

FOR RELEASE MAY 23, 2025

What Web Browsing Data Tells Us About How AI Appears Online

An analysis of one month of browsing data finds that around six-in-ten respondents visited a search page with an AI-generated summary. But visits to more in-depth content about AI were relatively rare.

BY Athena Chapekis, Anna Lieb, Sono Shah and Aaron Smith

FOR MEDIA OR OTHER INQUIRIES:

Aaron Smith, Director, Data Labs
Sogand Afkari, Communications Manager

202.419.43

www.pewresearch.org

RECOMMENDED CITATION

Pew Research Center, May 2025, "What Web Browsing Data Tells Us About How AI Appears Online"

About Pew Research Center

Pew Research Center is a nonpartisan, nonadvocacy fact tank that informs the public about the issues, attitudes and trends shaping the world. It does not take policy positions. The Center conducts public opinion polling, demographic research, computational social science research and other data-driven research. It studies politics and policy; news habits and media; the internet and technology; religion; race and ethnicity; international affairs; social, demographic and economic trends; science; research methodology and data science; and immigration and migration. Pew Research Center is a subsidiary of The Pew Charitable Trusts, its primary funder.

© Pew Research Center 2025

How we did this

The goal of this study is to better understand how, where and in what context Americans are encountering artificial intelligence as they browse the web. For this analysis, we purchased the March 2025 web browsing data of 900 U.S. adults. Each respondent is a member of KnowledgePanel Digital, an online survey panel whose respondents agree to install an app that tracks their online browsing behavior.

For each respondent, we received a list of all the URLs they visited using a web browser on a tracked device during the March 1-31, 2025, study period. For each of these approximately 2.5 million visits to 1.1 million unique URLs, we ran an automated script that retrieved the text a visitor to that URL might see, along with the page title and description.

First, we wanted to see how many of these URLs mention AI anywhere on the page. To do that, we checked if the page text included any on a long list of AI-related terms (refer to [Appendix A](#) for the full list of terms).

This is our broadest possible measure of whether a page mentions AI or offers some sort of AI functionality. It captures anything from a website for a generative AI tool like chatgpt.com, to a website with a single reference to AI on the page.

In addition to this broad read on how many webpages mention AI at all, we also wanted to see how many mention AI more prominently or indicate an active interest in AI on the part of the visitor. Some examples might include searches for AI-related terms on a search engine, or visits to sites like:

- A generative AI tool such as chatgpt.com
- A news article that focuses on or extensively discusses AI-related issues
- A page for a product or service that extensively discusses its AI features

To identify these pages, we trained a model to indicate whether each page that mentioned AI did so in an incidental context or in a more substantive context along the lines of those listed above.

Although this study is based on the actual webpages respondents visited, it does have limitations that should be taken into consideration when interpreting its findings. First, the respondents were required to download a VPN to register personal devices for use in this study. As such, browsing on work devices or other untracked devices they might use to access the internet may not be captured. Second, the data only covers activity in web browsers, not other applications. Third, it does not include real-time algorithmically curated content (like the “For You” pages on X or TikTok) or pages that require a login to access.

What Web Browsing Data Tells Us About How AI Appears Online

An analysis of one month of browsing data finds that around six-in-ten respondents visited a search page with an AI-generated summary. But visits to more in-depth content about AI were relatively rare.

As artificial intelligence (AI) advances rapidly and becomes increasingly commonplace in online spaces, [the American public has shown a clear wariness](#) about its potential impact on individuals and society.

Public opinion polling can help us understand Americans' broad usage patterns and attitudes about AI, but we have less insight into what they see and read about this technology as they go about their daily online lives. To help shed light on this, we examined roughly 2.5 million webpage visits to 1.1 million unique URLs from a group of 900 U.S. adults who agreed to share their March 2025 browsing data.

Here are our main findings about how ordinary internet users are crossing paths with AI-related tools and content in their online browsing habits:

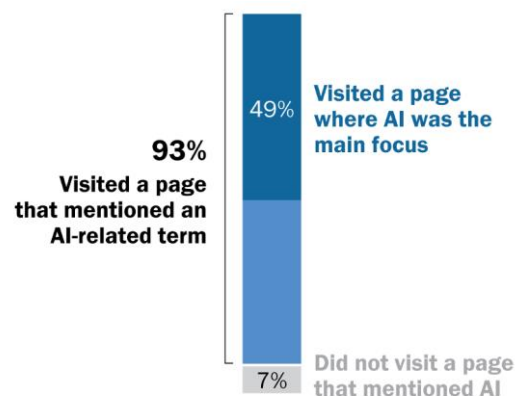
Most respondents encountered at least some AI-related content during the month – although most references were not especially prominent or in-depth.

