to generative AI. We adapt the task-exposure logic of Eloundou et al. (2024) from occupations to Product Hunt categories. Each of the 16 categories receives an integer score from 1 to 7. A category receives a high score when typical product concepts can be developed from general language and common knowledge. A category receives a low score when its product concepts depend more on tacit craft, specialized infrastructure, or regulatory expertise. For example, artificial intelligence receives 7 because LLMs are central to the category. Games receives 2 because product concepts rely heavily on design craft. Fintech receives 2 because many concepts depend on compliance and domain-specific constraints. Appendix Table A1 reports all scores and the rating prompt.

We construct the index as the median score across fifteen independent rater passes, five each from OpenAI gpt-4o, Google gemini-2.5-pro, and Anthropic claude-sonnet-4-6 under the same prompt, and treat it as an ex-ante measure. The passes agree closely within each model and rank-correlate across the three families (Spearman 0.83–0.93), so the index does not hinge on any single model. As a human benchmark, two raters independently scored all 16 categories from the same rubric, and their ratings align with the index.[6]

[4] We fix each category's cutoff from its retained pre-ChatGPT months, weighting every pre-period category-month equally (each launch weight $1/n_{c,t}$) so that high-volume months do not dominate, and hold it fixed afterward. Family and lifestyle have no pre-period month meeting the 30-launch minimum, so the distinctive-tail outcome covers fourteen categories.

[5] The order-1 Vendi score of a 30-launch cohort A is $V_1(A) = \exp(-\sum_k \lambda_{Ak} \log \lambda_{Ak})$, where the λ_{Ak} are the eigenvalues of the cosine Gram matrix S_A normalized by its trace $\mathrm{tr}(S_A)$; we average V_1 over 200 random 30-launch cohorts drawn with fixed cell-specific seeds. A score of m means the cohort holds the variety, under the cosine kernel, of m mutually orthogonal launches, an effective count rather than a count of labeled idea types.

[6] Measurement error in Exposure_c would attenuate the exposure gradient, so rater noise works against the result.

4.3 Empirical Methodology

We estimate the following panel model:

$$\text{sim}_{c,t} = \alpha_c + \gamma_t + \beta \cdot \left(\text{PostChatGPT}_t \times \text{Exposure}_c\right) + \varepsilon_{c,t}, \tag{1}$$

where PostChatGPT_t equals one for calendar months after November 30, 2022, and Exposure_c is the standardized exposure index (one standard deviation corresponds to about 1.6 points on the 1–7 scale). The specification includes category fixed effects (α_c), which control for time-invariant differences in baseline similarity across categories, and calendar-month fixed effects (γ_t), which control for shocks common to all categories in a given month, including entry waves, post-COVID demand shifts, platform redesigns, and the average effect of ChatGPT itself. The coefficient of interest, β, measures whether the post-ChatGPT change in within-category similarity is larger in categories with greater ex-ante LLM exposure.

Interpreting β as evidence that generative AI shapes the ideas that reach the market requires a differential parallel-trends assumption: absent the diffusion of generative AI, categories with different levels of ex-ante LLM exposure would not have experienced systematically different changes in idea similarity around ChatGPT's release. We assess the plausibility of this assumption using pre-treatment trends. We find no evidence of a systematic positive differential pre-trend (Appendix C.1).

The baseline specification gives each category-month equal weight.[7] We cluster standard errors by category. There are only 16 clusters, so we also report wild-cluster bootstrap p-values for headline estimates, using Webb six-point weights with 9,999 replications (Cameron et al., 2008). The baseline estimate has a wild-cluster bootstrap p-value of 0.010.

The exposure index is an ex-ante measure of a category's susceptibility to generative AI, not an experimentally assigned treatment. We therefore interpret the estimates as evidence that generative AI shapes the ideas that reach the market, rather than as a randomized treatment effect. The sample covers Product Hunt's major categories, not all startups; founders who do not launch publicly are outside the sample frame.

5 Findings

Our analyses proceed as a sequence of increasingly demanding tests of the proposed mechanism. We first establish whether startup ideas become more similar after ChatGPT. We then use cross-category differences in ex-ante LLM exposure to assess whether the shift is strongest where a category's ideas are most susceptible to AI. Next, we examine the shape of the idea distribution to determine

[7] The gradient does not depend on how cells are weighted. Weighting cells by launch count gives $\beta = +0.0164$ (cluster SE 0.0039, $t = 4.20$), about 89% of the unweighted estimate; weighting by log launch count gives +0.0181 ($t = 4.52$), and down-weighting large cells by $1/n$ gives +0.0203 ($t = 4.00$).

whether convergence reflects a broad movement across all ideas or a center-pulling decline in distinctive entry. Finally, we compare first-time and experienced teams.

5.1 Within-Category Ideas Become More Similar

Within-category idea similarity rises after ChatGPT (Figure 3). For each category, we fit a linear trend to its pre-ChatGPT monthly similarity from January 2018 to November 2022, extrapolate the trend into the post period, and calculate the average difference between observed and predicted similarity. Across the 14 categories with sufficient pre-period observations, this detrended excess is +0.027 in mean-cosine units, about 15% of the pre-period mean of 0.185. This positive excess is robust across categories, holding true for 12 out of the 14 categories analyzed (Table 2). The increase persists through April 2026, more than three years after ChatGPT's release. Its duration is difficult to reconcile with a brief release-period entry wave alone and indicates that the convergence is not confined to the initial diffusion of the technology. The excess is a descriptive summary; the identification rests on the panel estimates in Section 5.2, which do not use it.

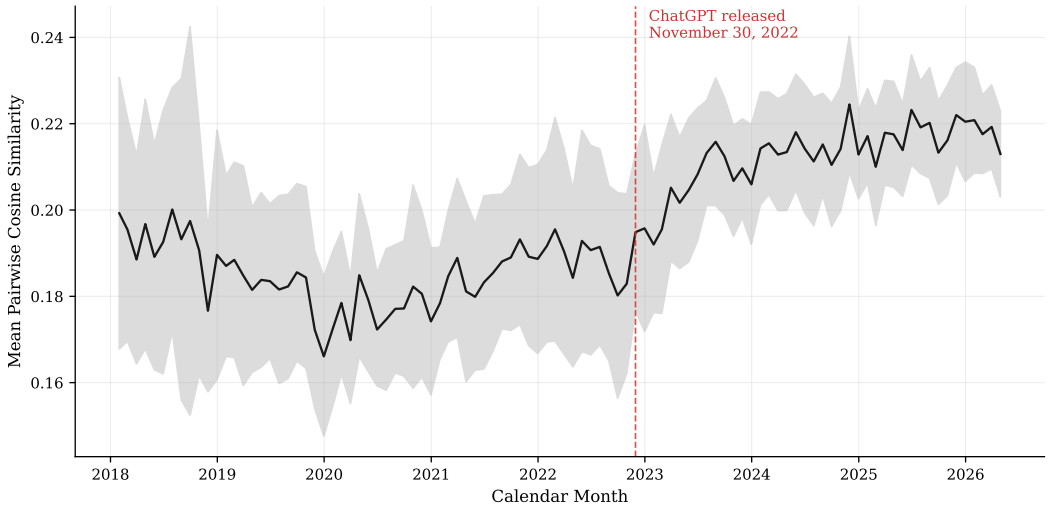


Figure 3. Aggregate within-category idea similarity by calendar month. Cross-category mean of within-category mean pairwise cosine similarity, January 2018–April 2026. The shaded region is a pointwise 95% band based on the cross-category standard error; the dashed line marks the November 30, 2022 release of ChatGPT.

Table 3 shows what the launch-level measure captures, using selected launches from categories outside artificial intelligence; each launch is scored against its own category's same-month peers. The high-similarity launches never mention AI, yet they read like the same pitch written for different domains—an all-in-one travel companion, an all-in-one conversion booster, an all-in-one habit tracker: each sits squarely on its category's modal concept. The low-similarity launches address problems far from that modal concept, from crowdsourced pint prices to a plastic-free toothpaste tube. The examples are selected for illustration; they show the contrast that the aggregate measure summarizes.

Table 2. Per-category post-ChatGPT excess in within-category similarity

Category	Exposure index	n_{pre} cells	n_{post} cells	Detrended excess (mean, z-score)
Higher exposure (index ≥ 5)				
Artificial intelligence	7	59	41	+0.0705 (36.7)
Marketing	6	59	41	+0.0897 (29.3)
Productivity	6	59	41	+0.0486 (36.5)
Education	6	59	41	+0.0236 (12.9)
Developer tools	5	59	41	+0.0391 (32.6)
SaaS	5	47	41	+0.0160 (9.7)
Lower exposure (index ≤ 4)				
E-commerce	4	51	41	−0.0036 (−1.0)
Travel	4	37	40	+0.0058 (2.0)
Social media	4	49	41	+0.0083 (3.9)
Design tools	3	59	41	+0.0094 (4.0)
Health and fitness	3	59	41	+0.0313 (11.5)
Fintech	2	59	41	+0.0244 (7.2)
Web3	2	47	40	+0.0147 (4.0)
Games	2	59	41	−0.0038 (−1.8)
Aggregate (mean across 14 categories)	–	–	–	+0.0267
Spearman ρ (excess vs. exposure)	–	–	–	0.61

Notes. For each category, we fit a linear trend to pre-period (January 2018–November 2022) cell-level mean similarity and define excess as the post-period mean of observed similarity minus the extrapolated trend. The z-score divides this excess by the category's pre-period residual standard deviation. The table includes the 14 categories with at least 18 pre-period cells containing at least 30 launches; lifestyle and family fall below that threshold.

Table 3. Selected post-ChatGPT launches in non-AI categories: modal versus distinctive ideas

Product	Category	Tagline	Cell (month, n)	$\bar{s}$
Modal (high within-cell similarity)				
GoTrip	Travel	Plan, share, travel: your all-in-one adventure companion	2024-08 (n=56)	0.364
SalesBuddy	E-commerce	Turn visitors into buyers with realtime offers, social proof	2025-01 (n=113)	0.362
CompleteDay	Health and fitness	Habit tracker with nutrition, exercise, and fasting tracker	2024-04 (n=122)	0.345
Prettify	Design tools	Create stunning visuals instantly	2024-10 (n=152)	0.338
Distinctive (low within-cell similarity)				
Compare Country Size	Education	Compare country sizes on one map	2025-06 (n=404)	0.010
PAIST	E-commerce	100% plastic-free toothpaste tube, made by dentists	2023-11 (n=86)	0.014
PintWatch Ireland	Social media	Crowdsourced pint prices across Irish pubs	2026-04 (n=391)	0.018
Inkle's Free Form 7004	Fintech	Extend a US startup's IRS tax deadline by six months	2023-03 (n=111)	0.020

Notes. Rows are selected illustrations; modal rows come from the high end of each category's within-category similarity distribution and distinctive rows from the low end. $\bar{s}$ is a launch's average cosine similarity to same-category, same-month peers, computed from the original (un-paraphrased) text so the displayed taglines correspond to the scores; the primary measure uses paraphrased text (Section 4.1). Taglines are lightly edited for length.

5.2 Similarity Rises More in Categories with Higher LLM Exposure

Table 4 reports the panel difference-in-differences estimates. In the baseline specification (column 1), a category one standard deviation higher in ex-ante LLM exposure is associated with a 0.018 larger post-ChatGPT increase in mean within-category cosine similarity (about half a standard deviation of monthly within-category similarity), relative to contemporaneous changes at lower exposure levels; the estimate is statistically significant ($p = 0.010$ under a wild-cluster bootstrap) and is about 10% of the pre-period mean. Columns 2 and 3 add an exposure-specific linear trend and report leave-one-out sensitivity.

