the bright initiative | by Bright Data

Impact Report 2024

Table of Contents

Letter from Bright Data CEO

Or Lenchner

Dear Partners,

2024 was a transformative year for Bright Data and the digital landscape, marked by significant advancements in AI. AI is only as powerful as the data that fuels it. As AI becomes the number one consumer of data, one truth is undeniable: the future of AI depends on unrestricted access to high-quality, real-time web data. At Bright Data, we are leading the web data industry, ensuring that AI models have the scale, accuracy, and adaptability they need to succeed.

Public web data is the backbone of AI innovation. It drives transparency, powers automation, and enables organizations to build smarter, more effective AI systems. But, without open access to this data, AI models will be outdated, inaccurate, and uncompetitive. The companies that access and scale web data effectively will define the future of AI. Our commitment remains unchanged—to make public web data accessible to all, ensuring that innovation is not restricted by lack of access to the world's information.

In 2025, the year of inference time, web data will be the key to building AI that is faster, more efficient, and more impactful than ever before. Bright Data is not just a data provider; we are shaping the future of AI by ensuring that the data it relies on remains open, ethical, and scalable.

This year, we also aim to empower our nonprofit partners to explore how AI can be developed and used for positive change. Together, we can shape the next generation of AI to build a more equitable world.

Sincerely,

Or Lenchner
CEO, Bright Data

Letter from The Bright Initiative General Manager: Dana Mazia

As we conclude another remarkable year at The Bright Initiative, it is with deep pride that I present this year's impact report.

In 2024, we continued to expand our global network of partners, totaling over 87 new organizations. From detecting bias and hateful content online to developing AI for good, our partners are using Bright Data's resources to create a more informed, safer, and equitable world.

2024 has been particularly critical in the context of elections. With political tensions rising globally, misinformation, disinformation, and the proliferation of AI-generated content are posing unprecedented risks to democracy and public trust. As a result, we've seen a dramatic surge in demand for tools that can help identify, track, and mitigate the spread of false narratives. The Bright Initiative has played an integral role in empowering organizations with the data needed to safeguard the integrity of public discourse.

This year with the expansion of our Hackathons, we attracted more data experts to participate in data for good initiatives, experiencing their power to promote positive change through their data skills.

Another key focus this year has been policy. In a world where data and AI are increasingly influencing every aspect of our lives, the importance of strong, ethical frameworks cannot be overstated. Throughout 2024, The Bright Initiative has worked tirelessly to engage in meaningful policy discussions around data ethics, AI accountability, and the responsible use of public web data. By bringing together diverse stakeholders—including policymakers, tech experts, and civil society groups—we are working towards an ecosystem where public web data serves the interests of humanity.

As we move into the next year, we remain focused on our core mission: empowering organizations to use public web data for social good, advancing ethical standards in data collection, and defending the integrity of our digital world.

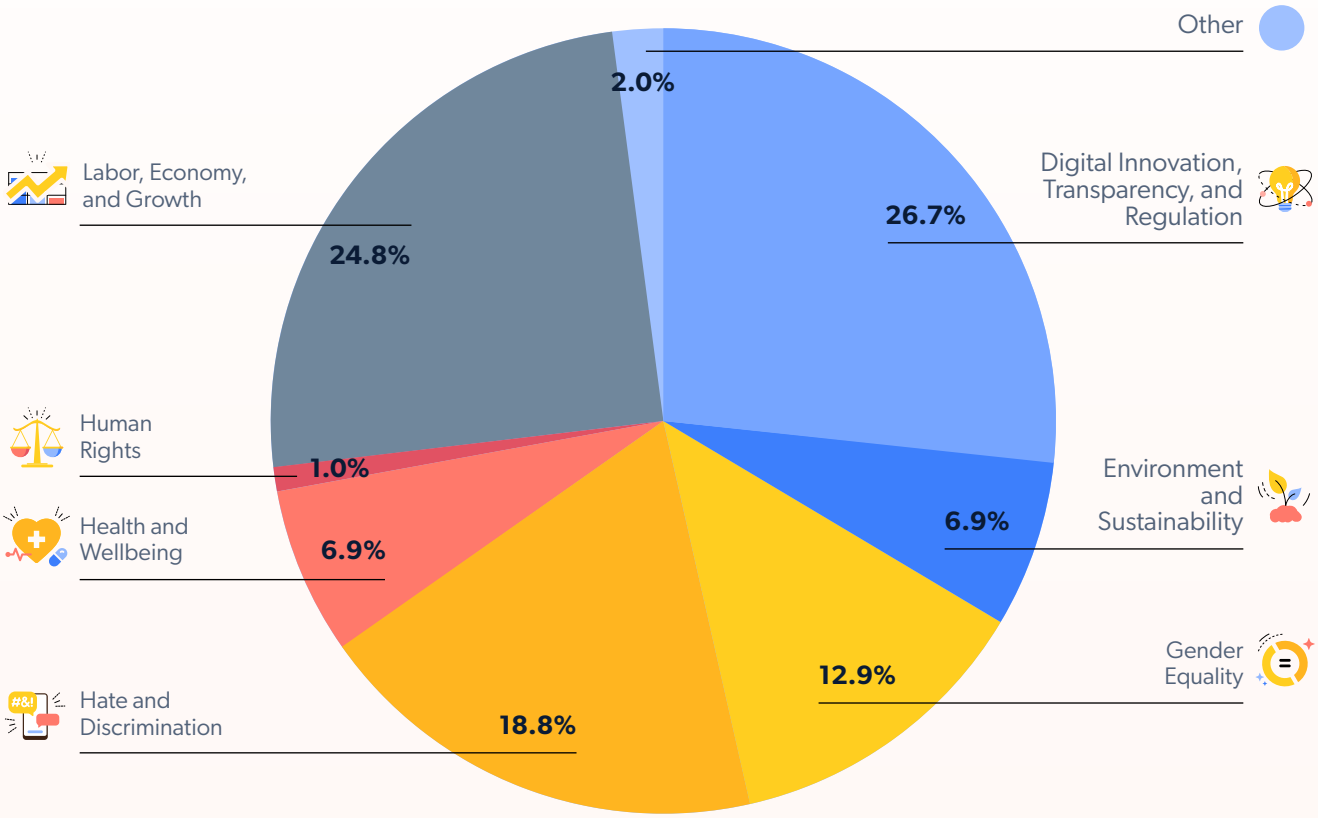
I invite you to explore the impact stories featured in this report, which reflect the powerful change we helped to achieve. We are more committed than ever to building a future where data, technology, and ethics intersect to serve humanity.

