heywa studio

Discovery IQ: The UK Retail Product Discovery Benchmark 2026

A study of 100 UK retail sites shows shoppers are well served when they can name a product and poorly served when they describe what they need.

"summer wedding guest dress under £30"

"christmas gift for my husband"

"champagne vs prosecco"

"my feet hurt in new shoes"

"stuff to stop my hair frizzing"

Going into Peak Trading season, Heywa Labs scored 100 UK retail sites on how well they help shoppers find, compare and choose.

43%

The average score from Discovery IQ, our measure on how well sites can deal with a range of different searches - from the specific to the more abstract.

What we did

Each site was tested with nine realistic searches, from an exact product name and a typo to a head-to-head comparison and a gift idea. Every result was rated on five dimensions, which gives the Discovery IQ score. We will run it again in six months.

At a glance

Sites find what shoppers can name. 73% returned a perfect result for an exact product name.

They lose shoppers who describe what they want. Only 11% did for an everyday problem like "my scalp is really flaky and itchy", and 66% failed outright. Not one of the 100 sites got a gift search or a head-to-head comparison right.

Some searches really missed the mark: "christmas gift for my husband" returned seven baby soft toys. "fleece under £100" returned an £8.75 headband.

Sometimes shoppers need bundles, not single items. A complete office outfit, a new shower, everything to put up a garden fence. Only 4% of sites gave a complete answer.

Nobody has this figured out. The Discovery IQ average is 43 out of 100 and the best is 67. We're at the beginning of the AI search & discovery and agentic commerce. We anticipate performance improving significantly, and quickly.

Why now

Agentic commerce, in which software agents search, compare and buy on a shopper's behalf, is arriving in the UK. Meta launched Muse, its personal agent, in the UK on 23 November 2026, with Christmas shopping among its headline uses [2]. An agent can only be as good as the site it searches. Gartner expects at least one new conversational AI interface to start disrupting traditional search and browse by the end of 2027 [1].

Run the Discovery IQ report on your own site heywalabs.com/discovery-iq

Foreword

“I need a waterproof jacket for the school run that I’d actually want to wear to work.”

That’s a perfectly ordinary thing to ask in a shop. Online, we’ve spent years turning requests like that into keywords and filters. As AI makes it easier to ask in our own words, shoppers will expect retailers’ sites to help them narrow the options and decide what to buy.

We wanted to see how well retailers could do that today. In the run-up to Black Friday, we tested 100 UK retail sites across ten sectors. The best scored 67 out of 100. Sites are good at understanding what a shopper asks for and far weaker at helping them choose. Nobody has this figured out yet, which means the best experience is still up for grabs.

Peak trading makes that gap especially visible. A shopper can find dozens of products in seconds and still be left wondering which one is right for them. I think the retailers that set the next standard for online shopping will be the ones that help people make that decision with confidence, even when they arrive with only a rough idea of what they want.

I hope this report helps you see where your own site does that well and where shoppers could use a little more help.

Milena Nikolic
Founder, Heywa Labs

Run the Discovery IQ report on your own site heywalabs.com/discovery-iq

Methodology

Heywa Labs built a tool called Discovery IQ, which scores how well a retail site helps shoppers find, compare and choose, out of 100. It scored 100 UK retail sites on 5 and 6 November 2026, three weeks before Black Friday (27 November). The sample is made up of leading UK retailers across 10 sectors (from beauty + fashion through to outdoor and DIY). We've kept all brands anonymous.

Each site was tested the way a shopper would use it, with nine searches that run from the easiest to the hardest for the brand to serve.