Table 4. Cross-category gradient: panel difference-in-differences on within-cell similarity

	(1) Monthly baseline	(2) Monthly + trend ctrl	(3) Leave-one-out range
$\text{PostChatGPT}_t \times \text{Exposure}_c$	+0.0184	+0.0178	[+0.015, +0.020]
	(0.0041)	(0.0044)	
$t \times \text{Exposure}_c$ (linear trend)		+0.0001	
		(0.0011)	
Wild-cluster bootstrap p	0.010	0.013	—
Category fixed effects	Yes	Yes	Yes
Calendar-month fixed effects	Yes	Yes	Yes
Cluster-robust SE on category	Yes	Yes	Yes
Observations	1,355	1,355	1,255–1,350
R^2	0.782	0.782	0.78–0.79
Adjusted R^2	0.762	0.762	0.76–0.77
Dep. var. mean (pre / post)	0.185 / 0.213	0.185 / 0.213	—

Notes. The dependent variable is the primary similarity measure at the category-month level (Section 4.1). Column 1 estimates equation (1) on the monthly panel (1,355 cells, 16 categories); column 2 adds a linear time trend interacted with the exposure index; column 3 reports the range of leave-one-out estimates, dropping each category in turn. Standard errors in parentheses are cluster-robust at the category level (CR1, 16 clusters). Wild-cluster bootstrap p-values impose the null and use Webb six-point weights with 9,999 replications (Cameron et al., 2008; Webb, 2023). A permutation test reassigning the exposure index across the 16 categories (Young, 2019) (9,999 permutations, two-sided) gives $p \leq 0.001$ in columns 1 and 2. Exposure_c is the standardized category-level LLM exposure index (Appendix Table A1); PostChatGPT_t indicates calendar months after November 30, 2022.

Additionally, we use effective idea variety, which counts how many distinct ideas a cohort of launches effectively contains, as a complementary, count-based view of the exposure gradient. Under the same specification as equation (1), a category one standard deviation higher in LLM exposure is associated with a 0.892-unit more negative post-ChatGPT change in effective variety, relative to contemporaneous changes at lower exposure levels ($\hat{\beta} = -0.892$, SE = 0.200; Webb wild cluster bootstrap $p = 0.0083$; permutation $p = 0.0055$; Appendix Table A2). For a cohort of 30 launches, this is roughly one fewer effective idea in a category one standard deviation more exposed.

For the primary mean-cosine outcome, the exposure gradient appears as a level break after ChatGPT.

Adding an exposure-specific linear trend gives $\hat{\beta} = +0.0178$ (wild-cluster bootstrap $p = 0.013$), while the trend interaction is small and statistically insignificant (Table 4, column 2). Leave-one-out estimates range from +0.015 when web3 is omitted to +0.020 when health and fitness is omitted. Every estimate remains significant at CR1 $p < 0.01$.

Figure 4 maps this cross-category gradient by plotting each category's detrended post-ChatGPT excess similarity against its ex-ante LLM exposure index, with the underlying values detailed in Table 2. The most pronounced increases in within-category similarity are concentrated among the four highest-exposure categories (index ≥ 6). Relative to their pre-period similarity mean of 0.158, the post-treatment excesses range from 15% in education to 57% in marketing. Conversely, games—among the lowest-exposure categories—exhibits a slightly negative excess (−0.004), while travel and e-commerce also stay near zero. The non-parametric Spearman rank correlation between the per-category excess and the exposure index is $\rho = 0.61$ ($p = 0.021$ via a two-sided permutation test), consistent with a monotonic cross-category gradient.

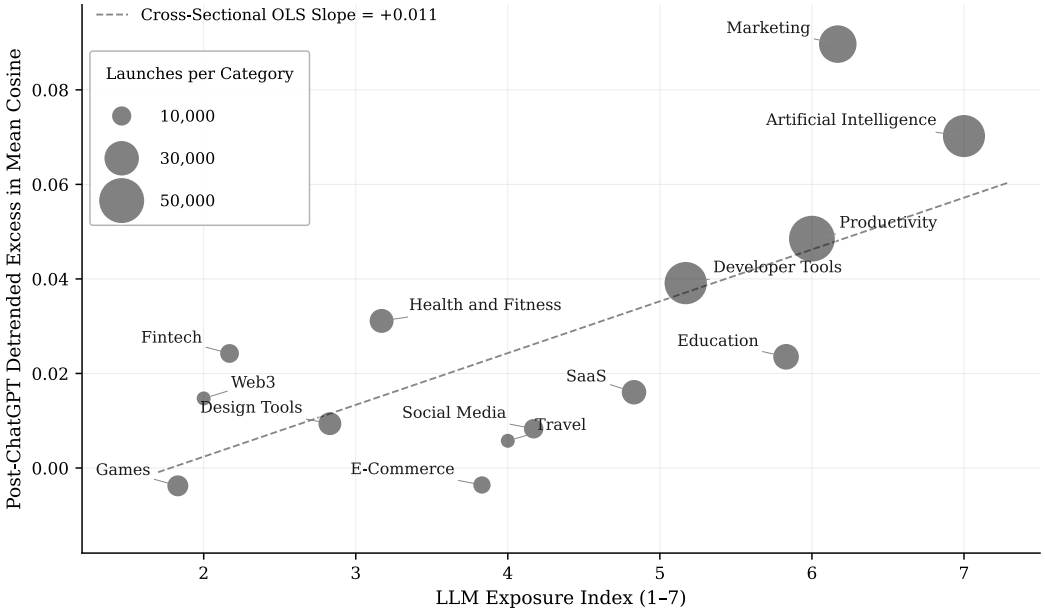


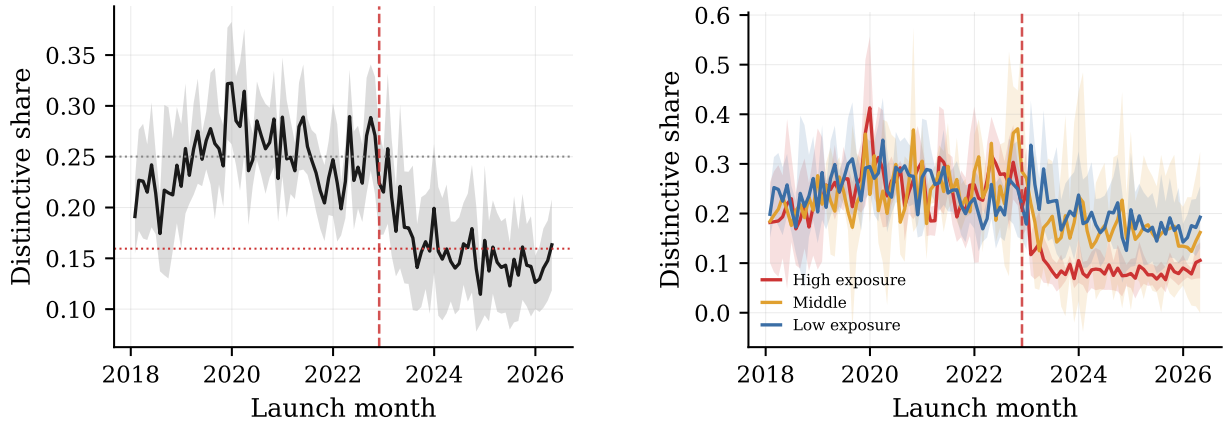
Figure 4. Post-ChatGPT change in within-category similarity by LLM exposure. Each bubble is a category. The vertical axis is the category's detrended post-ChatGPT excess in mean within-category cosine; the horizontal axis is the ex-ante exposure index; bubble area is proportional to launch count. The dashed grey line is a descriptive OLS fit through the 14 category-level points (slope +0.011). Categories with the same exposure score are horizontally jittered for legibility.

Interestingly, fintech and web3—two low-exposure categories—show noticeable positive excesses (+0.024 and +0.015, respectively), running against the direction of the gradient. Because these categories score low on exposure, their positive excesses flatten rather than steepen the cross-category gradient, so if anything they make the baseline estimate conservative. We speculate that the late-2022 cryptocurrency market crash contributed to the anomaly: both sectors contracted around the treatment window, which could have removed crypto-adjacent products and lowered

the within-category baseline against which the detrended excess is measured. We do not test this channel directly and offer it only as a possible reading of these two off-gradient categories.

The cross-category gradient is consistent with generative AI shaping the ideas that reach the market. Calendar-month fixed effects absorb shocks common to all categories, so the estimate comes from differential changes across ex-ante LLM exposure levels. Similarity rises more where a category's ideas are more susceptible to AI. This ordering is consistent with product ideas moving toward common category concepts after ChatGPT. The panel gradient does not by itself reveal whether this convergence comes from a changing mix of product types or from convergence within types. Descriptive evidence from the four highest-exposure categories indicates that both margins contribute, though it does not cleanly separate their shares (Appendix C.2).

5.3 The Distinctive Tail Thins Most in LLM-Exposed Categories



(A) Distinctive-launch share, 14 categories with baselines

(B) Distinctive share by exposure group

Figure 5. Distinctive-launch share before and after ChatGPT. A launch is distinctive when its average cosine to same-category, same-month peers falls below its category's pre-ChatGPT bottom-quartile cutoff. **(A)** Equal-category mean across the 14 categories with pre-period baselines. **(B)** Means for exposure groups fixed by the index: high (≥ 6, four categories), middle (= 5, two categories), and low (≤ 4, eight categories). Family and lifestyle lack a qualifying pre-period baseline and are excluded. Shaded regions are pointwise 95% bands from the cross-category standard error; dashed lines mark November 30, 2022.

Innovation returns are highly skewed: a small number of breakthroughs account for most of the value (Kogan et al., 2017; Scherer & Harhoff, 2000), and atypical combinations are more likely to produce high-impact science and inventions (Arts & Veugelers, 2015; Uzzi et al., 2013). In recombinant accounts of innovation, a thicker distinctive tail widens the set of inputs available for future discovery (Romer, 1990; Weitzman, 1998). Cheaper ideation opens two opposing possibilities. It may let founders explore concepts they would otherwise have left untouched, or it may draw founders who lean on the same models toward the same familiar concepts. This section asks which distributional pattern appears among products that reach the market.

As defined in Section 4.1, a launch is distinctive when its average similarity to same-category, same-month peers falls in the bottom quartile of its category's pre-ChatGPT launches. Figure 5, Panel A, plots that share across the fourteen categories with a pre-ChatGPT baseline. The share falls from 25.0% before ChatGPT to 16.0% after, a decline of more than a third. On the figure's equal-category scale, the rate moves from roughly one in four to one in six. Ideas far from a category's familiar concepts appear less often in the flow of new products.

Panel B asks whether the tail thins evenly or more where LLM exposure is higher. Before ChatGPT, the high-, middle-, and low-exposure groups all track around a quarter, as the category-specific pre-period cutoffs imply. After the release all three groups decline, ordered by the index: the distinctive share falls to 9.1% in the four high-exposure categories and to 16.9% in the two middle categories, while the eight low-exposure categories fall only to 19.2%. This ordering is the distinctive-tail counterpart of the exposure gradient established in Section 5.2.

We put that ordering into a single coefficient by re-estimating Section 5.2's difference-in-differences specification with the distinctive share as the outcome. A one-standard-deviation increase in LLM exposure is associated with a 5.3-percentage-point larger post-ChatGPT decline in the distinctive share ($\hat{\beta} = -0.0532$, cluster-robust SE 0.0132, wild-cluster bootstrap $p = 0.0102$, permutation $p = 0.0052$), about a fifth of the aggregate 25.0% pre-period share. This regression records the corresponding gradient in the distinctive share under the comparison developed in Section 5.2.