Thank you for your continued support. Together, we will continue to rise to meet the world's most urgent challenges.

Dana Mazia
General Manager, The Bright Initiative

2024 at a Glance

New Projects: Areas of Focus Breakdown



The Bright Initiative by Bright Data's ***impact on the academic community is profound.*** *By alleviating the burden of data collection, they empower researchers to dedicate more time to critical thinking and complex analysis,* ***pushing the boundaries of knowledge and innovation.***

Professor Rohit Aggarwal, University of Utah

CSR Excellence Award 2024
Investment of Corporate Time, Effort, and Funding

Snowflake's Data Drivers
Data For Good, EMEA

British Data Awards
Data for Good Consulting Initiative of the Year

2024 at a Glance

Partner Highlights

Exposed online hate in realtime on social media platforms

Uncovered bias in media outlets and search engines at scale along with the help of advanced AI models

Promoted transparency in election cycles around the world against false narratives, misinformation and disinformation

Accelerated climate action towards cleaner future regulations and market advancement

> *They have helped my research tremendously and I definitely would recommend Bright Data to anyone in academia…**Their support to academia is generous and amazing.***
>
> Professor Guangli Lu, The Chinese University of Hong Kong

> *The account managers are not just there to assist; they act as **real partners in your data extraction journey,** providing valuable insights and support to ensure your success.*
>
> Ariel Kirtchuk, CEO, Fakeoff

Founded the Alliance for Responsible Data Collection

ARDC Alliance For Responsible Data Collection

New Partners

Driving Policy and Regulation in Data and AI

In 2024, Bright Data solidified its leadership in shaping policy and regulation for data in general and data for AI specifically.

Founding the Alliance for Responsible Data Collection (ARDC)

A primary focus of our work this year was the creation of the Alliance for Responsible Data Collection (ARDC), with Bright Data as a founding member. ARDC unites leaders from businesses, nonprofits, and academia to establish responsible data collection standards. Its mission includes guiding best practices, offering third-party assurance for responsibly sourced data, and safeguarding open access to public web data.

With a steering committee featuring Common Crawl, Authors Alliance, and others, ARDC represents a shared commitment to ethical and transparent data practices.

Since its launch this year, the ARDC has submitted a draft of guidelines to the Internet Engineering Task Force, suggesting technical and governance standards for responsible data collection. It has also participated in robust conversations and workshops alongside the GPAI, OECD, and other leaders in the field.

For more information click here

Advancing Policy and Legislative Engagement

Bright Data actively engaged policymakers and contributed to key regulatory initiatives. We produced white papers and briefings demonstrating the societal benefits of public web data and participated in consultations with agencies like the UK Information Commissioner's Office and the French National Commission on Informatics and Liberty. Our insights influenced discussions on significant legislation, including the UK's Data Use and Access Bill.

Bright Data's policy efforts in 2024 reaffirm our commitment to shaping a fair, open, and ethical data ecosystem.

2024 Partner Spotlights

Focus Area:

Digital Innovation, Transparency & Regulation

During a year of global elections, the Initiative promoted Internet transparency while strengthening democratic values by combating the spread of misinformation, disinformation, and bias on social platforms and search engines, in the age of AI.

SCIENCE FEEDBACK

Science Feedback

Uncovering Disinformation Trends and Platform Biases

Science Feedback is a nonprofit organization dedicated to verifying the accuracy of scientific content, especially in areas like climate science and health. Their goal is to foster an Internet environment filled with reliable, scientifically vetted information by engaging experts to review and offer feedback on media pieces. Their reviews focus on objectivity, accuracy, and logic. This past year, Science Feedback used Bright Data for two notable research projects.

Use Case 1

Goal: Science Feedback looked to analyze the disparity in engagement between low-credibility and high-credibility accounts across major social media platforms, particularly with EU and election-related content.

Data Collected: It gathered engagement metrics from Instagram, Facebook, and YouTube, including likes, shares, and views for both low-credibility and high-credibility accounts.

Highlights:

- Posts from low-credibility accounts achieved significantly higher engagement per 1,000 followers: 2.3 times more on Instagram, 3.2 times more on Facebook, and 1.4 times more views on YouTube compared to high-credibility accounts.
- When analyzing EU- and election-related content, the disparity was wider: 1.7 times more engagement on Instagram, 7.6 times more on Facebook, and 2.4 times more views on YouTube for low-credibility accounts.

Impact: These findings suggest that platform algorithms prioritize engagement over credibility, promoting disinformation and raising concerns about compliance with the Digital Services Act and fairness for high-credibility content creators.

> ***It's simple, really: Bright Data's services just work.*** *Anyone working in this space knows how platforms' unanticipated design changes can make data collection impossible overnight, sometimes leading to the outright cancellation of research projects.* ***Using Bright Data is letting us build confidently:*** *we know that the data that is available today will still be available tomorrow. Not only does this mean that we save a lot of time when carrying out our work, but this confidence also means that we can expand the scope of the projects we can take on.* ***I'm very grateful that Bright Data is supporting our organization.***
>
> Bastien Carniel, Science Feedback

Use Case 2

Goal: Science Feedback set out to monitor whether X removes posts and accounts known to be involved in spreading disinformation.

Data Collected: It identified 1,366 posts on X that were part of a Russian disinformation campaign, Doppelganger.

