

# Model-Driven UI: Why Drilldowns Die

Machine Learning · UX · Recommender Systems

*From recommendation feeds to intent-to-execution: what a decade of platform data says about menus, and the arithmetic that follows when a model drives the interface. Top-k chips replace the tree, every transition becomes suggest and confirm, and exact search is the escape hatch.*

doi: 10.5281/zenodo.23059359

---

## Assets & Materials

|                                                                         |                                                                         |
|-------------------------------------------------------------------------|-------------------------------------------------------------------------|
| A small e-commerce demo demonstrating UX driven by a recommender model. | <a href="https://booster-demo.vski.ai">https://booster-demo.vski.ai</a> |
|-------------------------------------------------------------------------|-------------------------------------------------------------------------|

Every interface is a search algorithm. Menus, facets, drilldowns: the user executes a ranking by hand, one click at a time. Put a model in charge of the ranking and the interface is forced to shrink. That is not a design taste. It is arithmetic, and the last decade of consumer software already ran the experiment at the largest possible scale.

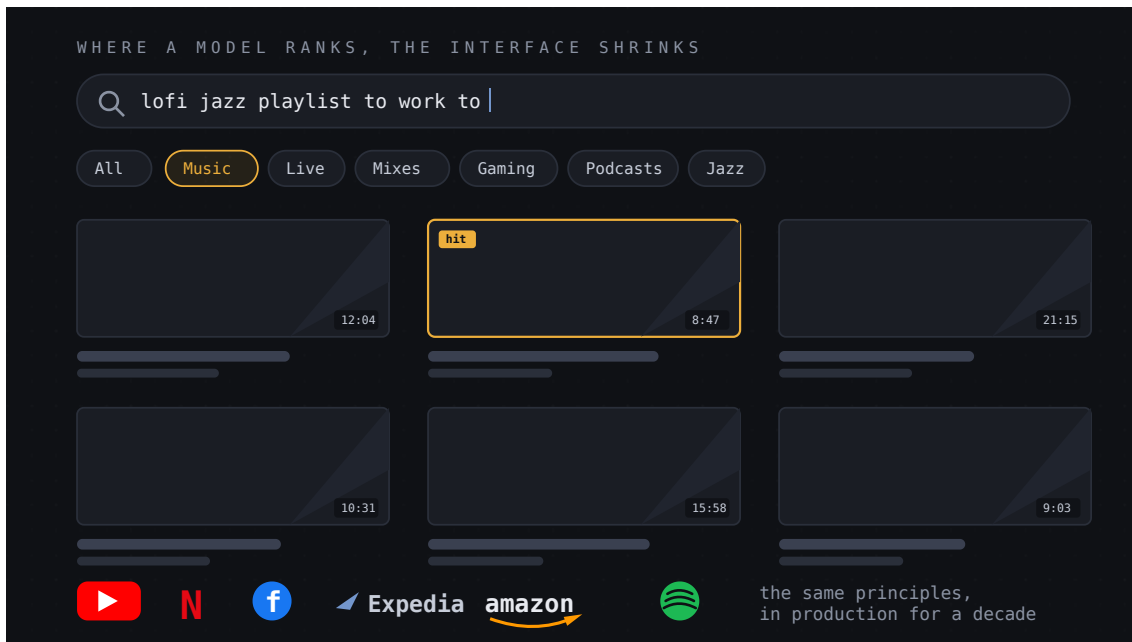
This post walks the arc in two waves: recommendation-driven interfaces, then intent-to-execution. Then it states the principles we build by, because the conclusions are more reusable than the examples.

## Wave one: where a model ranks, navigation loses

The numbers are old enough to be boring. Netflix, in its own recommender paper (Gomez-Uribe and Hunt, ACM TMIS 2015): about 80% of hours streamed come from recommendations, under 20% from search, and the company valued the personalization stack at a billion dollars a year. YouTube: recommendations drive the majority of all watch time, reaffirmed in CEO Neal Mohan's 2025 annual letter, on a screen (the TV) that has no menus worth using. TikTok took the same idea further: the feed is the entire product, and average usage climbed from 27 minutes a day in 2019 to about an hour and a half by 2025 (Data Reportal, via Exploding Topics).