Post-ChatGPT cells are also much larger, allowing each launch's average similarity to its peers more precisely, but this could move the share below a fixed cutoff on its own. To hold the comparison set constant, we draw one reference set of 30 launches per category-month and score every launch against that cell's common reference set in both periods. This fixed-reference measure calibrates its own equal-cell-weighted pre-period cutoff. Its distinctive share falls from 25.0% to 16.2%, and the exposure coefficient is essentially unchanged at -0.0545. The changing number of peers used to construct a launch's score therefore does not account for the thinning gradient. The quartile itself is not doing the work either: bottom-decile, bottom-quintile, and bottom-third cutoffs give the same significant contraction (Online Appendix OA2.3).

The pattern survives without reducing the distribution to the quartile indicator. Figure 6 plots the full launch-level densities for the four highest-exposure categories and the four lowest observed in both periods. In the high-exposure row, equal-cell-weighted mean similarity rises from 0.158 to 0.209, the post-ChatGPT density moves toward higher similarity, and the low-similarity shoulder where distinctive launches sit thins. The low-exposure row is comparatively stable, with smaller and mixed changes across categories. The dotted line in each panel marks the category-specific distinctive cutoff. Online Appendix OA2.2 shows all sixteen categories.

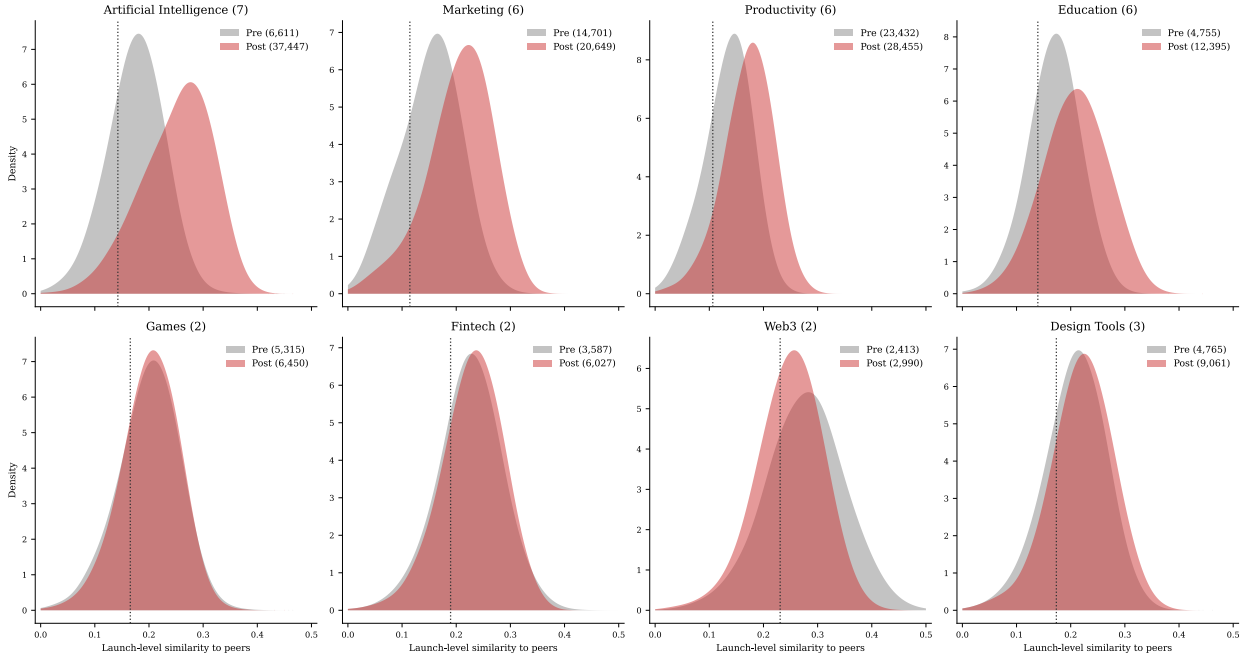


Figure 6. Launch-level similarity before and after ChatGPT by category. The top row contains the four categories with exposure scores of 6 or 7; the bottom row contains four selected low-exposure categories observed in both periods. Curves average cell-level distributions with equal weight per cell; exposure scores appear in parentheses and launch counts in the legend. Dotted lines mark category-specific pre-ChatGPT bottom-quartile cutoffs. Because the curves are kernel-smoothed, a cutoff need not leave exactly one quarter of the plotted pre-period density to its left.

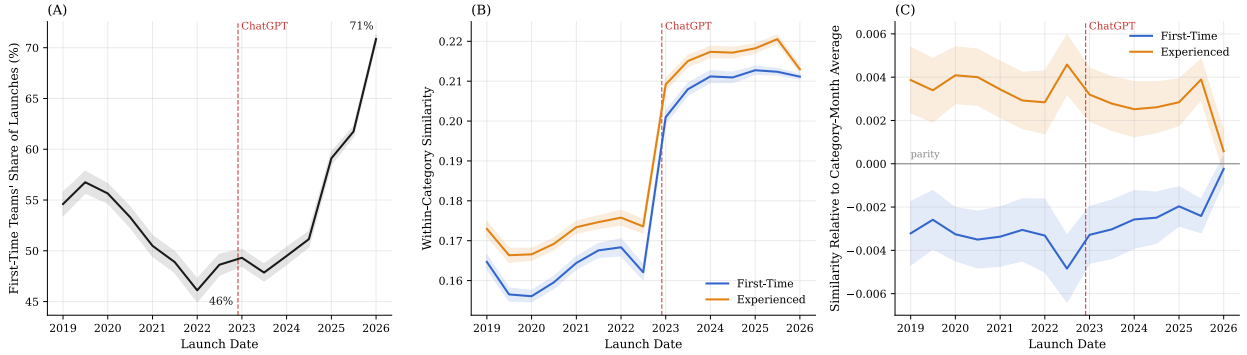


Figure 7. Founder experience and within-category similarity. **(A)** First-time teams' share of maker-classified launches by half-year since 2019; the series uses cells with at least 30 launches, with mixed-experience teams coded as experienced. **(B)** Raw within-category similarity for first-time and experienced teams. **(C)** Similarity for each group relative to the category-month average. All panels use half-year bins.

5.4 Similarity Rises Across First-Time and Experienced Teams

We next ask how the changing mix of founders (i.e., first-time vs. experienced founders) and changes in similarity within each founder type jointly shape the market-wide rise. We classify a launch as *first-time* when no maker on the team has a previously observed Product Hunt launch and as *experienced* otherwise.

Our model-free evidence shows that, after ChatGPT, similarity rises for both first-time and experienced teams, from around 0.16–0.17 to above 0.21 (Figure 7, Panel B). At the same time, the founder mix shifts sharply toward newcomers: first-time teams' share of maker-classified launches rose from 46% in 2022 to 71% by 2026 (Panel A). First-time teams remain more distinctive; their point-estimate similarity gap relative to experienced teams narrows from about 0.009 to 0.005 (about 0.14 to 0.08 standard deviations; Panel C). The shift toward the more distinctive group pulls aggregate similarity downward, but the rise within both groups more than offsets this compositional pull, so market-wide similarity rises.

The exposure gradient is statistically indistinguishable across founder types. A direct test does not reject equal gradients for the two groups; reweighting each category to hold its first-time share at its pre-period level leaves the gradient essentially unchanged; and within each founder type, the gradient remains positive (Online Appendix OA3.3). Launches grow more alike across both experience groups despite founders' incentives to differentiate, a pattern consistent with a common, publicly available knowledge source shaping the ideas that reach the market across experience levels.

5.5 Robustness Checks

We conduct several robustness checks to ensure that the documented cross-category exposure gradient is not driven by measurement artifacts, encoder choice, or alternative sample constructions.

First, to check whether the result reflects convergence in product ideas rather than convergence in writing style, we re-estimate the baseline model using the original, un-paraphrased launch text. The exposure gradient is nearly unchanged ($\hat{\beta}$ = +0.0177, SE = 0.0042). The aggregate level shift and the tail decline also retain about 90% of their magnitudes after paraphrasing, suggesting that the documented convergence is mainly about product concepts rather than shared wording.

Second, to check whether the result depends on the sentence encoder, we re-embed the full corpus using BGE-large-en-v1.5, a 1024-dimensional open-source embedding model with a different architecture and training corpus. The exposure gradient remains positive and statistically significant ($\hat{\beta}$ = +0.0120, SE = 0.0027). The estimate is smaller than the baseline but points to the same conclusion: the cross-category gradient is not specific to the primary encoder.

Third, to ensure the exposure gradient is not driven by low-attention launches in the long tail of

Product Hunt, we re-estimate equation (1) keeping only the more-upvoted half of each category-month cell. The coefficient is $\hat{\beta} = +0.0144$ ($t = 3.62$, wild-cluster bootstrap $p = 0.026$), so the result is not an artifact of low-visibility launches or of including the full long tail of posted products.

6 Discussion

This paper examines whether generative AI broadens or narrows the product ideas that entrepreneurs bring to market. We study 291,314 Product Hunt launches from January 2018 to April 2026, measure within-category idea similarity from launch text paraphrased to standardize writing style, and compare changes around ChatGPT's release across categories with different levels of LLM exposure. Our results show that product ideas launched within the same category become more similar after ChatGPT. The share of distinctive launches falls by more than one-third, and the contraction is larger in categories with greater ex-ante LLM exposure. The gradient remains when we hold founder composition fixed and estimate similarity within first-time and experienced founder groups.

The exposure index orders categories by how effectively a 2022-era general-purpose LLM could help an entrepreneur develop a typical product idea. We fixed the rating rubric and category scores before estimating the outcome gradients, so the exposure ordering was specified separately from the outcome analysis. It provides an ex-ante category-level test of whether ideas change more where LLM assistance is more relevant. The outcome places each launch in an idea space defined by the problem, intended user, and function presented to the market, after paraphrasing standardizes writing style. This construction lets us compare the distribution of ideas reaching the market across categories and over time. We do not observe founders' ideation processes or their actual use of LLMs. Accordingly, the exposure index captures the potential relevance of LLM assistance at the category level, whereas the outcome captures only the ideas ultimately presented to the market.

One possible interpretation is that founders draw on a common, publicly available knowledge source. The stylized model illustrates how ideas can become more concentrated when a common input is precise enough to receive substantial weight and shared enough to move founders in correlated directions (Appendix B). The cross-category ordering and placebo estimates weigh against a generic post-2022 shock. The gradient among launches that never mention AI weighs against a surge confined to explicitly AI products. Its persistence within both founder groups and after composition reweighting weighs against a changing founder mix. These tests are consistent with the common-input interpretation. They do not establish that individual founders used an LLM or identify the mechanism.

A change in technological opportunities can produce the same aggregate pattern. ChatGPT may have raised the returns to a focal set of applications in high-exposure categories, leading founders toward those applications without direct model assistance. The thinner distinctive tail does not

separate the explanations. A broad expansion of the feasible set could support more distinctive products, but an opportunity shift that raises returns to a narrow set could reduce the tail. We also cannot rule out imitation of visible launches or locate a common input in any particular layer of the model-to-launch pipeline. The common-input mechanism is therefore one explanation consistent with the evidence. The research design does not identify it separately from the opportunity response.

These distinctions make the implication for platforms conditional. If concentration reflects an accurate response to a focal opportunity, interventions that disperse entry could move the portfolio away from the applications with the highest expected returns. If correlated reliance on common inputs matters, tools that are useful to individual founders may nevertheless reduce variety in the incoming portfolio. For platforms that value breadth and discovery, the results motivate monitoring the distribution of ideas and testing portfolio-level responses, such as crowding feedback at submission or diversity-aware discovery. Because the gradient appears within both founder groups, changing the founder mix would not address the within-group component. Under the common-input interpretation, founders may gain more differentiation from private information, domain expertise, and proprietary data.