Highlights:

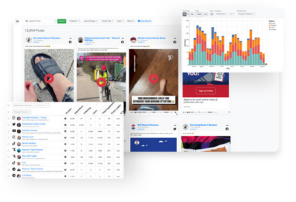
- 663 posts, along with their associated accounts, remained active on X since being reported.
- This contradicts X's community guidelines and its obligations under the Digital Services Act, which requires action against online harm.

Impact: These findings reinforce the urgent need for stronger enforcement of online platform regulations to prevent the spread of harmful disinformation.

National Conference on Citizenship

Scaling Impact Across Organizations and Researchers Using Shared Social Media Analysis

The National Conference on Citizenship (NCoC) is a nonpartisan nonprofit organization that aims to strengthen civic life in America. The Algorithmic Transparency Institute (ATI), a project of NCoC, focuses on enhancing the transparency of digital platforms that influence civic discourse. They achieve this by developing data collection and analysis tools such as Junkipedia. The tool is a monitoring and analysis dashboard for social media investigators. Junkipedia is powered by Bright Data's proxies and Web Scraper API, in its mission to track and analyze public content across 15 different social media platforms.



In 2024, ATI supported various researchers, journalists, and nonprofits in a range of research projects focused on understanding political speech, exploring how election officials communicate, analyzing greenwashing, tracking rumors, and educating the public through Junkipedia.

CANDIDATA24

CandiData24

Capturing Political Discourse in the Digital Age

CandiData is a collaborative project by the Center for Tech and Civic Life, the Media and Democracy Data Cooperative, the Social Media Archive at the University of Michigan, and the Political Science Department at the University of Kentucky. The project leverages Junkipedia, powered by Bright Data, and shares a structured repository of candidates' social media communications, offering accurate data across platforms.

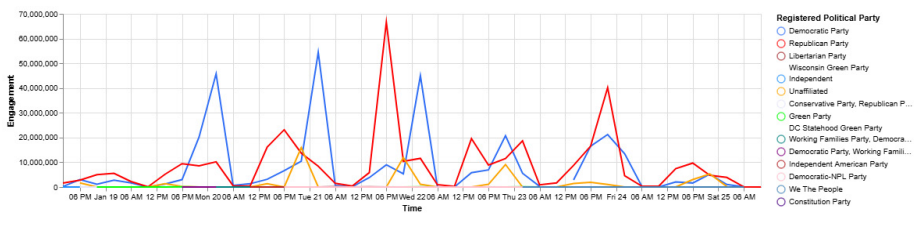
Goal: The organization aims to bridge a critical gap between unstructured online political discourse and researchers seeking reliable data for analysis.

Data Collected: During the 2024 US elections, it collected social media posts from accounts of U.S. federal and state candidates.

Highlights:

- Posts Created an open-access dashboard with aggregated content and data shared by U.S. federal and state candidates across various social media platforms.

Impact: CandiData has developed a detailed database of candidate social media accounts, that will be made publicly available on SOMAR and GitHub. These resources provide researchers, journalists, and civic organizations with valuable tools to analyze political speech and its influence on public discourse.



Total Engagement by Political Party, Taken from the CandiData Dashboard

The Washington Post

Washington Post

Ineffectiveness of a Fact-Checking Program on Social Media in Addressing Political Falsehoods

The Washington Post produced a series of investigative reports throughout the 2024 election cycle in the US, leveraging Junkipedia, powered by Bright Data and Bright Data's journalism program. One investigation focused on the effectiveness of X's (formerly Twitter) crowdsourced fact-checking program.

Goal: The Washington Post investigated the effectiveness of Community Notes, X's (formerly Twitter) crowdsourced fact-checking program, in combating election-related misinformation and addressing false narratives on the platform.

Data Collected: Researchers collected and analyzed 283 election-related posts flagged as false or misleading by independent fact-checkers, focusing on posts with proposed Community Notes, categorizing notes based on relevant terms like "Trump," "Biden," and "vote."

Highlights:

- Only 7.4% of proposed Notes related to elections were made public in 2024, dropping to 5.7% in October 2024, showcasing a decline in effectiveness.
- Even when a Community Note was eventually added, delays in the process—averaging over 11 hours for election-related posts—meant the false content had already reached millions of users.
- Posts by Republican politicians were four times more likely to have an approved Community Note compared to those by Democrats, despite more Notes being proposed for Democratic posts.

Impact: The analysis and findings suggest that Community Notes is failing as a standalone fact-checking program, especially in the context of politically charged misinformation. Its reliance on crowdsourcing and a consensus-based voting algorithm leads to inefficiencies due to partisan disagreement and delays in moderation.

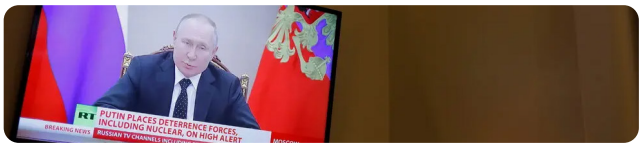
BROOKINGS

Brookings

How Russian State Media Uses TikTok to Influence Youth Audiences

The Brookings Institution is a nonprofit organization with the mission to conduct in-depth, nonpartisan research to improve policy and governance at local, national, and global levels. The Institute used Junkipedia powered by Bright Data to conduct thorough research during the 2024 US presidential election year and expose foreign influence.

Goal: The investigation focused on how Russian state-affiliated media use TikTok to spread politically charged narratives. By focusing on TikTok's unique algorithmic features, it documented how disinformation campaigns adapt to engage younger, highly active audiences.



Data Collected: The researchers collected posts created by Russian-affiliated accounts from Tiktok. They examined 70 affiliated accounts and compared them to previous years.

Highlights:

- About 5% percent of the TikTok content posted by the accounts was tied to U.S. political topics. Those posts had a much higher rate of engagement when compared to the same posts from the same account owners on other platforms.
- 22 of the top 30 most engaged TikTok posts tied to U.S. politics were in Spanish.