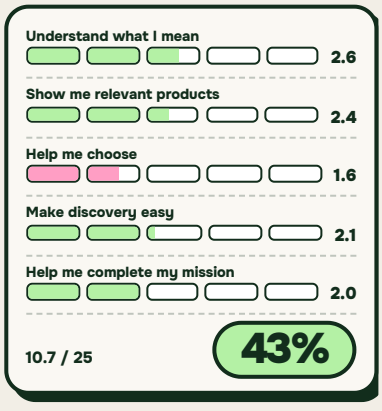
KIND OF SEARCH	WHAT THE SHOPPER IS DOING
1 **Exact name**	Types the full name of the product
2 **Typo**	Misspells the product or brand
3 **Everyday words**	Describes the product in plain words
4 **Details**	Adds specifics such as size, colour or material
5 **Budget**	Sets a price limit
6 **Everyday problem**	Describes a problem, not a product
7 **Whole job**	Asks for everything needed for a project or occasion
8 **Head-to-head**	Asks the site to compare two products
9 **Gift**	Shops for someone else

Each result was scored on five dimensions, each out of 5. The dimensions are the qualities of a search and discovery experience that matter most to a shopper, measured on every search and listed in the table below.

DIMENSION	WHAT IT MEASURES
1 **Understand what I mean**	Whether the site grasps the whole request, including limits such as price, not just the keywords.
2 **Show me relevant products**	Whether the products shown match what was asked for, when matching products exist.
3 **Help me choose**	Whether the site gives a relevant shortlist or targeted guidance, rather than generic suggestions.
4 **Make discovery easy**	How much extra work the shopper has to do, such as rewording the search or applying filters, to reach the right products.
5 **Help me complete my mission**	Whether the shopper ends up with what they set out to get, with their requirements applied.

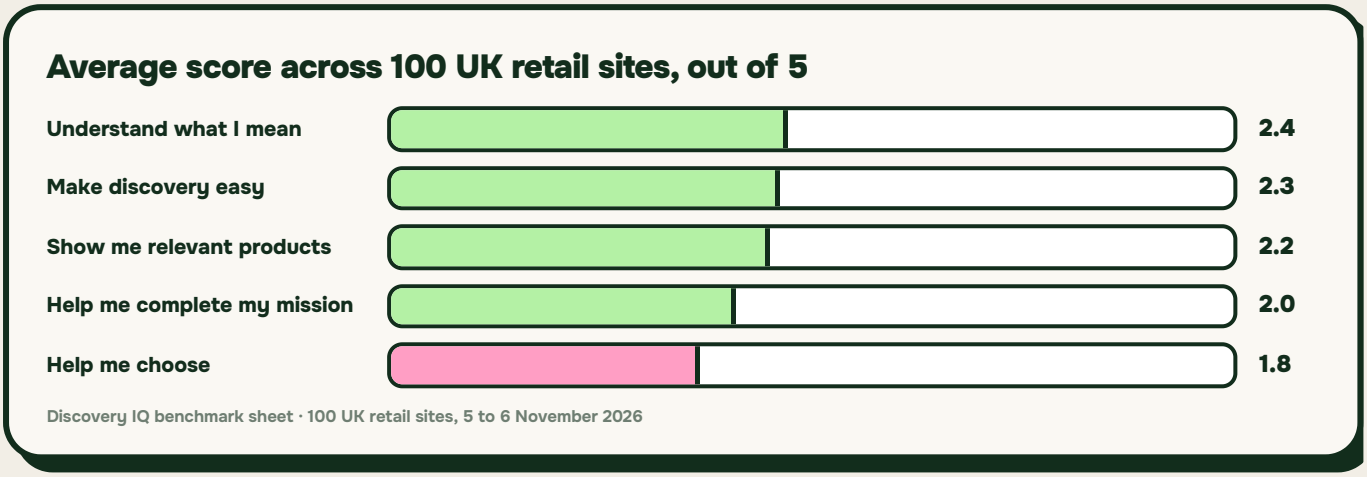
A dimension's score is its average across the nine searches, so it also sits between 0 and 5. The five dimension scores are added, out of a maximum of 25, and the Discovery IQ score is that total as a percentage. A site that scores 2.5 on every dimension therefore scores 50.