The vast majority (93%) visited at least one page that mentioned AI, even if that mention was brief or unrelated to the main focus of the page.

This share drops substantially for the subset of pages that *primarily focus on AI* – like the website of an AI chatbot, or a news article that extensively discusses AI-related issues. Around

Most respondents visited at least 1 page that mentioned AI, but they did not often visit pages focused on AI

% of respondents who ____ in March 2025



Of those who **visited pages that mentioned AI**, they typically visited that type of page ...

60 times

Of those who **visited pages where AI was the main focus**, they typically visited that type of page ...

3 times

Note: Shares may not total 100% due to rounding.

Source: Pew Research Center analysis of online browsing activity from 900 U.S. adults. Analysis includes URLs that respondents accessed March 1-31, 2025, using a web browser on a personal desktop, laptop or mobile device. Webpage data was collected April 7-17, 2025.

"What Web Browsing Data Tells Us About How AI Appears Online"

PEW RESEARCH CENTER

half of all respondents (49%) visited one such page during the month, though the typical person who looked at this type of content did so only a handful of times.

This pattern is found across several different types of web browsing activity, including querying search engines, reading the news and shopping online.

A small share of respondents used search engines to get *information about AI*. One-in-ten conducted an AI-related query on a search engine during the month.

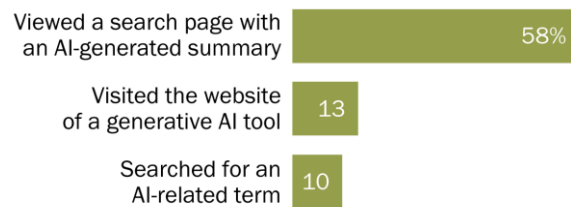
But more often, search engines were a source for material *generated by AI*.

The vast majority of respondents conducted a search using a search engine like Google. And for 58% of respondents, at least one of these searches produced an AI-generated summary along with the traditional search results.¹

Some 13% of respondents visited a website for an AI tool like a chatbot, image generator or other generative AI tool.

Around 6 in 10 respondents visited a search result page with an AI-generated summary

% of respondents who ____ in March 2025



Note: Findings for AI-generated summaries include only searches on google.com.

Source: Pew Research Center analysis of online browsing activity from 900 U.S. adults. Analysis includes URLs that respondents accessed March 1-31, 2025, using a web browser on a personal desktop, laptop or mobile device. Webpage data was collected April 7-17, 2025.

"What Web Browsing Data Tells Us About How AI Appears Online"

PEW RESEARCH CENTER

To the extent that respondents visit news pages mentioning AI, such pages don't typically discuss the technology at length. Around half of all respondents (52%) visited at least one page from a news outlet that contained an AI-related term. But most of these articles only mentioned AI in passing – for example, a brief reference to an AI company in a story about the stock market, or an AI-related term in a sidebar to the main article.

A far smaller share of respondents (8%) visited a news article in which AI was discussed in depth or was the primary focus of the story.

Respondents tended to visit pages that describe AI using generic terms, brand names, marketing language and product descriptions. The most common AI-related terms were broad references like "AI" or "artificial intelligence." Other relatively common terms

¹ Due to technical limitations in identifying the presence of AI-generated summaries on other search engines, findings for AI summaries include only searches on google.com

included “ChatGPT” and “OpenAI;” references to products “powered,” “assisted” or “enhanced” by AI; and “AI images” or “AI image generator.”

To be sure, these findings paint only a partial picture of all the ways Americans might encounter references to AI in their daily lives. Notably, our analysis does not capture things like content in mobile apps or web browsing on untracked devices – to say nothing of the AI references people might come across on TV or in their day-to-day interactions with others. But these findings help add context to how the public is learning about this complex and rapidly changing topic.

How common are visits to websites that mention AI?

The 900 respondents in this study visited a very large number of webpages during the month of March: 2,457,176 in total. Of these roughly 2.5 million page visits, 7% were to a page that contained at least one AI-related term.

Nearly every member of the study (93%) visited at least one page that mentioned an AI-related term. Among this group, the typical (median) respondent visited this type of page 60 times during the month.² And as is true of many online behaviors, some respondents visited these pages far more than others: The top 20% of respondents by browsing activity visited pages that mention AI about 300 times, on average.³

Respondents across a range of demographic groups are equally likely to see this sort of content, with no major differences by age, educational attainment or gender.