Impact: Using Junkipedia's data, Brookings identified key patterns in Russian state media's TikTok strategies. Its insights shed light on how these narratives penetrate U.S. political conversations, equipping policymakers, and researchers with actionable insights to combat the spread of foreign disinformation.

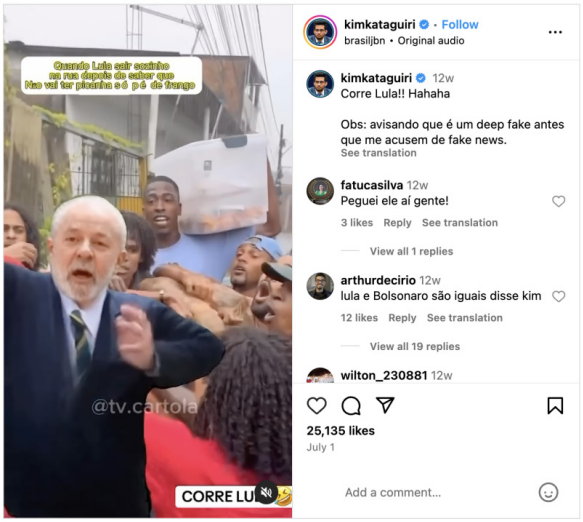
DFRLab

The Digital Forensic Research Lab (DFRLab)

Identifying Deepfakes Ahead of the 2024 Brazil Elections

The Digital Forensic Research Lab (DFRLab) is a nonprofit organization with technical and policy expertise on disinformation, connective technologies, democracy, and the future of digital rights. The DFRLab conducted research that monitors the use of AI-generated content in Brazil's 2024 municipal elections, collaborating with partners like NetLab at the Federal University of Rio de Janeiro. Using Junkipedia, powered by Bright Data, they were able to track and analyze manipulated media across social platforms to better understand its role in shaping public opinion during elections.

Screencap of a fake video featuring President Luiz Inácio Lula da Silva that was shared on July 1, 2024 by Kim Kataguiri, then a pre-candidate for mayor of São Paulo. (Source: DFRLab, Kim Kataguiri/Instagram)



Goal: DFRLab looked to study the landscape of research methods to help identify electoral deepfakes during the pre-campaign period.

Data Collected: For the study, they used Junkipedia to collect data from YouTube and Instagram (enabled by Bright Data), including videos and images during the election period, as well as engagement metrics to measure their effect.

Highlights:

- According to the study, social media platforms are hotspots for AI-driven disinformation.
- Even when a Community Note was eventually added, delays in the process—averaging over 11 hours for election-related posts—meant the false content had already reached millions of users.

Impact: With access to Junkipedia's advanced data collection and analytical tools, DFRLab effectively identified and documented AI-generated media's influence on Brazilian voters. The findings shed light on the growing role of social platforms in spreading disinformation, providing election bodies and international policymakers with the data to counter these threats. By offering actionable recommendations, DFRLab's study reinforces the importance of proactive measures to preserve democratic integrity in the face of AI's rapidly expanding role in digital media.

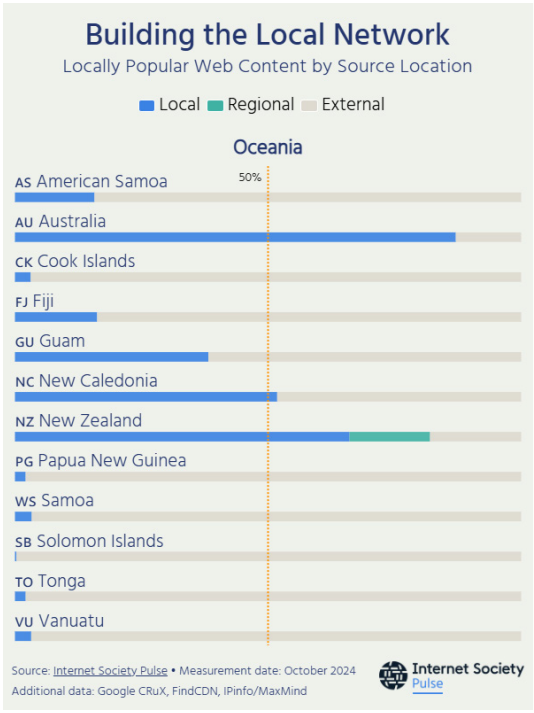
Internet Society

Internet Society (ISOC)

Transforming the Internet's Future Towards Equitable Ecosystems

The Internet Society (ISOC) stands at the forefront of advancing global Internet infrastructure. Committed to creating an Internet ecosystem that is open, secure, and beneficial, the Internet Society is a global charitable organization that drives the development of technologies, standards, and policy. Through its innovative project Pulse, ISOC provides global stakeholders with tools to monitor and shape the health, availability, and resiliency of Internet systems in real-time. One of its most ambitious initiatives, the 50/50 Vision, seeks to keep half of Internet traffic local in selected economies by localizing internet infrastructure resulting in faster, stronger, and cheaper Internet access. In addition to building Internet exchange points (IXPs) to improve traffic flow and Internet services, the team at ISOC collects web content from top websites, analyzes HTTP responses to derive location information (Geohints), and monitors responses in real-time.

To analyze the geographic distribution of Internet content effectively, the Internet Society faced limitations in bandwidth, complex regulatory environments, and difficulty scaling measurements across multiple locations. This challenge particularly impacted their ability to perform country-specific analysis in regions with less-developed Internet infrastructure. By leveraging Bright Data's extensive proxy networks, the Internet Society team at Pulse gained the means to capture reliable Geohints across various countries.