The lesson people took from wave one is “feeds win”. The actual lesson is narrower and more useful: **wherever a model ranks the options, the navigation in front of it loses**. The feed is simply what remains when you delete the menus entirely. The same force works far below the scale of a content platform, on tags and categories and filters, and that is where most software still lives.

Users noticed before builders did. On the open web, 68% of Google searches now end without a click (SparkToro/Datos, early 2026; only about 276 of every 1,000 searches send a visit to the open web at all), and Gartner predicted in February 2024 that traditional search volume would fall 25% by 2026. On e-commerce sites, 43% of shoppers go straight to the search bar on arrival, and visitors who use site search convert at 4.63% versus 2.77% for those who browse (Econsultancy via Algolia), the much-cited 2-3x lift. People state intent when you let them. The menu tree is a tax most of them pay only because nobody removed it.

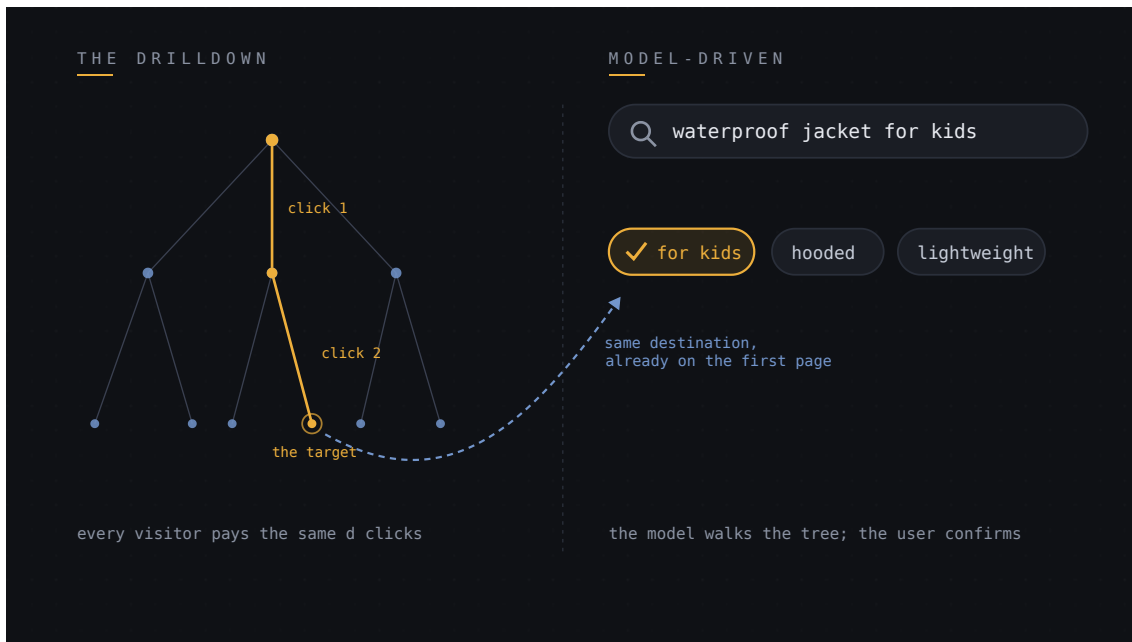


## The drilldown is a decision tree evaluated by hand

Now the arithmetic that makes the tax visible.

Take a catalog of  $N$  items behind a faceted tree with branching factor  $b$  and depth  $d$ , so  $b^d \geq N$ . Consider the best-case user: they know exactly what they want. They still execute  $d$  rounds, reading  $b$  labels per round, clicking one:  $\sim d \cdot b$  label reads and  $d$  clicks in total. The cost is identical for every visitor because the tree cannot use context. A drilldown is a search algorithm in which the human is the processor.