In-depth mentions of AI

Some pages that mention AI may further:

- **Highlight AI as a central focus of the page.** Examples include an article that discusses new developments in AI technology and their implications for the global economy, or a story that talks extensively about a controversial AI-generated viral video.
- **Indicate an active interest in AI on the part of the person visiting the page.** This could include things like a search query for an AI-related topic, or a visit to an AI tool like chatgpt.com.

Visits to these kinds of pages accounted for fewer than 0.05% of all page visits in this analysis. Although these visits are a tiny share of respondents’ web browsing overall, around half of

² Page visits include instances where respondents visited the same URL more than once.

³ Overall browsing activity is measured by the total number of page visits.

respondents (49%) visited *at least one* such page over the course of the month. However, this was not a central part of most respondents' web browsing activity. The typical respondent who visited at least one of these pages went to just three in total during the study period.

Mentions of AI on different types of sites

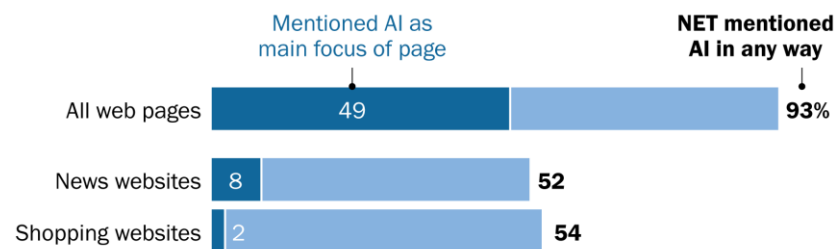
The roughly 2.5 million page visits in this study encompass a wide range of web browsing habits. Here are some of the ways that AI appears in specific corners of the web, based on the browsing data in this analysis. Read the [methodology](#) for more details about the specific sites that were included in each category.

News websites

Around half of the respondents in this study (52%) visited a news webpage that mentioned AI. By and large, these pages tend to mention AI only in ways that are incidental to the main story or article. This might include a single mention of AI in an otherwise unrelated story, or an AI-related term in a recommended headline on the page's sidebar.

Relatively few respondents visited a news article that discussed AI in a meaningful context

% of respondents who visited at least one webpage on the following types of sites that ____ in March 2025



Source: Pew Research Center analysis of online browsing activity from 900 U.S. adults. Analysis includes URLs that respondents accessed March 1-31, 2025, using a web browser on a personal desktop, laptop or mobile device. Webpage data was collected April 7-17, 2025.

"What Web Browsing Data Tells Us About How AI Appears Online"

PEW RESEARCH CENTER

Just 8% of respondents visited a news webpage that discussed AI in a meaningful context. Some examples actually visited by study participants include:

- An article about the U.S. stock market that [discusses how graphics processing units made by the hardware company Nvidia](#) are used in modern AI applications
- An article about Hollywood celebrities [petitioning President Donald Trump on potential AI copyright regulations](#)
- A story about a man who filed a complaint because [an AI chatbot falsely indicated he had committed a serious crime](#)

Shopping websites

Around half of respondents (54%) went to a page on one of 18 major shopping websites that contained a reference to AI. These references were often mentions of AI-powered product features or AI-generated summaries of customer reviews.

Amazon.com, a domain visited by 63% of respondents in March, accounts for a large portion of pages in the latter category. One of the most-visited sites in the world, Amazon has [AI-generated summaries of customer reviews](#) for many products.

Social media

A great deal of content on social media today is algorithmically curated or requires a login to access it. As such, much of the social media content viewed by respondents was inaccessible to our data collection methods. Even so, three-quarters of all respondents visited a social media page with an AI reference. Many of these visits were to pages on facebook.com that, during the study period, contained a link to the company's Meta AI tool in the sidebar.

Search results and AI-generated search summaries

Around six in ten respondents (58%) conducted at least one search engine query that produced an AI-generated summary of results.⁴ Around two-thirds (65%) conducted a search that produced an AI reference anywhere on the results page, regardless of whether the search itself pertained to AI or not.

One-in-ten respondents made a search that directly related to AI. The most common queries involved various combinations of the words “AI” and “ChatGPT.”

Generative AI tools

Some 13% of respondents – and 20% of those ages 18 to 29 – visited the website of an AI chatbot at least once during the month of March. This category includes tools like OpenAI's ChatGPT, Google Gemini and Perplexity AI.

[In an August 2024 Pew Research Center survey](#), one-third of U.S. adults reported having *ever* used an AI chatbot.

⁴ Due to technical limitations in identifying the presence of AI-generated summaries on other search engines, findings for AI summaries include only searches on google.com.