Locally Popular Web Content by Source Location, Pacific's Connectivity Hub. Region's Content Capital?, October 2024

The collaboration with The Bright Initiative has been pivotal in advancing the Internet Society's 50/50 Vision. Using Bright Data's solutions, the Internet Society gains invaluable insights into the localization of web traffic and can deliver real-time insights from countries around the world. Insights derived from its investigations are informing policies aimed at reducing latency, improving Internet costs, and enhancing access speeds in target economies. The collaborative success is showcased in ISOC's published reports and demonstrates the practical implications of studying traffic flows for both commercial and civic use cases. Through this collaboration, the Internet Society continues to push boundaries in creating equitable and efficient Internet ecosystems.

2024 Partner Spotlights

Focus Area:

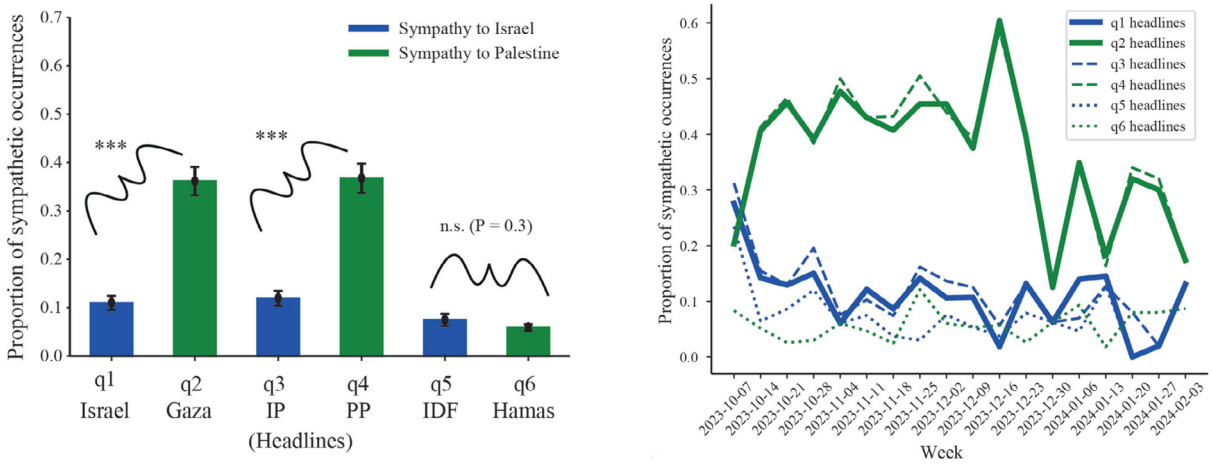
Hate and Discrimination

Fighting against hate, racism, xenophobia, discrimination, and corruption in both online and offline environments.

Research for Impartial Media (RIME) and Asserson Law Firm

Measuring Media Bias with LLMs

Research led by Trevor Asserson from Asserson Law Firm and Dr. Haran Shani-Narkiss from Research for Impartial Media (RIME) examined the BBC's portrayal of the Israel-Gaza conflict that started on October 7th, 2023. Its research employed innovative methods using ChatGPT-4 to evaluate and analyze the BBC's coverage of the war for bias, focusing on whether sympathy is expressed equally for both sides. The study examined the articles' main text and headlines, comparing BBC English and Arabic reporting, and extended the analysis to podcasts, radio, and TV.



Proportion of sympathetic occurrences by headlines of BBC articles | Proportion of sympathetic occurrences by headlines of BBC articles in weeks, 'Using Large Language Models to Measure Impartiality in the Media: ChatGpt-4 and the BBC's Coverage of the Israel-Gaza War as a Case Study'

To validate the comprehensive datasets manually collected by the Asserson team, the researchers aimed to replicate its findings using an automated data collection solution. By rigorously capturing all relevant publicly available data this approach mitigates the risk of explicit and implicit cherry-picking. To prevent any manual methods, they have used Bright Data's Web Scraper API and were able to collect all relevant articles spanning from October 7th. Researchers collected and validated data which was then fed into ChatGPT-4's large language model to identify bias.

The research achieved critical visibility, influencing global media discussions about bias and transparency. Since its publication, this research has been mentioned more than 0.5 billion times in global social media and was widely acknowledged by outlets such as Sky News, the Telegraph, Fox News, the New York Post, the Daily Mail, and more. In addition, the research has secured a method using LLM to uncover bias that could potentially be used in future studies and share a transparent perspective on various media sources around the world.

Anti-Corruption Data Collective

Exposing Transnational Corruption with Public Web Data

Anti-Corruption Data Collective (ACDC), a nonprofit coalition of investigative journalists, data analysts, academics, and policy advocates, committed to exposing and combating transnational corruption. In their recent report, ACDC was looking to reveal the influence of major actors on climate action.

Goal: The organization focused on corruption and undue political influence during the UN Climate Change Conference held in Baku, Azerbaijan, in November 2024.