This study has two limitations that define the scope of the evidence. First, the data come from one English-language, software-heavy platform and cover product concepts that founders choose to launch publicly. Second, because we do not observe founders' actual use of LLMs, the evidence cannot distinguish direct LLM assistance from contemporaneous changes in category opportunities. Founder-level use data or randomized access to LLM assistance could help separate these explanations. Experiments that independently vary the informativeness and overlap of model outputs could distinguish the precision and sharedness channels. Future work could link launch ideas to downstream outcomes to test whether a thinner distinctive tail predicts product success or limits future recombination. More broadly, the pattern points to a tension between individual capability and collective variety: tools that expand what each entrepreneur can produce may also concentrate the ideas that reach the market. The entrepreneurial consequences of generative AI therefore encompass individual productivity and the composition of the market-wide idea pool. The breadth of that pool shapes the possibilities available for future recombination and discovery.

References

Agarwal, D., Naaman, M., & Vashistha, A. (2025). AI Suggestions Homogenize Writing Toward Western Styles and Diminish Cultural Nuances. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–21. https://doi.org/10.1145/3706598.3713564

Akcigit, U., & Kerr, W. R. (2018). Growth through Heterogeneous Innovations. *Journal of Political Economy*, *126*(4), 1374–1443. https://doi.org/10.1086/697901

Anderson, B. R., Shah, J. H., & Kreminski, M. (2024). Homogenization Effects of Large Language Models on Human Creative Ideation. *Creativity and Cognition*, 413–425. https://doi.org/10.1145/3635636.3656204

Angeletos, G.-M., & Pavan, A. (2007). Efficient Use of Information and Social Value of Information. *Econometrica*, *75*(4), 1103–1142. https://doi.org/10.1111/j.1468-0262.2007.00783.x

Arthur, W. B. (2011). *The Nature of Technology: What It Is and How It Evolves*. Free Press.

Arts, S., & Veugelers, R. (2015). Technology familiarity, recombinant novelty, and breakthrough invention. *Industrial and Corporate Change*, *24*(6), 1215–1246. https://doi.org/10.1093/icc/dtu029

Ashkinaze, J., Mendelsohn, J., Qiwei, L., Budak, C., & Gilbert, E. (2025). How AI Ideas Affect the Creativity, Diversity, and Evolution of Human Ideas: Evidence From a Large, Dynamic Experiment. *Proceedings of the ACM Collective Intelligence Conference*, 198–213. https://doi.org/10.1145/3715928.3737481

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The Skill Content of Recent Technological Change: An Empirical Exploration. *The Quarterly Journal of Economics*, *118*(4), 1279–1333. https://doi.org/10.1162/003355303322552801

Banerjee, A. V. (1992). A Simple Model of Herd Behavior. *The Quarterly Journal of Economics*, *107*(3), 797–817. https://doi.org/10.2307/2118364

Bick, A., Blandin, A., & Deming, D. (2024, September). *The Rapid Adoption of Generative AI* (tech. rep. No. w32966). National Bureau of Economic Research. Cambridge, MA. https://doi.org/10.3386/w32966

Bikhchandani, S., Hirshleifer, D., Tamuz, O., & Welch, I. (2024). Information Cascades and Social Learning. *Journal of Economic Literature*, *62*(3), 1040–1093. https://doi.org/10.1257/jel.20241472

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. *Journal of Political Economy*, *100*(5), 992–1026. https://doi.org/10.1086/261849

Boden, M. A. (1998). Creativity and artificial intelligence. *Artificial Intelligence*, *103*(1-2), 347–356. https://doi.org/10.1016/S0004-3702(98)00055-1

Bommasani, R., Creel, K. A., Kumar, A., Jurafsky, D., & Liang, P. (2022). Picking on the Same Person: Does Algorithmic Monoculture lead to Outcome Homogenization? https://doi.org/10.48550/ARXIV.2211.13972

Botelho, T., Fehder, D., & Hochberg, Y. (2021, July). *Innovation-Driven Entrepreneurship* (tech. rep. No. w28990). National Bureau of Economic Research. Cambridge, MA. https://doi.org/10.3386/w28990

Bresnahan, T. F., & Trajtenberg, M. (1995). General purpose technologies 'Engines of growth'? *Journal of Econometrics*, *65*(1), 83–108. https://doi.org/10.1016/0304-4076(94)01598-T

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at Work. *The Quarterly Journal of Economics*, *140*(2), 889–942. https://doi.org/10.1093/qje/qjae044

Burtch, G., Lee, D., & Chen, Z. (2024). The consequences of generative AI for online knowledge communities. *Scientific Reports*, *14*(1), 10413. https://doi.org/10.1038/s41598-024-61221-0

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. *Review of Economics and Statistics*, *90*(3), 414–427. https://doi.org/10.1162/rest.90.3.414

Choudhury, P., Starr, E., & Agarwal, R. (2020). Machine learning and human capital complementarities: Experimental evidence on bias mitigation. *Strategic Management Journal*, *41*(8), 1381–1411. https://doi.org/10.1002/smj.3152

Cui, K. Z., Demirer, M., Jaffe, S., Musolff, L., Peng, S., & Salz, T. (2026). The Effects of Generative AI on High-Skilled Work: Evidence from Three Field Experiments with Software Developers. *Management Science*, mnsc.2025.00535. https://doi.org/10.1287/mnsc.2025.00535

Da, Z., & Huang, X. (2020). Harnessing the Wisdom of Crowds. *Management Science*, *66*(5), 1847–1867. https://doi.org/10.1287/mnsc.2019.3294

Deephouse, D. L. (1999). To be different, or to be the same? It's a question (and theory) of strategic balance. *Strategic Management Journal*, *20*(2), 147–166. https://doi.org/10.1002/(SICI)1097-0266(199902)20:2<147::AID-SMJ11>3.0.CO;2-Q

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4573321

Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, *10*(28), eadn5290. https://doi.org/10.1126/sciadv.adn5290

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2024). GPTs are GPTs: Labor market impact potential of LLMs. *Science*, *384*(6702), 1306–1308. https://doi.org/10.1126/science.adj0998

Felin, T., & Zenger, T. R. (2017). The Theory-Based View: Economic Actors as Theorists. *Strategy Science*, *2*(4), 258–271. https://doi.org/10.1287/stsc.2017.0048

Felten, E., Raj, M., & Seamans, R. (2021). Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. *Strategic Management Journal*, *42*(12), 2195–2217. https://doi.org/10.1002/smj.3286

Fleming, L. (2001). Recombinant Uncertainty in Technological Search. *Management Science*, *47*(1), 117–132. https://doi.org/10.1287/mnsc.47.1.117.10671

Friedman, D., & Dieng, A. B. (2022). The Vendi Score: A Diversity Evaluation Metric for Machine Learning. https://doi.org/10.48550/ARXIV.2210.02410

Gans, J. S., Stern, S., & Wu, J. (2019). Foundations of entrepreneurial strategy. *Strategic Management Journal*, *40*(5), 736–756. https://doi.org/10.1002/smj.3010

Girotra, K., Meincke, L., Terwiesch, C., & Ulrich, K. T. (2023). Ideas are Dimes a Dozen: Large Language Models for Idea Generation in Innovation. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4526071

Goldfarb, A., Taska, B., & Teodoridis, F. (2023). Could machine learning be a general purpose technology? A comparison of emerging technologies using data from online job postings. *Research Policy*, *52*(1), 104653. https://doi.org/10.1016/j.respol.2022.104653

Goldfarb, B., & Kirsch, D. A. (2020, December). *Bubbles and Crashes: The Boom and Bust of Technological Innovation*. Stanford University Press. https://doi.org/10.1515/9781503607934

Golub, B., & Jackson, M. O. (2010). Naïve Learning in Social Networks and the Wisdom of Crowds. *American Economic Journal: Microeconomics*, *2*(1), 112–149. https://doi.org/10.1257/mic.2.1.112

Gotelli, N. J., & Colwell, R. K. (2001). Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, *4*(4), 379–391. https://doi.org/10.1046/j.1461-0248.2001.00230.x

Guzman, J., & Li, A. (2023). Measuring Founding Strategy. *Management Science*, *69*(1), 101–118. https://doi.org/10.1287/mnsc.2022.4369

Guzman, J., & Stern, S. (2020). The State of American Entrepreneurship: New Estimates of the Quantity and Quality of Entrepreneurship for 32 US States, 1988–2014. *American Economic Journal: Economic Policy*, *12*(4), 212–243. https://doi.org/10.1257/pol.20170498

Hellwig, C., & Veldkamp, L. (2009). Knowing What Others Know: Coordination Motives in Information Acquisition. *Review of Economic Studies*, *76*(1), 223–251. https://doi.org/10.1111/j.1467-937X.2008.00515.x

Henderson, R. M., & Clark, K. B. (1990). Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms. *Administrative Science Quarterly*, *35*(1), 9. https://doi.org/10.2307/2393549

Hoberg, G., & Phillips, G. (2016). Text-Based Network Industries and Endogenous Product Differentiation. *Journal of Political Economy*, *124*(5), 1423–1465. https://doi.org/10.1086/688176

Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences*, *101*(46), 16385–16389. https://doi.org/10.1073/pnas.0403723101

Hotelling, H. (1929). Stability in Competition. *The Economic Journal*, *39*(153), 41. https://doi.org/10.2307/2224214

Kaplan, S., & Vakili, K. (2015). The double-edged sword of recombination in breakthrough innovation. *Strategic Management Journal*, *36*(10), 1435–1457. https://doi.org/10.1002/smj.2294

Kerr, W. R., Nanda, R., & Rhodes-Kropf, M. (2014). Entrepreneurship as Experimentation. *Journal of Economic Perspectives*, *28*(3), 25–48. https://doi.org/10.1257/jep.28.3.25

Kim, H., Kang, H., & Song, J. (2026, May). Generative AI Fuels Solo Entrepreneurship, but Teams Still Lead at the Top. https://doi.org/10.48550/arXiv.2605.10291

Kleinberg, J., & Raghavan, M. (2021). Algorithmic monoculture and social welfare. *Proceedings of the National Academy of Sciences*, *118*(22), e2018340118. https://doi.org/10.1073/pnas.2018340118