A worked example. Take a beauty retailer that scores 43. Each of its nine searches is graded out of 5 on all five dimensions. Averaged across the nine searches, it scores 2.6 on understanding, 2.4 on relevance, 1.6 on choosing, 2.1 on ease and 2.0 on completing the mission. Those five add up to 10.7 out of 25, giving a total Discovery IQ score of 43%.



Overall findings

The average site scores 43%. The best site scores 67%, only a handful get past 60% and about a third score below 40%. The advantages brands can gain from AI discovery & search are still in their infancy.



Three dimensions are relative strengths and two are clear gaps, but none overperform: no dimension averages above 2.4 out of 5. This is no surprise. AI search and discovery is still in its early days, so most sites are still testing these tools in the sandbox, with few really starting to roll it out across their entire site.

Gartner describes the same early stage in their June 2026 Magic Quadrant for Search and Product Discovery. They say ranking by relevance alone has become a commodity and is no longer enough, that the market is shifting towards detecting what a shopper intends, and that settled patterns for generative AI discovery interfaces are still elusive [1].

What sites do well

Understanding the intent of a customer who knows what they are looking for

Sites are good at qualifying demand the shopper already has: a clear request, typed in, even a messy one. **"nkie air max"** returned a page of Nike Air Max trainers with the spelling corrected in the heading, and **"hadly ottomen bed frame"** put the right bed frame first. This is the part conversational discovery will make more powerful, because it can read vaguer, longer and more natural requests.

Where improvements could happen: a shopper can land on a product that matches their search and still be left with questions. A shopper who searches **"my dog is scared of fireworks"** could reach a calming product and still wonder which one suits their dog. The gain is in guided selling: once a shopper reaches a product page that matches their search, a sales agent could answer the last few questions and help them settle.

Showing relevant products from slightly vaguer requests

Matching works about as well as understanding does. Shoppers can usually find relevant products. **"worktop for kitchen"** returned 481 worktops, all of them worktops, from £110 to £310, with filters for width, length and material. The trouble starts afterwards, when they have to pick one.

Where improvements could happen: relevance falls away once a request has a limit attached. **"fleece under £100"** returned a headband at £8.75 and **"white trainers for everyday"** returned five packs of white socks. Respecting budgets, sizes and product type is the quickest fix.

Make discovery easy

Search boxes, filters and navigation mostly do their job, which is why this scores second only to understanding. It is a relative strength, not a solved problem. When a search lands on the right products, most sites present them with filters that fit the category, such as width, length and material on worktops, or brand, colour and size on parkas. **"fence paint for wood"** returned exterior wood paints that were easy to narrow by colour or price.

Where improvements could happen: narrowing very large result sets. Guiding the shopper with a few smart questions, rather than a Sort and Filter button, would turn a search like **"going out top"**, which returned 8,437 results with only a Sort and Filter button, into a short list.

Where sites need to improve

Help me complete my mission

Shoppers arrive with a goal, such as a kitchen, a gift or a weekend away, and sites answer one query at a time. Only a handful of sites do this well. The thread of what the shopper is trying to do gets lost between steps. **"everything to put up a new garden fence"** returned 1,752 screens and panels with no posts, concrete or fixings. **"complete office outfit"** returned a single blazer. **"outfit for a winter wedding guest"** returned men's oxford shoes and a summer maxi dress. The best answer we saw was **"fit a new bathroom shower enclosure and tray"**, which put kits that bundle the enclosure, tray and waste first.

Bundling. Sites that put what a shopper needs for the whole job together, as the shower enclosure kits did, are the model. Fences, outfits and gift ideas are where most sites still return single items.