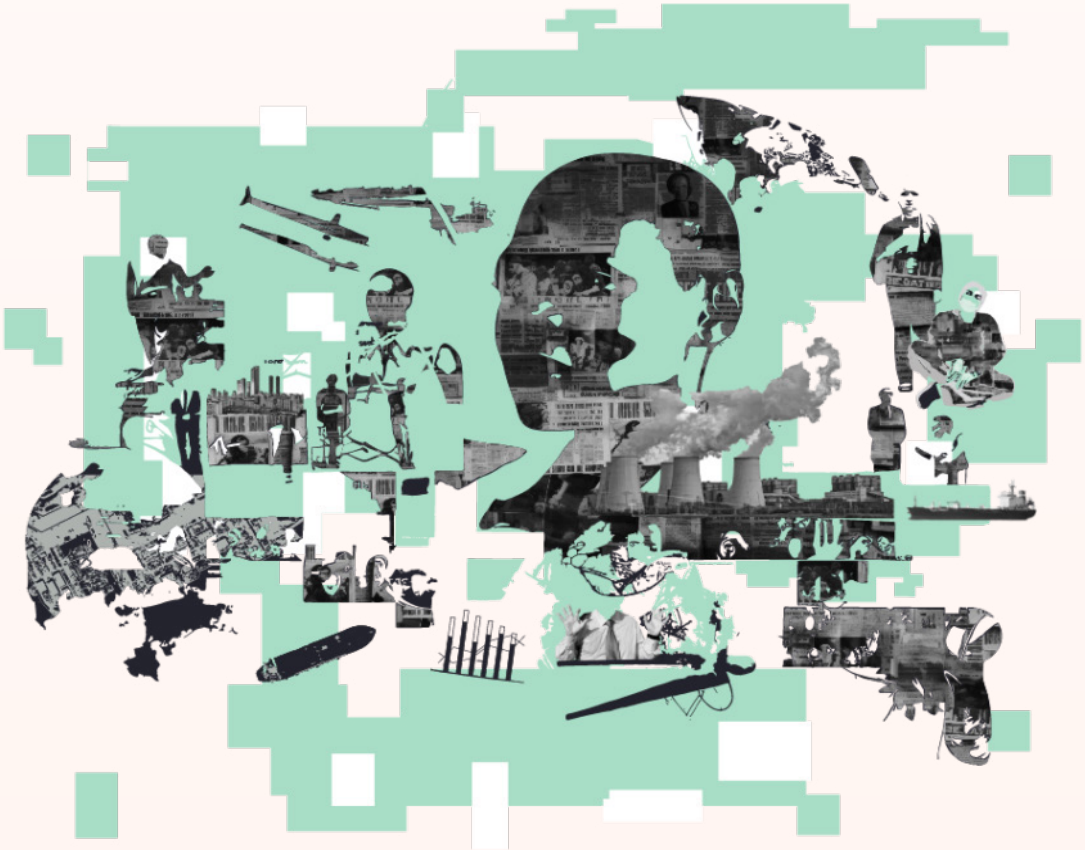
Data Collected: ACDC built datasets to uncover various large fossil fuels companies and workers' information, as well as public sector relevant actors to illustrate several risks to the integrity of UN climate conferences.

Highlights:

- It documented how lobbying and institutional loopholes allowed certain industries to dilute critical climate agreements.
- It revealed the influence of political and corporate actors in controlling funding allocation and project priorities at the UN Climate Conference.

Impact: The findings have been referenced by policymakers, nonprofit organizations, and academics advocating for greater scrutiny of international conference proceedings. ACDC's impactful work in this sphere continues to drive calls for more accountability within global climate governance.

> *I've been floored by the effectiveness of the Bright Data platform. I have used a number of their products for several years now. We have had* ***astonishing success using their proxies to collect web data for our research*** *and continuously rely on their ISP services to update our databases.*
>
> David Szakonyi, Co-Founder, Anti-Corruption Data Collective



Togethering
GATHER TOGETHER

Togethering Dashboard

Mapping Insights of Antisemitism Across the US

The Togethering platform provides a real-time intelligence dashboard to nonprofits working together to make the world better.

Goal: It aims to collect, analyze, and share data about real-world antisemitic incidents affecting people across the US.

Data Collected: The team at Togethering scrapes websites, videos, images, sound, and social media content and places it on an interactive map that can be sorted and filtered by several dimensions.

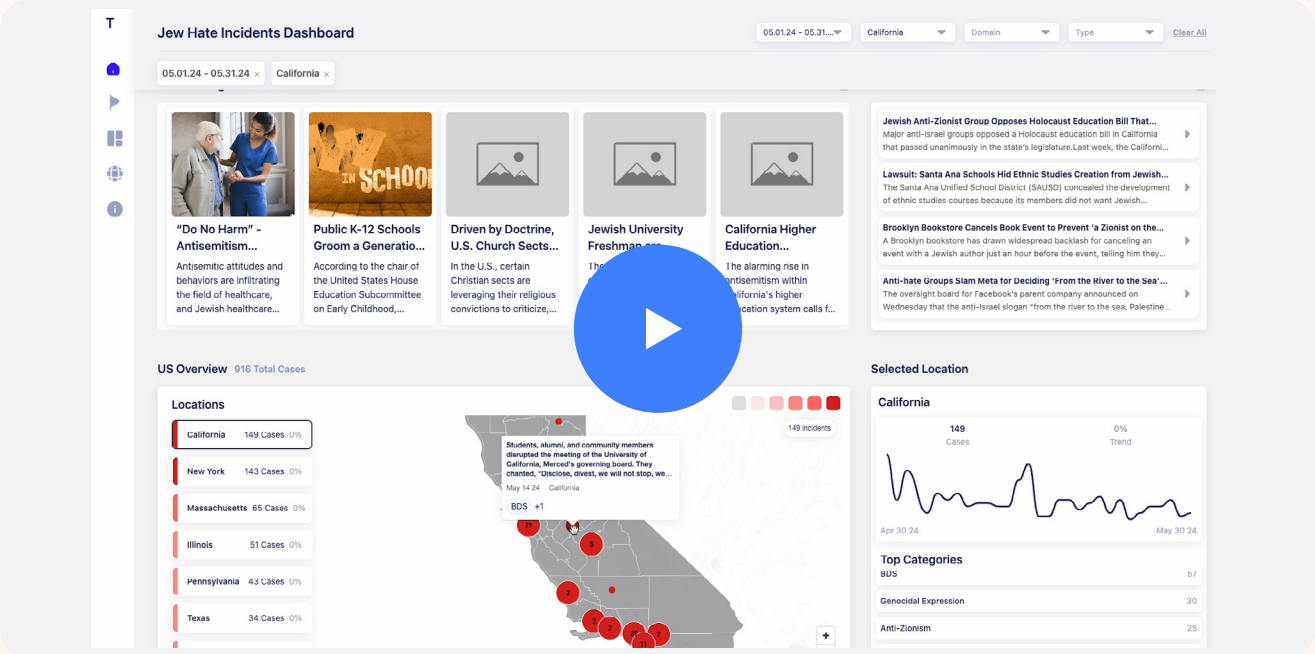
Highlights:

- The interactive map and the insights shared on the platform are powered by an AI analysis pipeline that processes collected data to identify, split, categorize, and locate incidents on a daily basis.
- This grants a strategic overview and insights for all partner organizations to combat hate and antisemitism.