Put a ranker in charge and the same work moves into the model. Given context (query, session, user), the interface offers the top- $k$  chips: tags, categories, filters. And note how much easier this problem is than what Netflix solves: ranking a few dozen labels over one domain is not ranking a million videos for a billion people. A trained model on a small label space drives  $P(\text{target} \in \text{top-}k)$  close to one. The chip the user is looking for is on the first page almost every time, so one confirm replaces the whole tree walk.



Misses are cheap too. A dismiss re-ranks and the next hop arrives. If  $p_1, p_2, p_3$  are the hit probabilities of successive hops, the fraction of sessions still unresolved after three hops is the product of the failures:

$$P(\text{miss after 3 hops}) = (1 - p_1)(1 - p_2)(1 - p_3)$$

With illustrative numbers, say  $p_1 = 0.9, p_2 = 0.6, p_3 = 0.4$ :  $(0.1)(0.4)(0.6) = 2.4\%$  of sessions left uncovered, and that tail walks to exact search, the escape hatch below. Even pessimistic numbers leave a small tail; the point of the arithmetic is the shape, not the constants.

So dropping the drilldown alone cuts interaction cost by factors: from  $d \cdot b$  label reads and  $d$  clicks to one to three confirms. Add quality semantic search and personalization and the same loop ranks items, not just labels.

## The transition function: suggest, confirm or dismiss

Strip the details and a model-driven interface has exactly one transition. The system suggests: chips, cards, applied filters. The user confirms or dismisses, one gesture each. A confirm executes a typed decision and logs a positive label. A dismiss logs a negative one. The interface is not adjacent to the learning system; it **is** the training loop, collecting supervision from every session (contextual bandits doing exactly this run in production at Expedia and Wayfair). The one-to-three-hop budget is not a hope, it is what the loop converges to as labels accumulate.

This is also what users already expect. 71% of consumers expect personalized interactions and 76% get frustrated without them (McKinsey, *Next in Personalization*, 2021), and personalization leaders take 10-15% more revenue off the same traffic. The demand side of this pattern is not in question.

## The escape hatch: exact search

The tiers, in order:

1. **Suggest and confirm** (default): top-k chips, one to three hops, covers the large majority of sessions.
2. **Exact search** (rescue): when the chips miss, the user types the exact thing. With real semantic search this tier is reliable, and it absorbs everything the suggest loop misses. Everything else is mitigated by exact search.
3. **Classic navigation** (rare): keep a shrunken map for corner cases and power users. It stops being the default tax every session pays.

Note the inversion. In the old interface, navigation was the default and search was a helper bolted on. In the model-driven interface, intent is the contract and navigation is a rescue module.

## Wave two: intent to execution, and the failure mode

The current wave pushes the same idea past discovery into execution. ChatGPT reached a billion weekly users in September 2026: an intent box is now the most-used new interface of the decade. The commerce rails are being standardized as we write: OpenAI's Instant Checkout on Stripe's Agentic Commerce Protocol, and Google and Shopify's Universal Commerce Protocol launched at NRF 2026. China is furthest along: WeChat's mini programs already move more than RMB 8 trillion a year through services with no install and no app menus, Tencent is routing AI agents through WeChat as the control surface, and ByteDance's Doubao sits at roughly 380M monthly users with 79% penetration as a default entry point (Citi / QuestMobile).

The failure mode is equally documented, and it is the reason this wave is not yet settled. Humane's AI Pin shut down in February 2025; the Rabbit R1 was reviewed into the ground. Gartner expected at least 30% of generative AI projects to be abandoned after proof of concept by end-2025 and now puts the failure rate above 50%. The best mature chat deployments contain around two-thirds of conversations; weak ones never clear a quarter. Forrester finds consumer adoption of agentic payments low and stagnant while the infrastructure races ahead.

What separates the winners from the failures is not model size. It is whether the interface renders confidence. A feed wins because a recommender knows when it might be wrong and shows another card: the cost of a miss is one scroll. A chat-only interface hides its confidence, so the cost of a miss is the whole task. When the model drives, the sure things must arrive as executable decisions and the unsure things as one precise question, and there must always be a way out.