Help me choose

It is the weakest area for roughly three in four sites, and almost none do it well. Even the strongest sites are far better at understanding what a shopper wants than at helping them decide. Choosing takes reasoning that keyword search and filters were never built for: weighing trade-offs, explaining differences and recommending with reasons. That is where conversational, generative discovery adds the most, and few sites do it today. Across the sites we tested, none handled a head-to-head comparison cleanly. **"champagne vs prosecco"** returned "Nothing Found". **"nike air force 1 vs new balance 530"** returned a screen of Air Force 1s, with no sign of the other shoe until far down. **"cashmere jumper vs merino jumper"** returned jumpers but nothing on which is warmer or softer. Budgets fared little better: only a handful handled a price limit cleanly, and **"dresses under £30"** returned a pair of £32 trainers.

This is the largest gap. Explaining how two products differ, recommending one with a reason, and answering **"which is better for me"** in the results themselves would change the experience most. Gartner expects guided selling assistants, which let shoppers discover and choose products for themselves, to be built into the core discovery experience rather than sit apart from it [1].

What shoppers actually typed, and what came back

Each site was tested with nine real shopper searches, from exact product names to half-formed ideas. They tell one story: sites reward the shopper who knows exactly what they want and let down the one who is browsing.

When searches worked

Sites do well when the shopper names the thing, even badly.

Typos are often forgiven

🔍 **"patagnia torrentshel jacket"**

"patagnia torrentshel jacket" returned the right jackets in both men's and women's versions, with gender, size, price and colour filters. **"hadly ottomen bed frame"** put the right bed frame first, with 19 more ottoman frames beneath it. **"hemp hand protecter"** put the right product first, from £8.

Exact names land

🔍 **"werner timberline loft access kit"**

"werner timberline loft access kit" returned that kit first, at £129 with 351 reviews. **"adidas originals samba"** opened on the right shoe in three colours with 100 styles to browse.

Plain descriptions can work too

🔍 **"chew sticks for dogs that clean teeth"**

"chew sticks for dogs that clean teeth" opened on dental chews with sizes and prices. **"warm dressing gown for winter"** returned warm fleece gowns with sensible filters.

When searches broke

The failures were not random. They cluster around five kinds of request, and all of them are how people actually shop.

Situations, not products

"my mum has a sweet tooth and it's her birthday"

"my mum has a sweet tooth and it's her birthday" returned nothing but generic tips. **"starting out with a new puppy"** returned training pads and toys, with no starter set. **"my feet hurt in new shoes"** returned ordinary trainers and flats with nothing on comfort. **"housewarming gift for a new homeowner"** returned building sand, moisture meters and brooms. **"stuff to stop my hair frizzing"** ended on an empty page.

Words matched, meaning missed

"white trainers for everyday"

"white trainers for everyday" returned five packs of white socks. **"pjs for summer"** returned fabric by the metre and floral bedding. **"the ultmate gift card"** was rewritten by the site into **"tee ultimate gift card"** and returned t-shirts and a fleece.

Budgets read as words

"dresses under £30"

"dresses under £30" returned one result, a £32 pair of trainers. **"shampoo under £5"** returned 77 results where only one of the top three was under £5, and the price filter started at £0 to £15. **"moisturiser under £10"** was sorted A to Z with no price filter applied.

Comparisons treated as one query

"keter store it out max vs einhell lawn mower"

"keter store it out max vs einhell lawn mower" returned a storage box and a pile of lawn mowers, with no side-by-side. A head-to-head of two named sofas returned generic sofas, and one of the two was not in the results at all.

Dead ends on specific requests

"women's waterproof walking boots gore-tex lightweight"