*We bring unity and cooperation to the fragmented world of nonprofit initiatives. Through Bright Data and The Bright Initiative, **we gain access to global media and social media content**, which we analyze using an AI-driven pipeline. Bright Data allows us to navigate challenging websites, collect and archive photos and videos from social media, and efficiently track critical channels. This data enables nonprofits to coordinate programs and funding and respond quickly to emerging needs and trends*

CTO of Togethering

Impact: While various organizations focus their efforts to combat antisemitism in specific spheres, the Togethering platform provides AI-driven and concrete analysis in real-time, based on a high-quality pipeline of data. This provides organizations with significant intelligence and allows them to expand its impact.



IronTruth

Fighting Against Disinformation, Illegal and Harmful Content

IronTruth is an initiative that combats disinformation, harmful content, antisemitic, and anti-Israel content online. It created a Telegram bot that receives direct post URLs from users across various social networks and reports them to social media platforms.

Goal: IronTruth reports misinformation, AI-fabricated, hateful, and antisemitic content across all social platforms for removal, while encouraging regulators to take policy actions.

Data Collected: It collects reported URLs of posts that are suspected as fake, hateful, or antisemitic by volunteers on the IronTruth Telegram chatbot.

Source: IronTruth Report

Highlights:

- They presented a full report to the European Union regarding violations of the Digital Services Act across social platforms while sharing examples of the hateful content collected.

Impact: By automating its verification and reporting platform, IronTruth was able to remove hateful content across platforms and share its thorough investigation with EU policymakers to encourage action.

Platform	Reports by Users	Authenticated Reports	Removed	%
Facebook	4615	509	364	71.51%
Instagram	19508	1547	312	20.17%
Tiktok	3535	337	223	66.17%
YouTube	1248	142	22	15.49%
LinkedIn	393	39	25	64.10%
X	6328	739	91	12.31%
Reddit	348	4	0	0.00%
Total	**35975**	**3317**	**1037**	**31.26%**

2024 Partner Spotlights

Focus Area:

Human Rights

Advocating for the inherent rights and freedoms of all individuals by identifying harmful materials and acts against children, promoting free will by combating slavery, forced labor, and sex trafficking, and supporting freedom and liberties.

C4ADS
innovation for peace

C4ADS Center for Advanced Defense Studies

Exposing Mining Contracts in Central Africa

C4ADS is a nonpartisan nonprofit organization with a mission to defeat the illicit networks that threaten global peace and security. Powered by their data and technological capabilities, they produce compelling analyses to target illicit networks that operate across borders to destabilize communities, prevent development, and devastate the environment. One of their released reports in 2024 exposed the extent of the foreign-owned mining concessions near protected areas in the Democratic Republic of Congo, and the effects of these concessions on the local communities and habitats.

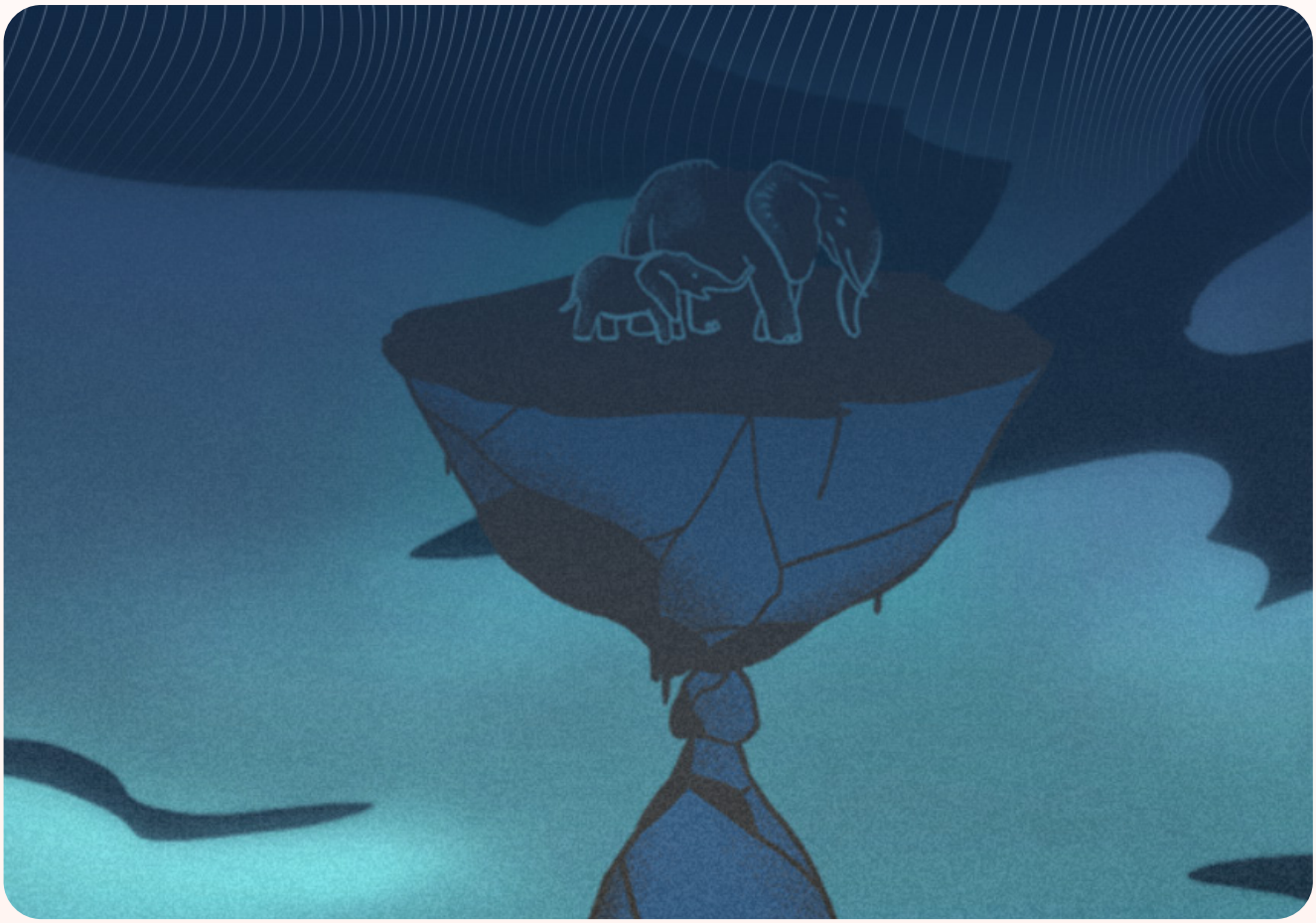
Goal: The organization investigated the effects of unregulated mining in the Democratic Republic of the Congo (DRC) while highlighting the rising pressure on the country's ecosystems.