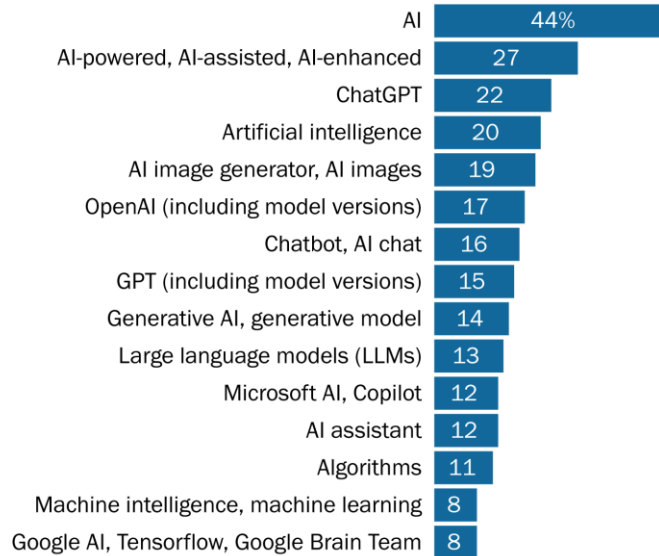
Language used on pages mentioning AI

The largest shares of respondents in this study visited pages that mentioned AI using generic descriptors like “AI” or “artificial intelligence,” as well as marketing phrases like “AI-powered,” “AI-enhanced” or “AI-assisted.” And around one-in-five visited a page that mentioned ChatGPT specifically.

At the page level, the distinct pages in this collection with *any* mention of AI included an average of two unique terms and three total terms [from our list](#). But around one-in-five of these pages (22%) only match for the single term “AI.”

Respondents tended to visit pages mentioning AI in generic terms, marketing language, brand names

% of respondents who visited a webpage that mentioned ____ in March 2025



Note: Term matches are not mutually exclusive. For example, a page that mentions “AI images” also counts as a page that mentions “AI.”

Source: Pew Research Center analysis of online browsing activity from 900 U.S. adults. Analysis includes URLs that respondents accessed March 1-31, 2025, using a web browser on a personal desktop, laptop or mobile device. Webpage data was collected April 7-17, 2025.

“What Web Browsing Data Tells Us About How AI Appears Online”

PEW RESEARCH CENTER

Acknowledgements

This report is a collaborative effort based on the input and analysis of the following individuals:

Primary researchers

Athena Chapekis, *Computational Social Science Analyst*

Anna Lieb, *Computational Social Science Assistant*

Sono Shah, *Associate Director, Data Labs*

Aaron Smith, *Director, Data Labs*

Research team

Regina Widjaya, *Computational Social Scientist*

Galen Stocking, *Senior Computational Social Scientist*

Brian Broderick, *Former Senior Data Engineer*

Gonzalo Rivero, *Former Associate Director, Data Labs*

Editorial and graphic design

Alissa Scheller, *Senior Information Graphics Designer*

Peter Bell, *Associate Director, Design and Production*

Anna Jackson, *Editorial Assistant*

Communications and web publishing

Sogand Afkari, *Communications Manager*

Mithila Samak, *Communications Associate*

Janakee Chavda, *Associate Digital Producer*

In addition, the project benefited greatly from the guidance of Courtney Kennedy, Andrew Mercer, Arnold Lau, Monica Anderson, Eileen Yam, Colleen McClain, Jeffrey Gottfried and Haley Nolan.

Methodology

The findings in this report are based on an analysis of the browsing behaviors of 900 U.S. adults who are members of the KnowledgePanel Digital online panel, a subset of Ipsos' KnowledgePanel. These panelists qualified for the study because they:

- Responded to a Pew Research Center pilot survey conducted on KnowledgePanel Digital in November 2023 (n=1,254)
- Were still active members of both KnowledgePanel (n=1,206) and KnowledgePanel Digital (n=1,102) as of March 2025
- Consented in 2025 to Ipsos providing their individual-level browsing data to clients (n=992)
- Were active on KnowledgePanel Digital at least once between March 1 and March 31, 2025 (n=900)

After the March monitoring period, panelists' web browsing activity logs were delivered on April 7, 2025. The dataset containing these logs included 2.5 million visited URLs with metadata, including an ID for the panelist who visited the URL, device information, the time when the URL was accessed and the duration of the visit.

About IPSOS KnowledgePanel and KnowledgePanel Digital

KnowledgePanel is a probability-based online panel designed to be representative of the adult U.S. population. The recruitment process employs an address-based sampling methodology from the latest Delivery Sequence File of the USPS, which has been estimated to cover as much as 98% of the population, although some studies suggest that the coverage could be in the low 90% range.⁵ Additional details about the KnowledgePanel can be found [here](#).