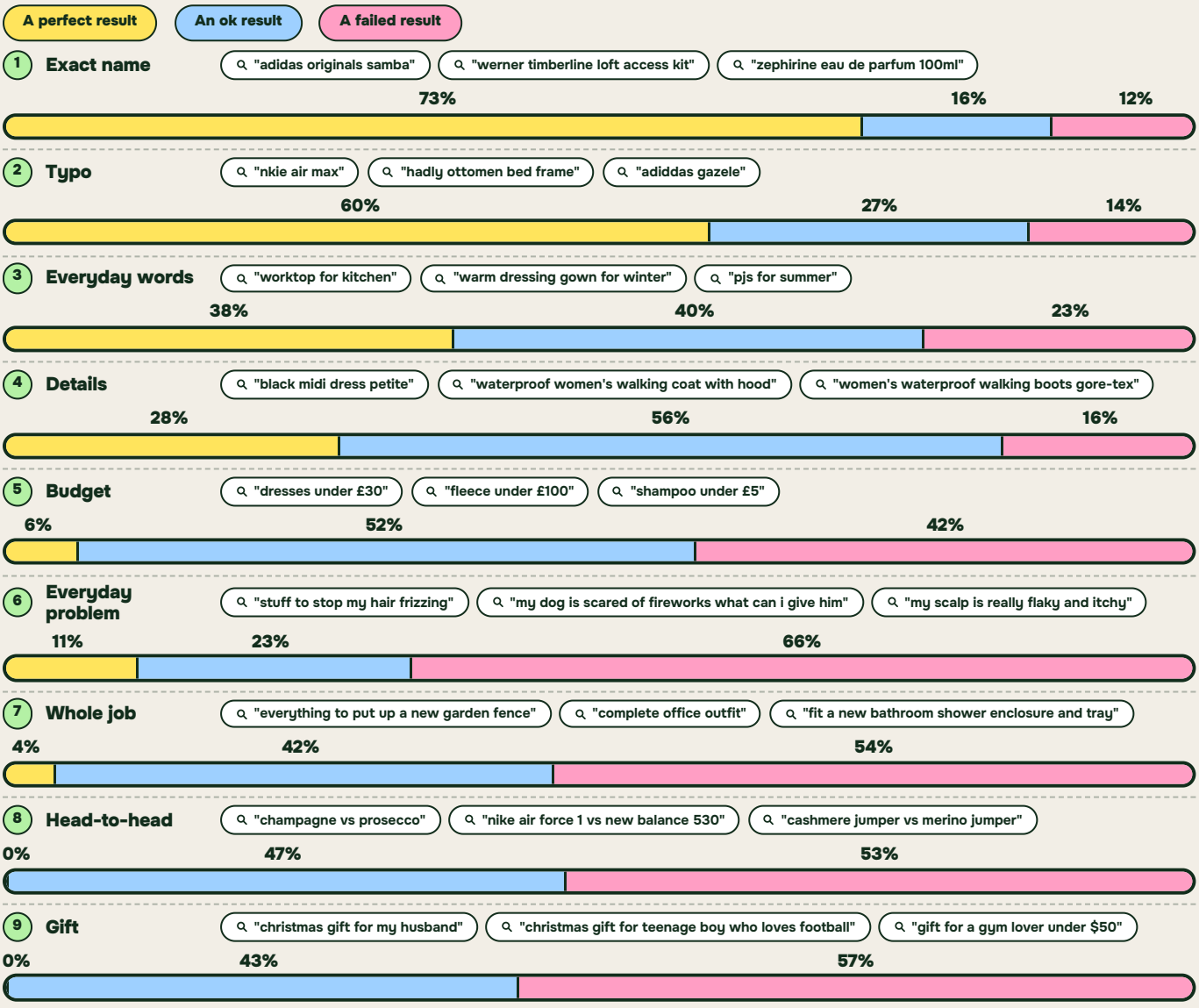
"women's waterproof walking boots gore-tex lightweight" returned a "no products matched" page with no boots on it. A full product name with dimensions for a 6ft fence panel returned a "please enter a valid search term" error. **"wooden garden fence"** returned 1,743 results that opened on decorative screens.

The more complex the search, the weaker the answers became

Every site was given the same nine searches, ordered from the simplest to the most complex. The first four are about matching: the shopper types a product name, a misspelling, a plain description or a few details, and the site only has to find those words. Most sites manage this well.