Kogan, L., Papanikolaou, D., Seru, A., & Stoffman, N. (2017). Technological Innovation, Resource Allocation, and Growth*. *The Quarterly Journal of Economics*, *132*(2), 665–712. https://doi.org/10.1093/qje/qjw040
Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, *74*(2), 132–157. https://doi.org/10.1086/259131
Laufer, B., Oderinwale, H., & Kleinberg, J. (2025, August). Anatomy of a Machine Learning Ecosystem: 2 Million Models on Hugging Face. https://doi.org/10.48550/arXiv.2508.06811
Liang, W., Zhang, Y., Codreanu, M., Wang, J., Cao, H., & Zou, J. (2025). The widespread adoption of large language model-assisted writing across society. *Patterns*, *6*(12), 101366. https://doi.org/10.1016/j.patter.2025.101366
Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences*, *108*(22), 9020–9025. https://doi.org/10.1073/pnas.1008636108
Meincke, L., Nave, G., & Terwiesch, C. (2025). ChatGPT decreases idea diversity in brainstorming. *Nature Human Behaviour*, *9*(6), 1107–1109. https://doi.org/10.1038/s41562-025-02173-x
Morris, S., & Shin, H. S. (2002). Social Value of Public Information. *American Economic Review*, *92*(5), 1521–1534. https://doi.org/10.1257/000282802762024610
Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, *381*(6654), 187–192. https://doi.org/10.1126/science.adh2586
Otis, N. G., Clarke, R. P., Delecourt, S., Holtz, D., & Koning, R. (2023, December). The Uneven Impact of Generative AI on Entrepreneurial Performance. https://doi.org/10.31219/osf.io/hdjpk
Padmakumar, V., & He, H. (2023). Does Writing with Language Models Reduce Content Diversity? https://doi.org/10.48550/ARXIV.2309.05196
Pontikes, E. G., & Barnett, W. P. (2017). The Non-consensus Entrepreneur: Organizational Responses to Vital Events. *Administrative Science Quarterly*, *62*(1), 140–178. https://doi.org/10.1177/0001839216661150
Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 3980–3990. https://doi.org/10.18653/v1/D19-1410
Romer, P. M. (1990). Endogenous Technological Change. *Journal of Political Economy*, *98*(5, Part 2), S71–S102. https://doi.org/10.1086/261725
Scherer, F., & Harhoff, D. (2000). Technology policy for a world of skew-distributed outcomes. *Research Policy*, *29*(4-5), 559–566. https://doi.org/10.1016/S0048-7333(99)00089-X
Shane, S., & Venkataraman, S. (2000). The Promise of Entrepreneurship as a Field of Research. *The Academy of Management Review*, *25*(1), 217–226. https://doi.org/10.2307/259271
Si, C., Yang, D., & Hashimoto, T. (2024). Can LLMs Generate Novel Research Ideas? A Large-Scale Human Study with 100+ NLP Researchers. https://doi.org/10.48550/ARXIV.2409.04109
Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of The Royal Society Interface*, *4*(15), 707–719. https://doi.org/10.1098/rsif.2007.0213
Taeuscher, K., Bouncken, R., & Pesch, R. (2021). Gaining Legitimacy by Being Different: Optimal Distinctiveness in Crowdfunding Platforms. *Academy of Management Journal*, *64*(1), 149–179. https://doi.org/10.5465/amj.2018.0620
Toner-Rodgers, A. (2024). Artificial Intelligence, Scientific Discovery, and Product Innovation. https://doi.org/10.48550/ARXIV.2412.17866
Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical Combinations and Scientific Impact. *Science*, *342*(6157), 468–472. https://doi.org/10.1126/science.1240474
Wang, J., Veugelers, R., & Stephan, P. (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy*, *46*(8), 1416–1436. https://doi.org/10.1016/j.respol.2017.06.006
Webb, M. D. (2023). Reworking wild bootstrap-based inference for clustered errors. *Canadian Journal of Economics/Revue canadienne d'économique*, *56*(3), 839–858. https://doi.org/10.1111/caje.12661
Weitzman, M. L. (1998). Recombinant Growth. *The Quarterly Journal of Economics*, *113*(2), 331–360. https://doi.org/10.1162/003355398555595
Welch, I. (1992). Sequential Sales, Learning, and Cascades. *The Journal of Finance*, *47*(2), 695–732. https://doi.org/10.1111/j.1540-6261.1992.tb04406.x

Young, A. (2019). Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results*. *The Quarterly Journal of Economics*, *134*(2), 557–598. https://doi.org/10.1093/qje/qjy029

Zhang, S., Xu, J., & Alvero, A. (2025). Generative AI Meets Open-Ended Survey Responses: Research Participant Use of AI and Homogenization. *Sociological Methods & Research*, *54*(3), 1197–1242. https://doi.org/10.1177/00491241251327130

Zhao, E. Y., Fisher, G., Lounsbury, M., & Miller, D. (2017). Optimal distinctiveness: Broadening the interface between institutional theory and strategic management. *Strategic Management Journal*, *38*(1), 93–113. https://doi.org/10.1002/smj.2589

Zhou, E., & Lee, D. (2024). Generative artificial intelligence, human creativity, and art (M. Harding, Ed.). *PNAS Nexus*, *3*(3), pgae052. https://doi.org/10.1093/pnasnexus/pgae052

Appendix

A Measure Construction and Validation

A.1 Category-level LLM exposure index

Each category's exposure score is an ex-ante rating—the median across fifteen prompted LLM passes—of how effectively a 2022-era general-purpose model could help an entrepreneur ideate a new product in that category. It adapts the task-exposure concept of Eloundou et al. (2024) (how effectively a model can do the work at equal quality without special domain access) to the category level. Table A1 lists the scores; the rating prompt follows.

Table A1. LLM exposure index by category (ex-ante rating)

Category	Index	Category	Index
Artificial intelligence	7	E-commerce	4
Marketing	6	Health and fitness	3
Productivity	6	Travel	4
Education	6	Lifestyle	5
Design tools	3	Fintech	2
Social media	4	Family	4
Developer tools	5	Web3	2
SaaS	5	Games	2

Notes. The table reports each category's ex-ante integer score in $\{1, \ldots, 7\}$ for how effectively a general-purpose LLM could assist product ideation. The panel regression standardizes these scores to mean zero and unit variance; Section 4.2 describes the construction, and the rating prompt follows.

Role. You are an expert annotator adapting the task-exposure concept of Eloundou et al. (2024) to product ideation. Given one product category, output an integer 1–7 answering: how effectively can a 2022-era general-purpose LLM (e.g., ChatGPT) help an entrepreneur *ideate* a new product in this category? Here "ideate" means generate, refine, and critique the product concept—not build, fund, market, or operate it.

Baseline. Assume a text-only general-purpose model at the late-2022 (ChatGPT) level with no domain-specific access (no fine-tuning, proprietary data, specialized tooling, regulated or licensed knowledge, or tacit hands-on craft). Rate the model's unaided ideation help for the *typical* concept in the category, holding output quality constant—the "equal quality, no extra software" spirit of the exposure criterion. Rate potential assistance, not adoption or any market outcome, and judge *ex ante*: do not use or anticipate any result.

How to score. Score higher when the category's typical concepts are well covered by general text and can be generated and refined from general knowledge alone, and highest (7) when a general-purpose LLM is plausibly the core enabling technology of the category's products. Score lower for barriers the bare model cannot supply: deep tacit or aesthetic craft, specialized technical infrastructure, regulatory, compliance, or licensing expertise, or dependence on proprietary or physiological data.

Scale. 7, the LLM is the dominant ideation aid (core technology); 6, ideation maps directly onto core LLM strengths (text, planning, tutoring, copy); 5, clear overlap but the model is an assistant, not the core; 4, a partial or still-nascent use case; 3, helps only some sub-areas, or meaningful barriers limit it; 2, ideation needs domain craft the 2022 model lacked; 1, essentially useless.

Output. "Score: n" for an integer 1–7, then a one- or two-sentence rationale naming the affordances or barriers that drove the score.

The index is the median score across fifteen rater passes, five each from OpenAI gpt-4o, Google gemini-2.5-pro, and Anthropic claude-sonnet-4-6 under the prompt above. The per-category scores are consistent within each model and rank-correlated across families (cross-model Spearman 0.83–0.93), so the index does not depend on any single model.

A.2 Style-normalizing paraphrase

We use two checks to assess whether the paraphrase preserves product meaning. First, we compare each launch's original-text embedding with its paraphrase embedding. Across all 291,314 launches, the mean original-versus-paraphrase cosine is 0.90 (median 0.91; 5th percentile 0.78), and 0.15% of launches fall below 0.50. The paraphrase stays close to the original content for essentially every launch while regenerating the prose. Second, near-duplicate launches remain near-duplicates after paraphrasing. Among the 773 within-category-month launch pairs whose original-text cosine is at least 0.90, the median paraphrase-text cosine is 0.92, 94% remain at or above 0.80, and no pair falls below 0.50. Paraphrasing does not split apart launches that describe essentially the same product. Both checks compare embeddings of the two text versions; they do not involve human judgment of meaning preservation.

We then validate whether the paraphrase-based cosine tracks idea proximity as a reader would judge it, using LLM raters as proxies. We drew 60 within-category-month launch pairs spanning the range of measured similarity (20 in each of the low, middle, and high cosine terciles) and had three independent LLM raters score each pair's conceptual closeness on a 1–7 scale, shown only the texts and never the cosine. Mean rated closeness rises monotonically across the terciles—1.4, 2.0, and 2.9 at mean cosine values of 0.084, 0.193, and 0.337—and the rank correlation between the cosine measure and the across-rater mean rating is Spearman 0.73 ($n = 60$). Inter-rater agreement is high, and distinct from that rank correlation (Krippendorff's ordinal $\alpha = 0.73$; average-measures ICC(2, k) = 0.94). The validation supports the measure's rank ordering; we report magnitudes in cosine units for interpretability, not as a calibrated distance scale.

To separate convergence in product *ideas* from convergence in writing *style*, we rewrite every launch with a large language model before embedding, preserving its meaning while regenerating the prose. We generate one paraphrase per launch with Qwen3-32B—an open-weight model we run locally—using sampling temperature 0.7, top-p 0.95, and up to 256 output tokens, and embed the paraphrase. We rely on a local model because paraphrasing all 291,314 launches through a frontier commercial API would be prohibitively expensive; to verify that nothing hinges on this choice, we re-paraphrased a random subsample of 1,000 launches with the OpenAI GPT API and obtained qualitatively the same results. The instruction is:

> *System.* You are a writing assistant that paraphrases product descriptions. Your task is to rewrite the given product description in your own words while (1) preserving the core meaning and key features, (2) maintaining a similar length (within 20% of the original), (3) using different vocabulary and sentence structure, and (4) keeping a professional, marketing-appropriate tone. Output only the paraphrased version, nothing else.

User. Paraphrase this product description: `{text}`

A.3 Effective idea variety: exposure-gradient estimate

Table A2 reports the exposure-gradient estimate under the specification of equation (1).

Table A2. Exposure gradient in effective idea variety among 30 launches

Outcome	$\hat{\beta}$	SE	95% CI	t	WCB p	Perm. p
Effective idea variety	−0.892	0.200	[−1.317, −0.467]	−4.47	0.0083	0.0055

Notes. Coefficient on Post × Exposure$_z$ with category and month fixed effects; the paraphrase-based panel of N = 1,355 equal-weight category-month cells and 16 category clusters; the post period begins in December 2022. The interval uses CR1 inference with $t(15)$ critical values. The wild cluster bootstrap is studentized and null-imposed with Webb six-point weights; the permutation test reassigns category exposure scores. Both procedures are two-sided with B = 9,999.

B A Stylized Common-Signal Model

To clarify one mechanism consistent with the evidence, we use a stylized signal-aggregation model. It adapts the wisdom-of-crowds logic of Da and Huang (2020) and the non-strategic information structure in Morris and Shin (2002) and Angeletos and Pavan (2007) to product launches. Founders combine idiosyncratic private information with an AI-assisted signal that contains both shared and founder-specific components. The model yields one directional comparative static: greater weight on a more precise signal reduces dispersion around a category's latent target, while a shared component makes the resulting movement correlated across founders. The model organizes the empirical prediction; it does not identify the mechanism or quantify its effect.

Let θ_c denote the latent target concept in category c. Founder i observes a private signal $s_i = \theta_c + e_i$ with precision τ_s and an AI-assisted signal $m_i = \theta_c + u_c + v_i$ with precision τ_m. The errors are mean zero and mutually independent, and e_i and v_i are independent across founders. The component u_c is common within the category, while v_i captures founder-specific variation in prompts, model versions, and edits. Write $\sigma_s^2 = 1/\tau_s$ and $\sigma_m^2 = 1/\tau_m$, and let $\phi \in [0, 1]$ denote the share of the AI signal's variance that is common: $\mathrm{Var}(u_c) = \phi\sigma_m^2$ and $\mathrm{Var}(v_i) = (1 - \phi)\sigma_m^2$. Thus ϕ is also the correlation between two founders' AI signals, conditional on θ_c. Exposure enters through $\tau_m = \tau_m(E_c)$, with $\tau_m'(E_c) > 0$: AI assistance is more informative in categories where it is more relevant. Before widespread AI use, the limiting case $\tau_m \downarrow 0$ gives the AI signal negligible weight.