KnowledgePanel Digital consists of members of the broader KnowledgePanel who meet the following conditions:

- Access the internet using an Android smartphone, Apple smartphone, Android tablet, Apple tablet, Windows computer or Apple computer.
- Have agreed to join KnowledgePanel Digital and have installed the RealityMeter app on at least one qualifying device to collect their device usage and internet activity.
- Have agreed to the [RealityMine Privacy Policy](#) and Terms & Conditions.

⁵ AAPOR Task Force on Address-based Sampling. 2016. "[AAPOR Report: Address-based Sampling](#)."

For households without internet service prior to joining KnowledgePanel, Ipsos provides web-enabled devices and free internet service. KnowledgePanel members from these households are not invited to join KnowledgePanel Digital.

Weighting

The data was weighted in a process that accounts for multiple stages of sampling and nonresponse that occur at different points in the panel survey process. First, each panelist begins with a base weight that reflects both their probability of recruitment into the panel and their probability of selection for this survey. Base weights for this study were provided by Ipsos. Next, these weights were calibrated to align with the population benchmarks in the accompanying table and trimmed at the 1st and 99th percentiles to reduce the loss in precision stemming from variance in the weights. Sampling errors and tests of statistical significance take into account the effect of weighting.

Weighting dimensions

Variable	Benchmark source
Age (detailed)	2022 American Community Survey (ACS)
Age x Gender	
Education x Gender	
Education x Age	
Race/Ethnicity x Education	
Born inside vs. outside the U.S. among Hispanics and Asian Americans	
Years lived in the U.S.	
Census region x Metropolitan status	
Volunteerism	2021 CPS Volunteering & Civic Life Supplement
Voter registration	2022 CPS Voting and Registration Supplement
Party affiliation x Race/Ethnicity	2023 National Public Opinion Reference Survey (NPORS)
Party affiliation among registered voters	
Frequency of internet use	
Religious affiliation (Protestant, Catholic, Religiously unaffiliated, Other)	

Note: Estimates from the ACS are based on noninstitutionalized adults. Voter registration is calculated using procedures from Hur, Achen (2013) and rescaled to include the total U.S. adult population.

PEW RESEARCH CENTER

The following table shows the unweighted sample sizes and the error attributable to sampling that would be expected at the 95% level of confidence for different groups in the survey.

Sample sizes and margins of error

Group	Unweighted sample size	Plus or minus ...
Total sample	900	4.9 percentage points
Age 18-29	91	13.7 percentage points
Age 30-49	315	7.9 percentage points
Age 50-64	290	8.9 percentage points
Age 65+	204	10.2 percentage points

Note: Unweighted sample sizes do not account for the sample design or weighting and do not describe a group's contribution to weighted estimates. See the Weighting section above for details.

PEW RESEARCH CENTER

Data collection and processing

Before collecting webpage content, we processed the dataset of URLs to prevent errors in the scraping process and avoid potentially harmful or explicit websites.

To ensure that each panelist meaningfully viewed each page in the dataset, we removed URL visits where the duration of the visit was zero seconds. We also combined “duplicate” records where a panelist visited the same URL twice within a one-second window.

For each URL in the dataset, we extracted a domain (for example, the URL “https://www.facebook.com/login” has the domain “facebook.com”). Based on the URLs and their domains, we filtered out certain categories of webpages that we were not interested in scraping. We excluded the following URL categories from this scraping process:

- **Known malware.** To identify malicious websites that are known to be used for malware distribution, we used a freely available [list of malware URLs](#). The list is compiled by [URLhaus, a platform that tracks malware URLs](#) and shares them with security providers.
- **Adult content.** Adult content websites were categorized as websites whose main purpose is to host pornographic or highly sexualized content, including file-sharing sites that are regularly used for pornography distribution. To identify adult websites, we used [an open-source URL list that is maintained primarily for network administrators at schools and workplaces](#).

- **Popular productivity tools that require a login.** Personal productivity tools like email and calendar apps were excluded, since the content we would be scraping was almost certain not to reflect the page content the panelist had viewed.

We removed these types of pages from the web scraping pipeline and did not check them for AI keywords. However, these pages were still included in each respondent's total visit count.

Scraping webpage content

After preprocessing the URLs from the web browsing activity logs to remove the pages which matched the criteria above, we used a web scraping pipeline based on the Requests Python library to retrieve all HTML and any metadata from each page. We collected this content April 7-11, 2025.