## Principles for model-driven UI

Collected from the two waves, the rules we develop against:

1. **Rank, don't tree.** If a model can order the options, do not bury them in a hierarchy. The taxonomy is training data and a fallback, not the interface.
2. **Suggest and confirm.** Every transition is a suggestion the user confirms or dismisses in one gesture. The budget is one to three hops.
3. **Every gesture is a label.** Confirms and dismissals train the ranker online. The UI is the training loop; there is no separate annotation pipeline.
4. **Render confidence.** What the model is sure of arrives as a typed, executable decision. What it is unsure of arrives as one structured question. Never fake certainty, never dead-end the user.
5. **Exact search is the escape hatch.** Always present, always one gesture away. It absorbs everything the suggest loop misses.
6. **Navigation is a rescue, not a default.** Keep a small map for the corner cases. Do not tax every session for the tail's needs.

## The demo

To demonstrate these principles we built a simple e-commerce storefront on open product datasets. One plain-words query in, results with apply-or-dismiss chips out; the unsure parts arrive as chips you can apply or drop, never as a wrong page. When the catalog has what the user means, the first query lands on it about 95% of the time, and the model runs on a CPU at the edge rather than behind an API. Wave one pointed these mechanics at content and kept the engagement for the platform. Pointed at transactions, the same mechanics hand the time back to the user.

## Sources

- SparkToro / Datos, “In 2026, Less than One Third of Google Searches Still Send a Click” (zero-click series, 2024-2026): <https://sparktoro.com/blog/in-2026-less-than-one-third-of-google-searches-still-send-a-click/>
- Gartner press release, February 2024: traditional search engine volume to drop 25% by 2026: <https://www.gartner.com>
- Gomez-Urbe, C. and Hunt, N., “The Netflix Recommender System: Algorithms, Business Value, and Innovation”, ACM TMIS 2015 (80% of hours from recommendations, stack valued at a billion dollars per year)
- Neal Mohan, YouTube CEO annual letter, February 2025 (recommendations drive the majority of watch time; TV is the primary US screen)
- Exploding Topics (Data Reportal figures), average daily TikTok time spent, 27 min (2019) to ~97 min (2025): <https://explodingtopics.com/blog/time-spent-on-tiktok>
- Econsultancy via Algolia, e-commerce site search benchmarks (4.63% vs 2.77% conversion): <https://www.algolia.com/blog/ecommerce/e-commerce-search-and-kpis->

## statistics

- McKinsey, “Next in Personalization” (71% expect, 76% frustrated, 10-15% lift):  
<https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/the-value-of-getting-personalization-right-or-wrong-is-multiplying>
- The Information, ChatGPT nears 1B weekly active users (Sep 2026):  
<https://www.theinformation.com/articles/openai-chatgpt-nears-1-billion-weekly-active-users-seven-months-target>
- OpenAI, “Buy it in ChatGPT” (Instant Checkout / Agentic Commerce Protocol):  
<https://openai.com/index/buy-it-in-chatgpt/>
- Forrester, agentic payments in B2C commerce (adoption low, stagnant):  
<https://www.forrester.com/blogs/agentic-payments-in-b2c-commerce-where-we-are-now/>
- CoinLaw, WeChat statistics compiled from Tencent filings (mini program GMV RMB 8T in 2024): <https://coinlaw.io/wechat-statistics/>
- Citi / QuestMobile via Honghe Business, China AI app rankings (Doubao ~380M MAU, 79% penetration): <https://www.honghebusiness.com>
- The Verge, “The Humane AI Pin never had a chance” (Feb 2025):  
<https://www.theverge.com>
- Gartner, “Top 10 Reasons Why Generative AI Projects Fail” (Nov 2025, >50% failure):  
<https://www.gartner.com>
- Expedia Group Tech, multi-variate web optimisation with linear contextual bandits:  
<https://medium.com/expedia-group-tech/multi-variate-web-optimisation-using-linear-contextual-bandits-567f563cb59>
- Wayfair Tech, contextual bandit for marketing treatment optimization:  
<https://www.aboutwayfair.com/careers/tech-blog/contextual-bandit-for-marketing-treatment-optimization>