A founder chooses a launch position x_i to minimize $(x_i - \theta_c)^2$, with no strategic reward for matching or differentiating from other founders. With Gaussian signals and a diffuse prior, the optimal

position is the posterior mean,

$$x_i \;=\; w(E_c)\, m_i + \big(1 - w(E_c)\big)\, s_i, \qquad w(E_c) = \frac{\tau_m(E_c)}{\tau_s + \tau_m(E_c)}, \quad w'(E_c) > 0. \tag{2}$$

This is the non-strategic, $r = 0$ case of Morris and Shin (2002). The weight w is the share of total precision carried by the AI signal. Founders place more weight on that signal as exposure rises, without any coordination motive.

To match the within-category empirical comparison, hold θ_c and the realized common component u_c fixed. Only e_i and v_i then vary across founders, so the cross-founder dispersion of launch positions is

$$D(E_c) \;=\; \mathrm{Var}_i(x_i \mid \theta_c, u_c) = (1-w)^2\, \sigma_s^2 + w^2\, (1-\phi)\, \sigma_m^2, \tag{3}$$

which is also one half of the expected squared distance between two founders' positions. Substituting the signal variances and the weight from equation (2) gives

$$D \;=\; \frac{\tau_s + (1-\phi)\,\tau_m}{(\tau_s + \tau_m)^2},$$

so dispersion falls with exposure, $dD/dE_c < 0$, holding τ_s and ϕ fixed.[8] Two forces contribute to this result. When $\phi = 0$, a more precise founder-specific signal still reduces dispersion around θ_c, since $D = 1/(\tau_s + \tau_m)$. At fixed precision, sharedness reduces dispersion further, $\partial D/\partial \phi = -\tau_m/(\tau_s + \tau_m)^2 < 0$, and makes founders' movements correlated through u_c. Precision and sharedness are therefore distinct channels.

In the empirical embedding space, lower within-category dispersion corresponds to higher mean pairwise cosine. The model therefore predicts a positive coefficient on $\text{PostChatGPT}_t \times \text{Exposure}_c$ in equation (1) and an exposure-ordered rise across categories. These predictions concern the sign and monotonicity of the gradient, not its magnitude. The model omits founder adoption, strategic differentiation, entry, changes in category opportunities, and imitation. Differentiation would push dispersion upward, while a focal opportunity could pull founders toward the same applications without direct AI assistance. The observed gradient is consistent with the common-signal mechanism, but the model does not identify it separately from these alternatives.

C Additional Results

C.1 Dynamic event study

This appendix traces the exposure gradient quarter by quarter. We replace the single $\text{PostChatGPT}_t \times \text{Exposure}_c$ interaction in equation (1) with quarterly event-time indicators interacted with exposure, so each coefficient shows how far high-exposure categories had moved relative to low-exposure

[8] $dD/d\tau_m = -\left[(1+\phi)\tau_s + (1-\phi)\tau_m\right]/(\tau_s + \tau_m)^3 < 0$. Because $\tau_m'(E_c) > 0$, the chain rule gives $dD/dE_c < 0$.

ones in that quarter (paraphrased panel, $N = 1{,}355$ cells; the reference bin is the quarter before the November 30, 2022 release; category-clustered standard errors). Figure A1 plots the result. Before ChatGPT the coefficients hover near zero, averaging 0.006 in absolute value. In the release quarter the gradient appears at +0.010, reaches +0.019 within two quarters, and stays between +0.012 and +0.019 through the last complete quarter of the sample—a level shift at the release that persists for three years, rather than a gradual drift that happens to straddle it.

A formal test of the pre-period coefficients supports this reading. Testing all nineteen quarterly leads jointly asks more than sixteen clusters can answer, so we test at the annual resolution: the joint test does not reject the null of no differential pre-trends on either panel ($F = 0.55$, $p = 0.70$ on the paraphrased panel; $F = 0.60$, $p = 0.67$ on the original-text panel), and no annual lead is individually significant. The one finer-grained (semi-annual) pre-period lead that is individually significant is negative, so any residual pre-trend works against the positive post-period gradient rather than manufacturing it. Consistent with this, the specification that absorbs exposure-specific trends outright leaves the gradient at +0.0178 (Table 4, column 2).

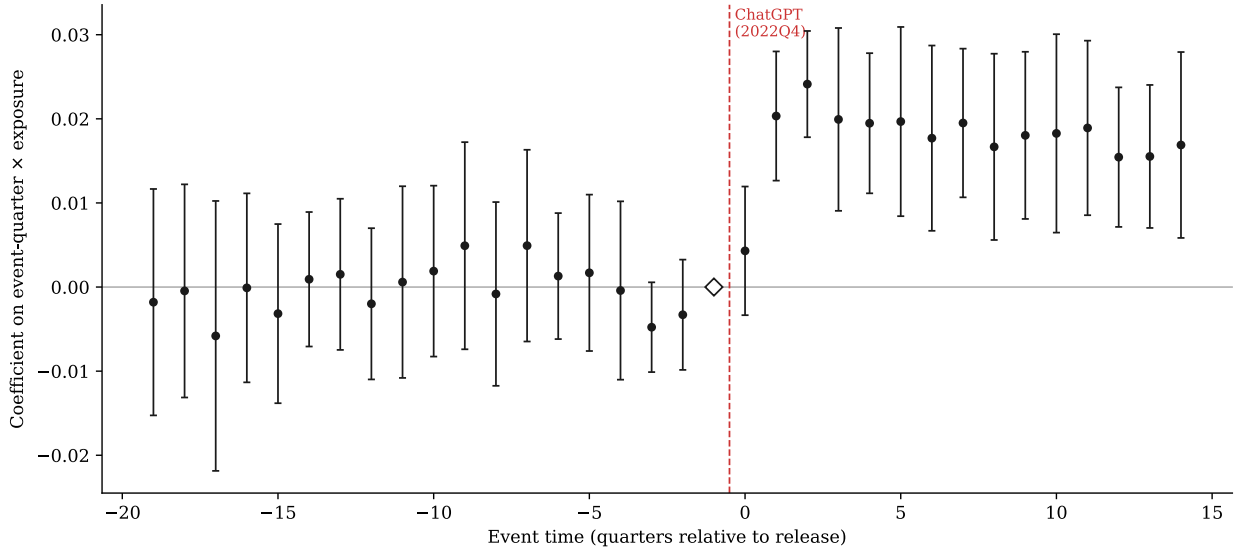


Figure A1. Dynamic event study of the exposure gradient. Coefficients on quarterly event-time indicators interacted with standardized LLM exposure, replacing the single post interaction in equation (1); paraphrased category-month panel ($N = 1{,}355$ cells). The reference bin (open diamond) is the quarter immediately before the November 30, 2022 release; bars are 95% confidence intervals from category-clustered standard errors (16 clusters).

C.2 Entry composition versus within-type convergence

Post-ChatGPT entry shifts sharply toward AI products: their share of launches rises from 5.6% to 42.0% overall, and from 8.7% to 52.5% in the four highest-exposure categories (index ≥ 6). Because AI products resemble one another, this shift alone would raise within-cell similarity. We separate the two margins with the block accounting formula $S \approx p^2 S_{AA} + (1-p)^2 S_{NN} + 2p(1-p)S_{AN}$, where

p is a cell's AI-product share and S_{AA}, S_{NN}, S_{AN} are the mean pairwise similarities within and across the two product types. Holding the block similarities at their pre-period values and moving only the AI share to its post-period value reproduces about 58.8% of the +0.0643 similarity rise in these categories (an approximate diagnostic rather than an exact decomposition of the distinct-pair cell mean); the remaining 41.2% comes from rising block similarities, both within and across the two product types (non-AI/non-AI +0.0283, AI/AI +0.0368, and cross-type +0.0199).

Whether that within-type convergence also produces the cross-category gradient is harder to pin down. Among launches that do not contain AI-related keywords, the exposure gradient is positive but modest ($\beta = +0.0074$, cluster-robust $t = 1.94$; wild-cluster bootstrap $p = 0.084$): within-type convergence tilts the right way with exposure but is no longer significant at conventional levels once the index pools multiple raters. The complementary AI-only subsample is too thin to estimate a stable gradient—few category-months clear the thirty-launch floor and the interaction is nearly collinear with the calendar-month effects—so we cannot cleanly separate how much of the panel gradient is within-type and how much is compositional. What the decomposition does establish is that both channels operate: more-exposed categories both attract more mutually similar AI products and see genuine convergence within product types. These calculations use the original-text similarity measure; the exposure gradient is nearly identical on original and paraphrased text (Section 5.5), so the diagnostic is informative about the primary result, though not a decomposition of it.

Online Appendix

Does Generative AI Make Startup Ideas Alike? Evidence from Product Hunt

This appendix supplements "Does Generative AI Make Startup Ideas Alike? Evidence from Product Hunt." It collects the robustness and supplementary analyses behind the headline cross-category difference-in-differences (DiD) estimate, the coefficient β on Post $\times$ Exposure$_z$ in the category-month panel, estimated from sim $= \alpha_c + \gamma_t + \beta(\text{Post} \times \text{Exposure}_z) + \varepsilon$ with standard errors clustered on category. Sections OA1–OA4 group the material by the question it answers: identification and timing (OA1), measurement (OA2), sample and inference (OA3), and the exposure index (OA4). Each table states the panel it uses (paraphrase-based primary or original-text), the post boundary, and the inference design; wild cluster bootstrap and exposure-permutation tests follow Cameron et al. (2008) and Young (2019). The map below links each manuscript claim to the section that supports it.

Manuscript claim or object	Here
Pre-trends and event-study sensitivity (Section 4.3)	OA1
Measure, effective variety, and embedder choice (Sections 4.1, 5.1–5.2)	OA2
Few clusters and high-vote sample (Section 5.5)	OA3
Founder-experience symmetry and within-type convergence (Section 5.4)	OA3.3
Cross-model stability of the exposure index (Section 4.2)	OA4

OA1 Identification: Pre-Trends and Placebo Timing

This section tests whether high- and low-exposure categories were on differential trends before ChatGPT, and whether the descriptive event-study excess depends on its timing and window choices. Unless a row states otherwise, estimates use the category-month panel (N = 1,355, 16 category clusters, post boundary December 1, 2022, so that November 2022 is coded pre-period); the pre-period joint test in Table S1 uses the 762 pre-period cells.

Pre-period joint test. The direct test runs on pre-ChatGPT months only, with linear and quadratic exposure-time interactions tested jointly. This smooth-trend test does not reject the null of no differential linear or quadratic pre-trends on either panel (Table S1). The annual event-study lead test also does not reject on either panel ($F = 0.55$, $p = 0.70$ on the primary panel; $F = 0.60$, $p = 0.67$

on the original-text panel; Appendix C.1 of the manuscript), and no annual lead is individually significant. The one finer-grained pre-period lead that is significant is negative, so any residual imbalance runs against—not toward—the positive post-ChatGPT gradient, and the specification that absorbs exposure-specific trends leaves the estimate at +0.0178.

Table S1. Pre-period differential-trend joint test

Panel	Linear, quadratic t (p)	Joint F (p)
Paraphrased (primary)	+0.29 (0.776); −0.27 (0.789)	0.04 (0.957)
Original text	+0.18 (0.859); −0.10 (0.921)	0.09 (0.917)

Notes. Pre-ChatGPT months only ($N_{\text{pre}} = 762$), on the paraphrase-based primary panel and the original-text panel; post boundary December 1, 2022 (November 2022 coded pre-period); 14 category clusters (family and lifestyle have no retained pre-period cell), category-clustered SE. Exposure-interacted linear and quadratic time terms are tested jointly (2 restrictions); a small-G correction to the F reference distribution leaves the p-values unchanged.