In some cases, the attempt to access a URL would time out or our servers otherwise failed to connect to the given web address. This includes those that returned HTTP error codes, such as 404 (Not Found) or 403 (Forbidden) errors. These pages were also not included in the AI keyword analysis but included in each respondent's total visit count.

There were a handful of pages that returned HTTP 403 (Forbidden) errors that we determined were important to try to retrieve for the analysis, including a number of AI chatbots (such as Google Gemini and Grok) and pages from Reddit. These URLs were backfilled April 16-17, 2025, using manual HTML download and the Reddit Data API, as applicable.

URLs that did not resolve or redirect to an existing webpage, according to [a list of valid top level domains](#), could not be verified as a valid visit to a page and were excluded from respondents' total visit count (n=2,402 URLs).

In addition to pages noted above that we did not scrape, the following types of content may have been visible to respondents but would have been missed by the automated scraping process:

- Non-text content like images, audio or video
- Content that loads dynamically using JavaScript
- Content hidden behind a paywall or login screen

Google search result pages are not retrievable through most traditional web scraping methods, including the one we implemented in this analysis. Because of this, we collected the content of Google search result pages using a third-party web scraping service.

After all data collection and preprocessing was complete, the total number of distinct URLs we were able to retrieve page content for and analyze for AI term matches was 965,136. A total of 2,457,176 page visits to 1,107,424 URLs were analyzed in the report. This includes 142,288 URLs that we were not able to analyze for AI term matches but included in the analysis of respondents' total page visits.

Identifying webpages related to AI

Preprocessing

To prepare the page content data (n=965,136) for analysis, we removed HTML tags, JavaScript and other code from the scraped webpage data using BeautifulSoup. By doing this, we ensured the page content that was analyzed was as close as possible to the content that panelists would have viewed when they visited the webpage.

Any scraped webpages that contained excessively large amounts of data (greater than 1 gigabyte or 128,000 tokens) were also excluded from the web content dataset. These websites often contained video data or other non-text content.

AI keyword matching

To identify AI content in the webpage data, we started by compiling a list of AI-related terms (read [Appendix A](#)). These terms included technical AI terminology, as well as the names of various generative AI tools and well-known AI companies. Keywords were selected based on their specificity and their prevalence in current discussions of AI. To avoid returning matches that were not truly related to AI, we filter this list to only terms that are exclusively (or nearly exclusively) used to describe AI-related concepts.

All webpages that contained text matching at least one keyword in the list were flagged as a page that mentions AI. Of the roughly 1.1 million distinct pages viewed by panelists over a period of one month, around 6% of them (71,144 in total) mention an AI keyword.

AI relevance classification

While keyword matching was used to identify pages with *any mention* of AI-related terms and concepts somewhere on the page, these webpages are extremely diverse in their focus on the topic. Some make only the briefest mention to AI or AI-related topics – such as a reference to an AI tool in a sidebar or footer. Others contain meaningful or substantive discussion of AI or are themselves AI-centric tools or services.

