FOR IMMEDIATE RELEASE

Everyone Has the Right to Receive and Impart Information, but Trusted Organisations in the UK and Europe Block AI From Giving People Full, Reliable Answers

New research across 394,031 websites in 239 countries and territories shows that newsrooms, governments and institutions across the United Kingdom, Ireland and Europe give different instructions to the crawlers of different artificial intelligence (AI) companies, and those crawlers collect the pages AI assistants can draw on when people ask a question.

LONDON, 26 November 2026. Article 19 of the Universal Declaration of Human Rights states that everyone has the right to seek, receive and impart information and ideas through any media and regardless of frontiers. When people in the United Kingdom ask an artificial intelligence (AI) assistant for the day's news, the answer they receive depends partly on which news websites that assistant's crawlers are allowed to open. The British Broadcasting Corporation (BBC) tells all three crawlers of OpenAI, a United States-based company, to stay out. The rules of 99 of the 434 United Kingdom news websites ranked among the million most popular websites in the world tell both of the crawlers that OpenAI uses to answer questions to stay out.

Invictus Communications this week published The Answer Gap, a global research report on digital inclusion and how people and organisations are represented in AI answers. Covering 394,031 websites in 239 countries and territories, it finds that websites grant different permissions to different AI companies, and that more than three in four United Kingdom websites with a clear result have no rule written for AI. Of the 1,506 United Kingdom websites that tell OpenAI's training crawler to stay out, 1,280 let in the training crawler of DeepSeek, a Chinese AI company.

Digital Inclusion in the Age of AI

Digital inclusion encompasses connectivity, access to technology, accessibility and digital skills. The Answer Gap examines further dimensions: access to information, digital identity, representation and informed choice.

Digital inclusion concerns both obtaining information and making it available to others. These dimensions apply to people with limited access to technology and to some of the world's largest institutions. An organisation's digital inclusion can also affect the people and communities it serves, because customers, employees, investors, students and citizens may use AI answers in their search for information.

"Open by purpose. Protected by choice. Digital inclusion means every person, company and institution deserves to be found, represented accurately in AI answers and in control of what it shares," said Lindsay McDonald of Invictus Communications.

What the Research Found in the United Kingdom

Every website can publish a public file, called robots.txt, that states which pages crawlers may access. Crawlers, or bots, collect pages for AI assistants such as ChatGPT, Gemini, Claude, Grok, Perplexity, DeepSeek, Copilot, Meta AI and Le Chat. The research checks each robots.txt file for 31 AI crawlers behind 16 AI assistants, models and data sets, made by companies and organisations in the United States, Canada, China and France.

Of the 10,546 United Kingdom websites with a clear result, 77.5 per cent have no robots.txt file, or one that names no AI crawler the Index recognises. The figure is 86.4 per cent (661 of 765) for government websites, 88.9 per cent (426 of 479) for research institutes and think tanks, and 76.0 per cent (6,610 of 8,692) for news and media websites.

Having no rule for AI does not mean being closed to AI. Of all United Kingdom websites with a clear result, 91.2 per cent allow all six crawlers that answer questions for OpenAI, Anthropic and Perplexity. For government websites the share is 98.6 per cent.

OpenAI uses two crawlers to answer questions: OAI-SearchBot, which finds pages for ChatGPT search, and ChatGPT-User, which fetches a page when a person asks. Of the 8,692 United Kingdom news and media websites with a clear result, 96.6 per cent allow Google's search crawler and 91.6 per cent allow OAI-SearchBot. Of the 8,500 of those news and media websites whose robots.txt file could be opened or was confirmed absent, 434 are ranked in the Tranco list of the million most popular websites, and 99 of them (22.8 per cent) tell both OAI-SearchBot and ChatGPT-User to stay out. Among the 8,066 others, 337 (4.2 per cent) do.

The robots.txt file of the British Broadcasting Corporation (BBC, bbc.co.uk) gives crawlers from the same company different instructions. It refuses ClaudeBot and Claude-SearchBot, which Anthropic, a United States-based company, uses for training and search, and lets in Claude-User, which fetches a page when a person asks Claude a question. It refuses all three crawlers of OpenAI and both of Perplexity, also United States-based companies, as well as CCBot of Common Crawl, a United States non-profit organisation, and Bytespider of ByteDance, a Chinese company. It uses Google-Extended to keep its pages out of the training of Gemini, made by United States-based Google. Its rules let in DeepSeekBot of DeepSeek, a Chinese company, GrokBot (the name crawler directories record for the crawler behind Grok, made in the United States by SpaceX Super Intelligence) and the crawler of Le Chat, made by Mistral AI in France.

The rules of The Telegraph (telegraph.co.uk) give the opposite instruction to the BBC's on three crawlers. The Telegraph refuses DeepSeekBot, which the BBC lets in. It refuses Claude-User, the crawler that fetches a page when a person asks Claude a question, which the BBC also lets in. It lets in Google-Extended, which the BBC uses to keep its pages out of Gemini training. The Telegraph's file names more than 30 crawlers. It refuses all three crawlers of OpenAI and of Anthropic, both of Perplexity's, Common Crawl's CCBot, ByteDance's Bytespider, the crawlers of Meta, Amazon's training crawler Amazonbot, Apple's training crawler Applebot-Extended and the crawler of Le Chat by Mistral AI. It also refuses the crawlers of Cohere, a Canadian company, and the Allen Institute for AI, a United States non-profit research institute, and a crawler named Grok. It lets in the search crawlers of Google, Microsoft's Bing, Apple and Amazon, all United States-based companies.