Placebo timing and window sensitivity. Table S2 collects two checks on the descriptive event-study excess. They answer two different questions: whether the exposure-aligned ordering predates ChatGPT (placebo timing), and how much the visualized excess depends on the window choice. At a December 2021 placebo cutoff, estimated on observations strictly before ChatGPT's release, the excess–exposure correlation is small and negative, even though part of the excess level remains by construction. Re-estimating the DiD at the fake boundaries yields interaction estimates near zero ($\hat{\beta} = -0.0011$, $t = -0.38$ one year early; $\hat{\beta} = -0.0015$, $t = -0.50$ one year late, using post-period data only), with wild-cluster bootstrap and permutation p-values above 0.63 in every case. The ordering appears in every window and is strongest in the first 6 to 12 months after the release, while the level grows mechanically with the window length.

Table S2. Event-study excess: placebo timing and window sensitivity

	Categories	Excess	Spearman	p
Panel A: placebo timing (placebo rows use data strictly before 2022-12-01)				
True cutoff, full window	14	+0.0267	+0.61	0.021
Placebo 2021-12-01	13	+0.0090	−0.23	0.420
Panel B: window sweep at the true cutoff				
First 6 months	13	+0.0147	+0.89	< 0.001
First 12 months	14	+0.0184	+0.83	< 0.001
Full window	14	+0.0267	+0.61	0.021

Notes. Event-study construction on the paraphrase-based primary panel. Excess is the aggregate post-cutoff excess over the extrapolated pre-trend; Spearman is the rank correlation of per-category excess with exposure. Placebo estimates use only data strictly before the true cutoff (data cut at 2022-11-30).

OA2 Measurement: Cosine Geometry and Effective Variety

The headline measure averages pairwise cosine similarity within a category-month cell. We first test two alternative similarity measures. We then use effective idea variety to express the same geometry on a fixed 30-launch scale. Raw cosine can overstate proximity because two launches can score high partly because both inherit a common direction from MiniLM's embedding space. Table S3 therefore tests whether the exposure gradient survives removing that direction or replacing MiniLM with a structurally different encoder. Both checks run on the original-text panel (baseline $\beta = +0.0177$), whose gradient is nearly identical to the paraphrase-based headline.

Panel A removes the direction shared by all MiniLM vectors and renormalizes before computing cosine similarity (Mu et al., 2017); the exposure gradient remains positive and precise ($\hat{\beta} = +0.0154$, $p = 0.001$).

Table S3. Alternative cosine constructions and a second sentence encoder

Measure	$\hat{\beta}$	SE	CR1 p
Panel A: pairwise similarity after geometry correction			
Mean cosine (MiniLM, reference)	+0.0177	0.0042	< 0.001
Anisotropy-corrected cosine	+0.0154	0.0047	0.001
Panel B: second sentence encoder			
Mean cosine (BGE-large-en-v1.5)	+0.0120	0.0027	< 0.001

Notes. Original-text panel (baseline $\hat{\beta} = +0.0177$); 1,355 category-month cells, 16 category clusters; the post period begins in December 2022. Each row estimates the same panel DiD; p-values are asymptotic CR1 category-clustered. Panel A estimates the first principal component of the original-text embeddings on the full corpus, subtracts its projection from every launch vector, and L2-renormalizes before recomputing cosine; Panel B re-embeds the same launch universe with a 1024-dimensional retrieval-tuned model. These are sensitivity checks for the primary mean-cosine outcome.

Panel B changes the embedding model. BGE-large-en-v1.5 uses a different encoder family, embedding dimension, and training objective, yet produces a similar gradient ($\hat{\beta} = +0.0120$, $p < 0.001$). Its category-level detrended excesses are closely correlated with those from MiniLM (Pearson $r = 0.940$ across the 14 categories with sufficient pre-period observations). MiniLM and BGE therefore yield the same positive cross-category ordering.

OA2.1 Effective Idea Variety Among 30 Launches

Mean pairwise cosine summarizes how close the typical pair of launches is. Effective idea variety translates the same similarity structure into an effective count (Friedman & Dieng, 2022). For each category-month cell, let A be a uniformly sampled subset of 30 launches, drawn without replacement, and let S_A be the cosine Gram matrix of their Qwen-paraphrased, L2-normalized MiniLM embeddings. If λ_{Ak} denotes the kth eigenvalue of $S_A/\mathrm{tr}(S_A)$, then

$$V_1(A) = \exp\left(-\sum_{k=1}^{30} \lambda_{Ak} \log \lambda_{Ak}\right), \qquad \theta_{ct} = \mathbb{E}_A[V_1(A)].$$

Because S_A is a Gram matrix, it is positive semidefinite: the normalized eigenvalues are nonnegative and sum to one even though individual off-diagonal cosines can be negative. The score ranges from 1 when all embeddings coincide to 30 when they are mutually zero-cosine. A score of m means that the cohort has the same spectral variety as m mutually zero-cosine launches under this kernel.

We estimate θ_{ct} by averaging 200 random 30-launch subsets per cell, drawn without replacement within each draw, using fixed cell-specific seeds. The mean across cells of the cell-level Monte Carlo standard error is 0.053. This quantity measures simulation error in the estimated cell outcome; it is not the regression standard error. Fixing the cohort at 30 gives every cell the same 1–30 support and the same opportunity to reveal variety as launch volume changes. This parallels the fixed-comparison principle of the manuscript's fixed-size peer check but is a different estimator.

Appendix A.3 of the manuscript reports the exposure-gradient estimate and inference. Under the existing design, the variety estimate is an associational companion. Figure S1 plots the descriptive monthly equal-category mean, which falls from 20.81 before ChatGPT to 19.47 afterward.

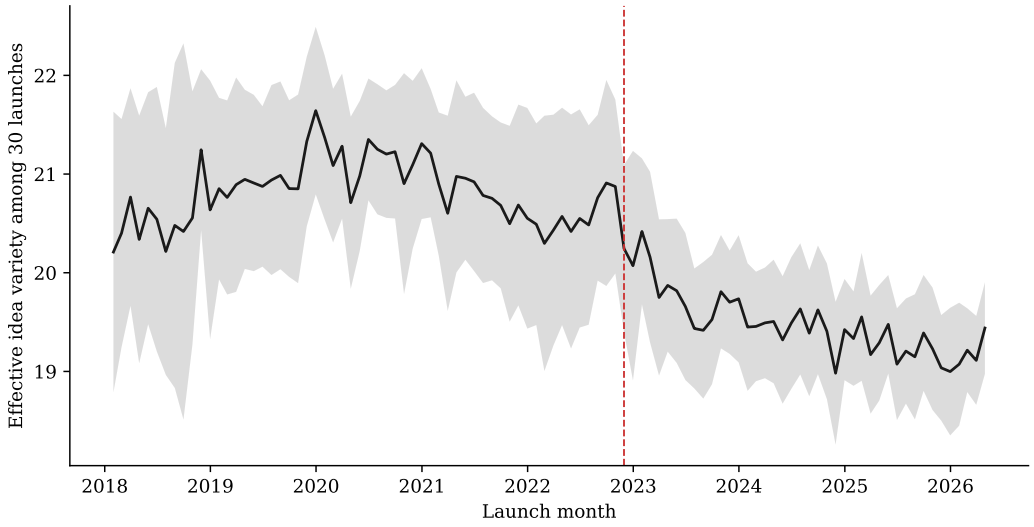


Figure S1: Effective idea variety among 30 launches, monthly equal-category means. Each monthly value is the equal-weight mean across categories observed that month; each category-month estimate averages 200 random subsets of 30 launches. The shaded region is a pointwise 95% band based on the cross-category standard error. The dashed line marks the November 30, 2022 release of ChatGPT. The series is descriptive.

OA2.2 Launch-Level Similarity by Category, All 16 Categories

The manuscript's category grid shows the launch-level similarity distribution shifting right in four high-exposure categories, with smaller and mixed changes in four low-exposure ones. Figure S2 extends the same construction to all sixteen categories, ordered by the exposure index. Each panel averages cell-level distributions with equal weight per cell, so months with many launches do not drive the shape; the exposure score is in parentheses and pre/post launch counts are in the legend. For each of the fourteen categories with a retained pre-period cell, a dotted vertical line marks the equal-cell-weighted pre-ChatGPT 25th-percentile cutoff. Because the curves are kernel-smoothed, the plotted pre-period density need not place exactly one quarter of its area to the left of the line. Family and lifestyle have no retained pre-period cell, so their panels show the post-period density only and no cutoff line. The rightward shift is largest among the high-exposure categories at the top of the grid. Middle-exposure categories are more heterogeneous, and the low-exposure categories at the bottom mostly overlap or move in mixed directions—web3, for instance, shifts left in the raw distribution.

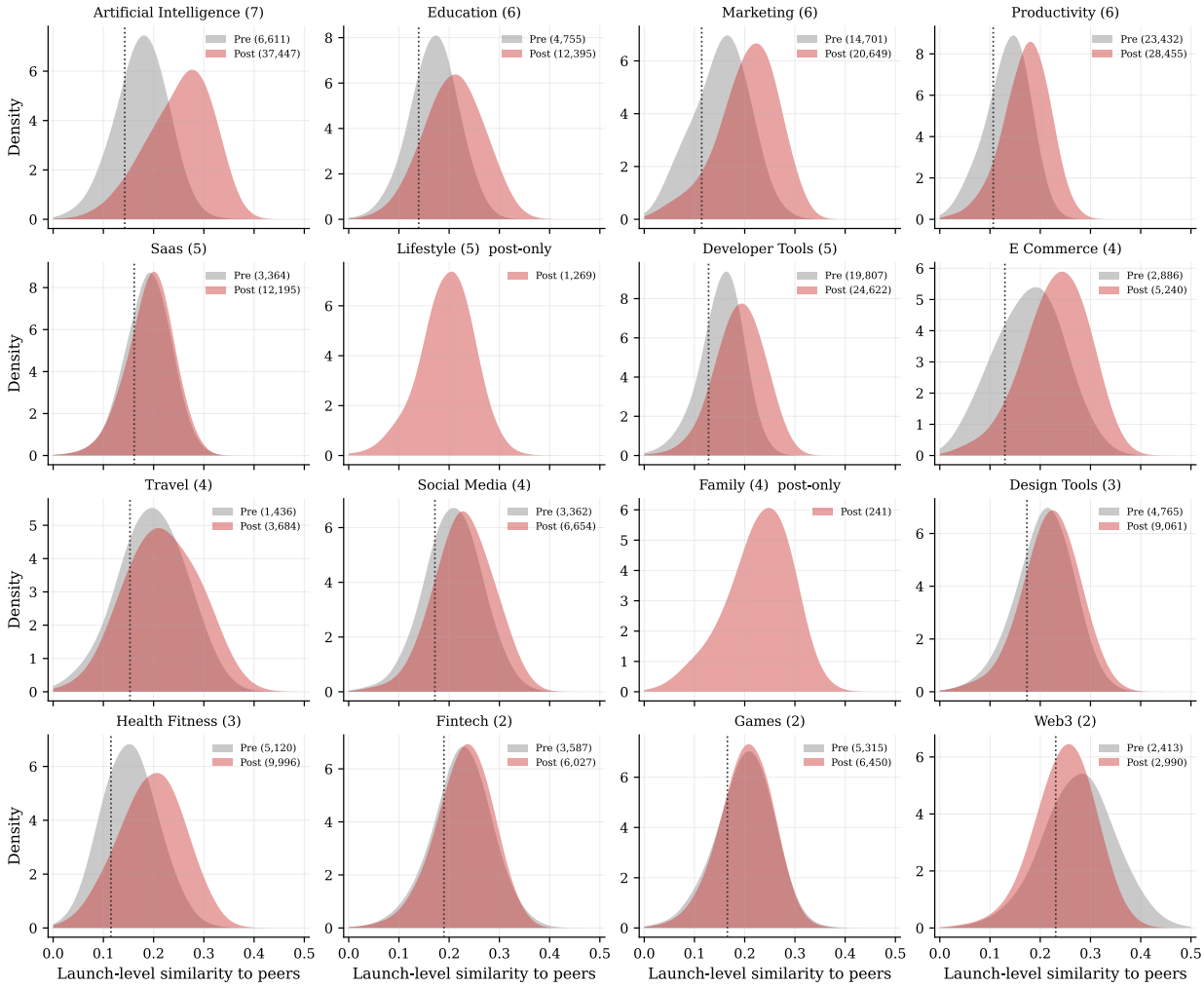


Figure S2: Launch-level similarity before and after ChatGPT, all 16 categories, ordered by the LLM exposure index (score in parentheses). Each panel averages cell-level distributions with equal weight per cell; legends give pre- and post-ChatGPT launch counts. Dotted vertical lines mark category-specific equal-cell-weighted pre-period 25th-percentile cutoffs. Family and lifestyle have no retained pre-period cell, so they show the post-period density only and no cutoff line.