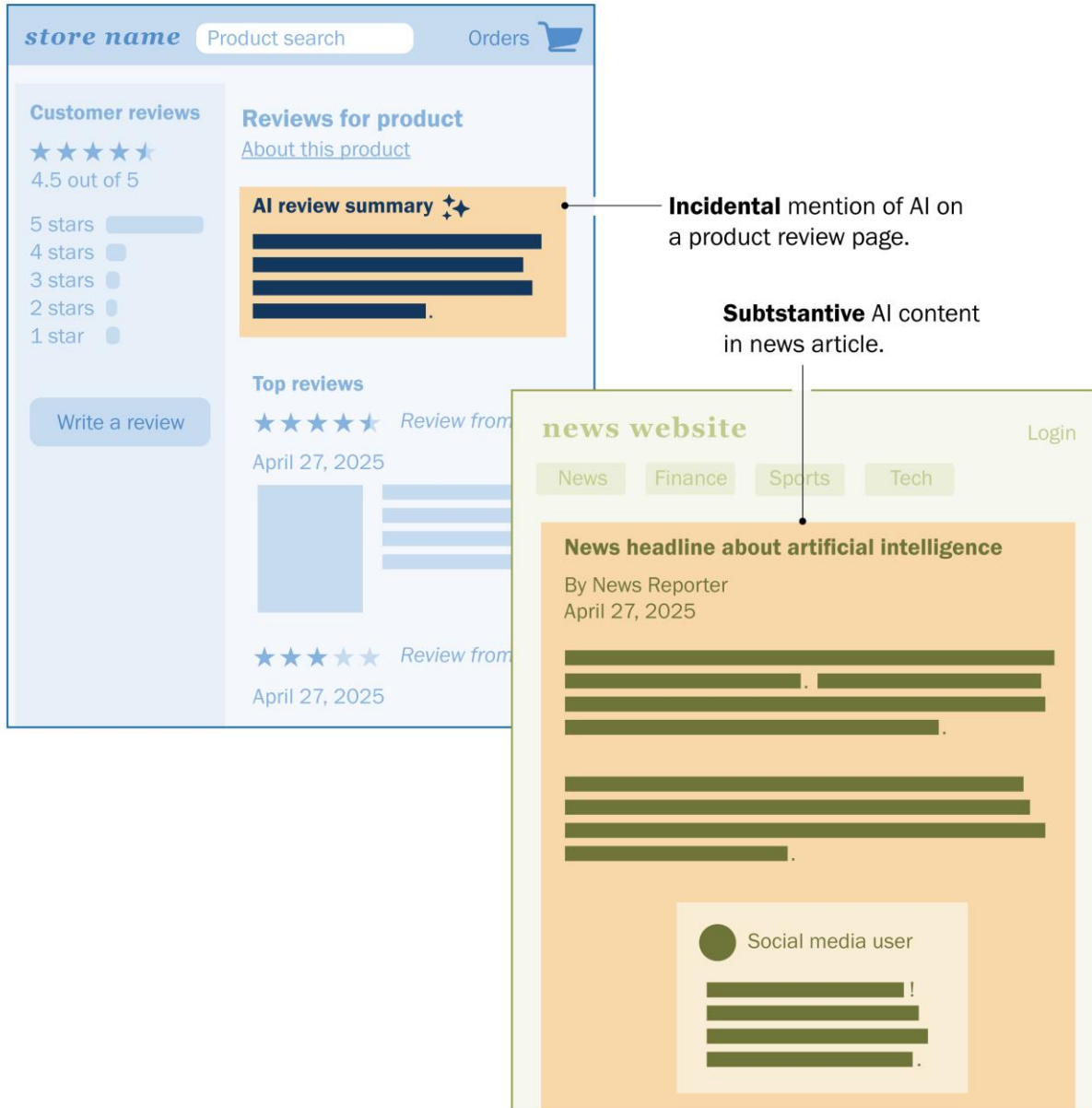
For this analysis, we used a logistic regression classifier to identify pages that mention AI in a substantive context. Common examples of “substantive mentions” might include pages such as:

- An AI chatbot interface, AI image generation site, AI-focused consulting service or other online tool that makes AI a central focus of its functionality
- A shopping website that prominently features AI functionality in its product descriptions
- An article or story that is about (or extensively discusses) AI or AI-related topics

The classifier was trained on a dataset of 509 webpages labeled as containing either a substantive AI mention or a minor AI mention. Two human annotators generated the labels and achieved high inter-rater reliability on the training dataset (Cohen's kappa = 0.877). The classifier took five input variables as features: the number of AI keyword matches, number of AI keyword matches *that were not* "AI," number of AI keyword matches in the website's title, number of AI keyword matches in the website's description, and the proportion of all words on the page that are AI keywords.

After training the classifier, model performance was measured using a separate evaluation dataset of 400 labeled webpages. For this dataset, human annotators again achieved high inter-rater reliability (Cohen's kappa = 0.837). When applied to the evaluation dataset, the classifier achieved an F1 score of 0.829. The model was highly reliable for identifying substantive AI mentions but was somewhat prone to false positives; the classifier's performance on the evaluation data gave a recall score of 0.970 and a precision score of 0.724. These metrics indicate that the model was sufficient for use in this report.

A page that mentions AI versus one that discusses it in more depth



"What Web Browsing Data Tells Us About How AI Appears Online"

PEW RESEARCH CENTER

Domain and page content analysis

In this report, we examine how respondents encountered mentions of AI on different types of websites. Here is how we created the different site categories for that analysis.

News websites include 2,317 domains categorized as “News/Information” by the measurement and audience metrics company [comScore](#).

Shopping websites include a list of 18 major shopping and e-commerce domains developed by the authors of this study. Researchers consulted two lists of popular shopping and e-commerce domains (via [Statista](#) and [Semrush](#)) and cross-referenced them against respondents’ most-visited domains. Our final list includes aliexpress.com, amazon.com, apple.com, bestbuy.com, chewy.com, costco.com, craigslist.com, ebay.com, etsy.com, homedepot.com, kohls.com, macys.com, mecari.com, target.com, temu.com, ticketmaster.com, walmart.com, and wayfair.com.