From the budget search down, the shopper is asking the site to understand something instead, such as a limit, a problem, a whole job, a choice or an occasion. Here results fall away quickly: on all five, no more than 11% of sites got a perfect result, and none did on a gift or a head-to-head comparison. On budgets, details, whole jobs and comparisons, roughly half of sites land on "an ok result": the right kind of product with the limit ignored, half the job missing, or the comparison left for the shopper to do.

Three real example searches are shown for each, and the table shows the share of sites that landed in each outcome, across every site tested on 5 and 6 November. These are the searches shoppers type when they are ready to buy, and on Black Friday they decide who wins the basket.



That is the same story the scores tell, seen from the shopper's chair. It is also what Black Friday looks like: gift hunts, "under £30" searches and "which of these two" questions. The searches that fail are the ones the sale will bring.

What great looks like: nobody is there yet

No site scored above 70, and none scored 4 out of 5 on a single dimension. The best experience in UK retail discovery has not been built yet. We are at the beginning of this journey.

Gartner makes the same point about the stakes. It says AI answer engines such as ChatGPT, Gemini and Perplexity are increasingly where people discover products, but that most traffic and transactions on commerce sites are still human, so brands that want to keep control of the experience and their data will need their own discovery experience, at least in the medium term [1].

What the best sites show us

Five sites reach 60 and the best scores 67. They sit in four different sectors, so no single category has cracked it. They win on breadth: the five do well on understanding, relevance, ease and mission. Even they are weakest at choosing. A score of 70 needs strength on every dimension, and the best site clears that bar on only two of the five.

Where the strongest sites pull ahead is on messy, real-life requests. **"my dog is scared of fireworks what can i give him"** returned calming supplements, chews and pastes. **"my scalp is really flaky and itchy"** returned the anti-dandruff shampoo, a scalp scrub and a scalp serum. Across the sites we tested, requests like these were let down more often than not: **"stuff to stop my hair frizzing"** returned a "no results" page. Even the leaders stumble on whole jobs: **"starting out with a new puppy"** returned training pads and toys, with no starter set.

What great would look like

This is our view of the target: a bar set well above anything in today's results, and not something the scores measure. Great is a site that thinks with the shopper, not one that waits to be told exactly what to find.

Understand what I mean

a half-formed thought works as well as a product name, and the site asks one sharp question when it needs more. A request like **"something for my dad, who walks a lot"** gets a short, confident list, not a keyword guess. Today, **"gift for dad who likes diy"** returns cement colourant, pipe lubricant and sabre saw blades.

Show me relevant products

results fit the person, not just the query. Budget, size, occasion and taste are all respected, and the first result is one the shopper would have picked themselves.

Help me choose

the site acts like a trusted expert. It compares the few options that matter, recommends one and says why in plain words. Today, none of the sites we tested does this cleanly for a plain "X vs Y".

Make discovery easy

no dead ends and no wall of 8,437 results. The shopper can refine, change their mind or switch how they ask without starting again.

Help me complete my mission

the site holds the whole goal from the first question to a finished basket, and finishes the job. **"everything for a weekend camping trip"** is the test: a complete kit, not a single tent.

What discovery outside ecommerce can teach retail

These are our analogies, not findings from the benchmark. The first three are the big shifts in AI, and the rest are the human-scale lessons.

Large language models have reset what people expect from a question

Shoppers who ask an AI assistant a plain sentence and get a reasoned answer will expect the same from a retail site. A results list that ignores what they typed will feel out of date.

Search is turning into asking

Conversational answer engines replace a list of links with one worked-out answer, and Gartner says they are increasingly where people discover products [1]. Retail sites now compete with that experience, not only with each other.

Assistants are moving from answering to doing

Agentic tools plan, compare and complete tasks for a person, and Gartner warns that the shift from answer engine to agent could threaten the traditional search and discovery market [1]. Meta launched its personal agent, Muse, in the UK on 23 November 2026, with Christmas shopping among its headline uses. Meta says that with some retail partners Muse can navigate the retailer's own website to find products, with the shopper approving each purchase [2]. A site that cannot hold a goal from start to finish will be bypassed by one that can.