OA2.3 Distinctive-Tail Sensitivity to Cutoff and Peer-Set Construction

The manuscript defines a distinctive launch by the bottom quartile of its category's pre-ChatGPT launch-level similarity. Table S4 re-runs the tail analysis with the cutoff set instead at the bottom decile, quintile, and third, holding everything else fixed (equal-cell calibration, 14-category sample, and inference design). The exposure gradient is negative and statistically significant at every cutoff, and the post-period share falls by a stable 34–38% of its pre-period level throughout. Panel B reports the same bottom-quartile analysis with a fixed 30-launch peer set in each cell.

Table S4. Distinctive-tail sensitivity to cutoff and peer-set construction

	Pre share	Post share	$\hat{\beta}$ (SE)	WCB p	Perm. p
Panel A: cutoff choice					
Bottom decile ($q = 0.10$)	0.100	0.062	−0.0207 (0.0055)	0.0110	0.0094
Bottom quintile ($q = 0.20$)	0.200	0.127	−0.0433 (0.0106)	0.0079	0.0044
Bottom quartile ($q = 0.25$, paper)	0.250	0.160	−0.0532 (0.0132)	0.0102	0.0052
Bottom third ($q = 1/3$)	0.333	0.219	−0.0741 (0.0166)	0.0080	0.0029
Panel B: fixed-reference peer construction					
Bottom quartile, fixed 30-launch reference	0.250	0.162	−0.0545 (0.0125)	0.0099	0.0039

Notes. Each row re-runs the equal-cell tail analysis with the distinctive-launch cutoff set to the equal-cell-weighted pre-period quantile q of launch-level similarity within each category, held fixed after ChatGPT; the paper uses the bottom quartile. In Panel B each launch is scored against a fixed set of 30 launches drawn within its cell. Outcome is the cell share of distinctive launches; DiD with category and month fixed effects, Post × Exposure$_z$, CR1 SE clustered on 14 categories (N = 1,334 cells). WCB p: null-imposed studentized Webb wild-cluster bootstrap; Perm. p: permutation of category exposure scores; B = 9,999 each.

OA3 Sample and Inference Robustness

OA3.1 Leave-One-Out Stability Across Category Clusters

With only 16 category clusters a single category could drive the result. We re-estimate the main monthly coefficient leaving out each category in turn. All leave-one-out estimates fall in a narrow positive band, from +0.015 (web3 excluded) to +0.020 (health and fitness excluded), and every estimate remains significant (CR1 $p < 0.01$).

OA3.2 Popularity Filtering

The manuscript reports the gradient keeping only the more-upvoted half of each category-month cell (Section 5.5). The complementary fixed-size check keeps the 30 most-upvoted launches per cell—the panel's minimum cell size, matching the fixed-reference-set check—so attention-based and full-cell selection are compared at identical cell sizes: $\beta = +0.0195$ ($t = 3.90$; Webb wild cluster bootstrap $p = 0.0236$, $B = 9{,}999$; paraphrase-based primary panel).

OA3.3 Founder Experience: Gradient Equality and Within-Type Convergence

The manuscript's founder analysis measures each launch against all of its same-cell peers. Because newcomers moving toward incumbents would mechanically raise incumbents' similarity to that changing peer pool, Table S5 adds three tests on the maker-classified sample (mixed teams coded as experienced). First, a launch-level test of the pre/post change in the first-time gap: the Post × FirstTime coefficient is positive but not significant under the wild-cluster bootstrap, so the manuscript's narrowing of the gap is reported as descriptive. Second, the two groups' exposure gradients are close in magnitude: +0.0179 for first-time teams and +0.0165 for experienced teams, each estimated on its own panel with its own category and month fixed effects. A test of the difference, which keeps those fixed effects group-specific and clusters on category across both panels, does not reject equality. The joint test accounts for the covariance between the two panels, so the difference is estimated more precisely than either gradient on its own. Third, similarity measured *only among each group's own launches* rises with exposure for both groups: first-time launches grow more similar to other first-time launches, and experienced launches to other experienced launches.

OA4 Cross-Model Stability of the LLM Exposure Index

The DiD interaction relies on a category-level exposure index, so we report how the ratings and the downstream estimate behave across raters. The index pools each category's 1–7 exposure score across three LLM families under an identical rubric, which adapts the Eloundou et al. (2024) task-exposure concept to the single activity of ideating a product in a category. Three independent LLM families rate every category (gpt-4o, gemini-2.5-pro, claude-sonnet-4-6), five passes each. We compute within-model reliability and cross-model rank agreement, and we re-estimate the DiD under each model's median index.

The rubric narrows the Eloundou et al. (2024) concept (rating work by how well an LLM performs or assists it at equal quality and without special domain access) to ideation alone, holding the capability baseline at the late-2022 ChatGPT release with no fine-tuning, proprietary data, tools, or tacit craft. The scale runs from 7, where a general-purpose LLM is plausibly the core enabling technology of the category's products, to 1, where ideation needs domain craft the 2022 model lacked; the rubric names no category, so the scores are not anchored to particular categories.

Table S6 reports the per-category scores, within-model reliability, cross-model agreement, and the DiD re-estimated under each re-elicited index.

Reliability is high and the scores agree closely across models. The headline index is the pooled median across all fifteen passes ($\hat{\beta} = +0.0184$); substituting any single model's median ($\hat{\beta}$ from +0.0165 to +0.0184) leaves the gradient positive and statistically significant in every case (bottom row of Table S6).

Table S5. Founder-experience tests: gap change, gradient equality, and within-type convergence

Test	$\hat{\beta}$ (SE)	t	WCB p	N
Panel A: change in the first-time gap				
Post × FirstTime (launch level, cell FE)	+0.0021 (0.0010)	2.01	0.115	248,602
Panel B: equality of exposure gradients				
Gradient equality: Post × Exposure$_z$ × FirstTime	+0.0014 (0.0010)	1.43	0.233	2,573
Panel C: within-type convergence and composition reweighting				
Within-type DiD, first-time only (S_{FF})	+0.0169 (0.0047)	3.63	0.013	1,231
Within-type DiD, experienced only (S_{EE})	+0.0133 (0.0035)	3.83	0.016	1,151
Pooled founder panel (unweighted)	+0.0178 (0.0042)	4.25	0.010	1,355
Composition reweighted to pre-period share	+0.0178 (0.0042)	4.29	0.011	1,355

Notes. Maker-classified launches in the paraphrase-based primary panel; mixed teams coded as experienced; post boundary December 1, 2022 (November 2022 coded pre-period). All tests use the 248,602 maker-classified launches that fall in retained ($n \geq 30$) cells (of 288,929 launches in those cells); the maker classification covers 250,539 launches corpus-wide. Panel A: launch-level average cosine to all same-cell peers regressed on FirstTime and Post × FirstTime with cell (category × month) fixed effects; the pre-period first-time gap is −0.0066 (the manuscript's additive category-and-month specification gives the −0.009 level cited in Section 5.4). Panel B: the equality row stacks the two founder types' cell-level panels (cells with at least 15 launches of the group) with *group-specific* category and month fixed effects, so that each group carries its own nuisance effects and the triple coefficient is exactly the first-time minus experienced difference in exposure gradients; the stacked fit reproduces the separately estimated gradients (+0.0179 and +0.0165) to machine precision, and clustering on category across both panels retains the covariance the two groups share. Panel C: the within-type rows compute cell-level mean pairwise cosine only among the group's own launches (cells with at least 20 such launches); the pooled row estimates the DiD on all maker-classified launches, and the reweighted row holds each category's first-time share at its pre-period value. All rows: CR1 category-clustered SE; Webb null-imposed studentized wild-cluster bootstrap, B = 9,999; 16 category clusters except the experienced-only within-type row, which uses 15 (one category has no cell with at least 20 experienced launches). Maker classification covers 250,539 of 291,314 launches (86.0% overall: 78.1% pre-ChatGPT and 90.3% post-ChatGPT, and 81–95% across categories); the tests here use classified launches only, and the lower pre-period coverage mainly affects the descriptive first-time entry share, not the within-type gradients, which are identified from post-versus-pre changes among classified launches.

Table S6. Exposure scores across rating models

	Median score (five passes per model)		
Category	GPT-4o	Gemini	Claude
Artificial intelligence	7	7	7
Marketing	6	6	6
Productivity	6	5	6
Education	6	6	6
Design tools	3	2	3
Social media	5	3	4
Developer tools	4	6	5
SaaS	5	4	5
E-commerce	5	3	4
Health & fitness	3	2	3
Travel	5	5	4
Lifestyle	5	5	5
Fintech	3	1	2
Family	4	4	4
Web3	2	2	2
Games	3	2	2
Krippendorff α (ordinal)	0.93	0.93	0.97
Categories with all passes within 1 pt of median	100%	88%	100%
DiD $\hat{\beta}$ (SE) using this index	+0.0183 (0.0045)	+0.0165 (0.0047)	+0.0184 (0.0041)

Notes. Scores use the 1–7 rubric described in this section. The main DiD builds Exposure_z from the pooled median across all fifteen passes (five per model family), which gives the headline $\hat{\beta} = +0.0184$; the three columns report each family's five-pass median. Reliability is computed over the five passes within each model family. The medians rank-correlate across families (Spearman 0.83–0.93), and average pairwise within-family agreement is 0.94–0.98. The final row re-estimates the manuscript DiD on the paraphrase-based primary panel, substituting each family's median index for the pooled median.

References

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. *Review of Economics and Statistics*, *90*(3), 414–427. https://doi.org/10.1162/rest.90.3.414

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2024). GPTs are GPTs: Labor market impact potential of LLMs. *Science*, *384*(6702), 1306–1308. https://doi.org/10.1126/science.adj0998

Friedman, D., & Dieng, A. B. (2022). The Vendi Score: A Diversity Evaluation Metric for Machine Learning. https://doi.org/10.48550/ARXIV.2210.02410

Mu, J., Bhat, S., & Viswanath, P. (2017). All-but-the-Top: Simple and Effective Postprocessing for Word Representations. https://doi.org/10.48550/ARXIV.1702.01417

Young, A. (2019). Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results*. *The Quarterly Journal of Economics*, *134*(2), 557–598. https://doi.org/10.1093/qje/qjy029