Search websites include google.com/search, bing.com/search, duckduckgo.com and search.yahoo.com. Analyses of AI-generated summaries include only pages from google.com/search.

Social media websites include Meta sites (facebook.com, instagram.com, threads.net, whatsapp.com), youtube.com, tiktok.com, bsky.app, pinterest.com, linkedin.com, x.com and reddit.com.

Visits to generative AI tools include those from OpenAI (openai.com, chatgpt.com, chat.openai.com, openai.com/chatgpt, openai.com/dall-e-2), Microsoft (copilot.microsoft.com, bing.com/chat), Google (bard.google.com, gemini.google.com), claude.ai, midjourney.com, perplexity.ai and character.ai.

Appendix A: AI keywords used in this study

A webpage was classified as containing an AI mention if it had at least one of the following keywords. These keywords were detected using regular expression matching, included common variants and were not case sensitive, unless noted otherwise.

- AI or A.I. (case sensitive)
- BERT or RoBERTa or roBERTa (case sensitive)
- GPT (including GPT 3, 4, 40, 401 or 403)
- LLM(s) or Large Language Model(s) or Language Model(s)
- NLP or Nature Language Processing
- X.AI or xAI (case sensitive)
- Grok (including Grok 1, 1.5, 2 or 3)
- AI Assistant
- AI Governance or AI Policy
- AI-Assisted or AI Assisted
- AI-Enhanced or AI Enhanced
- AI-Powered or AI Powered
- AI Regulation or AI Regulatory Compliance
- AI For Good or AI For Social Good
- AI Image Generator or AI Image(s)
- AI Overview
- AI Prompt or Prompt Engineering
- AI Summary
- AI Solution(s)
- Algorithm(s)
- Alibaba Cloud AI
- Anthropic
- Amazon AI or Amazon Alexa
- AWS Machine Learning
- Apple Machine Learning or Core ML
- Apple Intelligence
- Artificial Intelligence
- AI Translation or Automatic Language Translation
- Azure AI
- Baidu AI
- Bias in AI or Algorithmic Bias
- Bing AI
- Character.AI or Character AI or CharacterAI
- Chatbot(s) or Chat bot(s) or AI Chat

- ChatGPT or Chat GPT
- Claude AI (including Claude 2, 2.1, 3, 3.5, Sonnet, 3.5 Sonnet, Haiku, 3.5 Haiku, Opus or 3.7)
- Cognitive Computing
- Computer Vision
- Contextual Embedding
- Copilot Answer
- Deep Learning
- DeepFake(s) or Deep Fake(s)
- DeepMind
- DeepSeek (including DeepSeek LLM, MoE, Math, V2, V3 or R1)
- DALL·E or DALL-E or DALLE
- Ethical Tech or Responsible Tech or Ethical AI or Responsible AI or Ethics in AI
- Explainable AI
- Facebook AI or Meta AI
- Feature Engineering
- Fine-tuned Model or Fine-tuning a Model
- Generative Model or Generative AI
- Generative Adversarial Network(s)
- Gemini AI or Google Gemini
- Global AI
- Google AI or Google Brain Team or Google Brain or TensorFlow
- Healthcare AI or AI in Healthcare
- Huawei HiAI
- Hyperparameter Tuning
- IBM Watson (case sensitive)
- Intel AI
- Knowledge Graph
- LLaMA (case sensitive)
- Meta Llama or Code Llama (including Llama 2, 3, 3.1, 3.2 or 3.3)
- Machine Intelligence or Machine Learning
- Microsoft AI or Microsoft Copilot
- MidJourney
- Neural Network(s)
- NotebookLM or Notebook LM
- NVIDIA or NVIDIA AI
- OpenAI or Open AI (including OpenAI o1, o1-mini, o3 or o3-mini)
- Oracle AI
- Perplexity.AI or Perplexity AI or PerplexityAI
- Pre-trained Model(s)

- Predictive Analytics or Predictive Modeling or Predictive Modelling
- Qwen (including Qwen 7B, 72B, 1.8B, 2, 2-Math, 2.5, VL, VL2, Audio or LLM)
- QwQ or QwQ-2.5 or QwQ 2.5
- Reinforcement Learning
- Robotic Process Automation
- Retrieval Augmented Generation
- Salesforce Einstein
- SAP Leonardo or Leonardo AI (case sensitive)
- Self-Attention Mechanism
- Semantic Analysis
- Sequence Modeling
- Stable Diffusion
- Summary by Copilot
- Supervised Learning
- Tencent AI
- TensorRT or Tensor.Art
- Transfer Learning
- Unsupervised Learning
- Workforce Automation