Two Chinese AI Companies, Two Different Instructions

Of the 10,546 United Kingdom websites with a clear result, 1,506 have robots.txt rules that refuse GPTBot, the crawler that United States-based OpenAI uses to collect text for training the models behind ChatGPT. The rules of 85.6 per cent of those 1,506 websites also refuse Bytespider, the training crawler of ByteDance, a Chinese company. The rules of 85.0 per cent (1,280) let in DeepSeekBot, the training crawler of DeepSeek, another Chinese company: none of those 1,280 files names DeepSeekBot, so their general rules allow it.

Ireland

Of the 969 Irish websites with a clear result, 80.0 per cent name no AI crawler the Index recognises, and the figure for government websites is 92.3 per cent (96 of 104). Of the 118 Irish websites whose rules refuse GPTBot, 99 (83.9 per cent) let DeepSeekBot in. The rules of Raidió Teilifís Éireann (RTÉ), the national public service broadcaster, give OpenAI's three crawlers different instructions: they refuse GPTBot and ChatGPT-User and let in OAI-SearchBot. They refuse Anthropic's training crawler and let in its search and user crawlers, and they let DeepSeekBot and GrokBot in.

Europe

Across 25 selected European countries, the United Kingdom, Ireland, Ukraine and Russia included, 9,811 websites refuse GPTBot, and the rules of 71.5 per cent (7,012) let DeepSeekBot in. In 13 countries of Western, Northern and Southern Europe the share is 69.0 per cent (3,815 of 5,530). It is 88.2 per cent in Portugal, 84.1 per cent in Italy, 77.3 per cent in France, 73.6 per cent in the Netherlands and 72.5 per cent in Spain. It is lower in Germany (55.0 per cent), Norway (57.1 per cent) and Sweden (50.0 per cent).

In eight countries of Central, Eastern and South-Eastern Europe, it is 64.6 per cent (1,052 of 1,628): 90.6 per cent in Romania, 81.7 per cent in Greece and 78.0 per cent in Poland. Czechia has the lowest share of any country with at least 100 such websites, at 30.4 per cent of 546: most of the Czech files that refuse DeepSeekBot do so through a rule that refuses every crawler not named, and 61 of the 546 are regional websites of one newspaper group, Deník.

Ukraine and Russia

Of the 2,390 Ukrainian websites with a clear result, 81.8 per cent name no AI crawler the Index recognises. Of the 299 Ukrainian websites that refuse GPTBot, 269 (90.0 per cent) let DeepSeekBot in, and 259 (86.6 per cent) refuse ByteDance's Bytespider by name while their general rules allow DeepSeekBot. Among Ukrainian news and media websites that refuse GPTBot, the share that lets DeepSeekBot in is 93.1 per cent (217 of 233).

Of the 6,839 Russian websites with a clear result, 86.3 per cent name no AI crawler the Index recognises. Of the 730 Russian websites whose rules refuse GPTBot, 62.3 per cent also refuse Bytespider, and 68.1 per cent let DeepSeekBot in. Of the 253 websites the Index classes as Russian government with a clear result, most of them pages of the Ministry of Foreign Affairs of the Russian Federation and its embassies, 147 refuse GPTBot, and 136 of those 147 belong to that ministry (mid.ru). The rules of 144 of the 147 let DeepSeekBot in, and the rules of 7 refuse Bytespider.

From Visibility to Representation

The findings contribute to search engine optimisation (SEO), answer engine optimisation (AEO), generative engine optimisation (GEO) and large language model optimisation (LLMO). The research gives these fields data and methods for checking crawler permissions, the sources assistants cite and the accuracy of an organisation's representation.

The new AI-Readable Sources Index shows which AI crawlers a website allows and compares the results by sector and country. The Answer Standard is a method for checking which pages assistants open, what they cite and how accurately they answer. Together, they help leadership, communications teams and industry bodies base decisions about access, representation and protection on evidence.

Access, Protection and Informed Choice

Digital inclusion does not require unrestricted access to information. It must also recognise the rights of information providers, including privacy, security, intellectual property, attribution and choice over how their information is used.

Under the report's guiding principle, each organisation decides what to open to AI, what to protect and whether its content may train AI. Those decisions should be informed by evidence about accessibility, representation and the organisation's responsibilities to the people it serves.

The Index is at https://invictuscomm.com/index/
The report summary is at https://invictuscomm.com/research/the-answer-gap/
Images free to publish with credit, including United Kingdom, Ireland, Europe, Ukraine and Russia infographics in English and in the national language of each country: https://invictuscomm.com/research/the-answer-gap/#media
Journalists, researchers and regulators can request the full report free of charge.

About Invictus Communications

Invictus Communications is a communications and research consultancy.

Media contact

Lindsay McDonald: lindsay@invictuscomm.com