A good shop assistant asks a couple of questions before recommending, then says why this one

Retail sites mostly show a list and stop.

A sommelier or stylist offers two options, not twenty

Fewer, better choices, matched to taste, budget and occasion.

A travel planner sells the trip, not the flight

The goal is the unit of work, which is the mission dimension.

Streaming and music services learn from what you do

Each suggestion gets better with use, so discovery improves the longer a shopper stays.

The lessons for future seasonal periods: protect your strengths, fix your weaknesses

Then focus far more on the weaknesses

Choosing and completing the mission are where vague intent can be turned into a sale. When prices are close and similar deals sit side by side, the site that helps a shopper decide takes the sale, and the site that leaves them to compare alone loses them to another tab. Gifting makes it sharper still: a shopper buying for five people has a mission, not a query. A site that can hold that thread keeps the whole basket. Black Friday searches look like the ones that failed. **"christmas gift for my husband"** returned seven baby soft toys. **"christmas gift for teenage boy who loves football"** returned everyday sneakers, baby clothes and junior trainers. **"leggings under $60"** returned "no results", and **"gift for a gym lover under $50"** found nothing at all.

Three lessons

1. **Treat choosing as the product**

   Help shoppers compare, narrow down and decide inside the site, not across five open tabs. **"keter store it out max vs einhell lawn mower"** returned a storage box and a pile of lawn mowers, with no side-by-side.

2. **Serve the mission, not just the query**

   Carry what the shopper is trying to do from one step to the next. **"everything for a cosy hot chocolate night in"** returned only boxed chocolates that matched the word "box".

3. **Spend where the gap is**

   Most investment has gone into finding products. The scores suggest the next return is in helping shoppers choose and finish.

Where brands have most to gain

Choosing has the most headroom: more than half of sites score poorly on "Help me choose", and more than one in four score very poorly.

Completing the mission comes next. Most sites lose the thread, and only a handful finish the job well. Searching well is not the same as finishing the job, and a shopper who arrives with a goal rather than a product name needs the site to hold the thread.

The market is moving this way. Gartner predicts that by the end of 2027, at least one new AI-based conversational interface will disrupt traditional search and browse, reaching at least 10% of a market it valued at nearly $18 billion in 2025 [1].

heywa studio

How Heywa Studio can support from here

Heywa Studio is built for the stage of discovery this benchmark finds weakest: helping a shopper go from "I'm looking for something" to "this is the one."

Studio is the AI search & discovery platform from Heywa Labs. Instead of a results page that looks the same for every query, the experience assembles itself around what the shopper is trying to do from their catalogue. That maps onto the three dimensions with the most headroom:

Understand and show relevant

Shoppers describe what they want in their own words, and the catalogue responds to that.

Help me choose

Options can be set side by side, with a clear reason why they differ.

Complete my mission

The conversation keeps the shopper's goal in view from first question to finish.

The simplest place to start is to find out where your own site stands. We can score any retail site on the same five dimensions used here. From there, we can show what a Studio experience looks like on your catalogue.

Reach out to us at heywalabs.com/studio

Sources

[1] Gartner, **Magic Quadrant for Search and Product Discovery**, Mike Lowndes, Sandy Shen and two other analysts, 22 June 2026, document ID G00839916. Gartner describes its research as the opinions of its research organisation, not statements of fact.

[2] Retail Insight Network, **Meta launches Muse AI agent in UK with several retail partners**, November 2026 (retail-insight-network.com). Launch date and approval requirement as reported in UK trade press.

Discovery IQ benchmark. Heywa Labs, Discovery IQ benchmark of 100 UK retail sites across 10 sectors, tested on 5 and 6 November 2026, three weeks before Black Friday. Each site was scored on nine searches across five dimensions, and brands are anonymised.