Data Collected: C4ADS collected publicly available corporate data and mining concessions.

Highlights:

- 45% of all foreign-owned mining concessions located within 10 kilometers of protected areas in the DRC are linked to Chinese enterprises.
- There has been a 69% annual increase (on average) in mining concessions granted to Chinese-owned enterprises near protected areas since 2015.

Impact: The report reveals the immense damage caused by concessions near conservation zones, offering actionable policy recommendations. These include the creation of stronger regulatory mechanisms in the DRC and greater accountability for international enterprises. By increasing awareness of the urgent need for responsible mining practices, the project promotes long-term sustainability for one of the world's most biodiverse regions.



2024 Partner Spotlights

Focus Area:

Environment and Sustainability

Protecting the environment through the promotion of clean energy, preservation of oceans and animals, and advancement of climate action.

Mycelium

Mycelium

Centralized Open Access Network for Carbon Transparency

Mycelium is leading a transformation in carbon accounting with its open-access emissions database. By creating a one-stop network powered by artificial intelligence, Mycelium consolidates emissions data from thousands of businesses into an accessible format.

Goal: It aims to empower companies, climate-focused apps, and decision-makers with real-time access to sustainability data that ensures transparency and accountability.

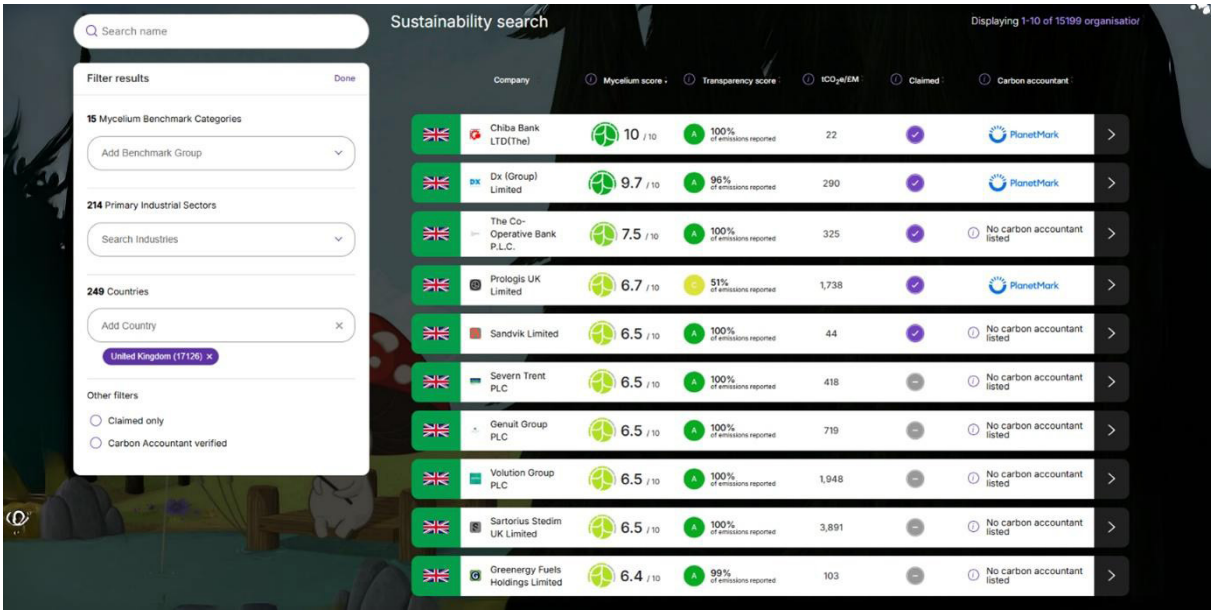
Data Collected: Mycelium collects publicly available sustainability and carbon reports, pulling verified emissions data from UK businesses.

As we're working in a space trying to democratize access to carbon emissions data, we were welcomed to The Bright Initiative. They gave us access to their data sets and tools which we have been able to ***harness and push our mission forward.***

It's incredible the amount of data they can handle and surface so quickly, and ***the breadth of their data collection is almost unfathomable.***

Tom Carpenter, Founder of Mycelium

Impact: By integrating Bright Data's advanced Scraping Browser tools and Residential Proxies, Mycelium's platform pulls verified data from public sources like Glassdoor, LinkedIn, and Yahoo Finance, creating an ecosystem of transparent carbon accounting. The Open-Access platform currently hosts 13,000 verified sustainability reports from UK companies and provides tools for live access through its API.



Mycelium's open-access database

Ambient

Accelerating the Electrification of Heating and Cooling in the UK

Ambient is an independent nonprofit organization advocating for data-driven policies that promote the adoption of electrified heating and cooling solutions in the UK and beyond. Ambient envisions a future where sustainable electrified heat is available to all, supported by a thriving local electrification industry. Ambient is working to dismantle barriers to mass adoption and advance bold electrification goals.

Goal: Identifying high-potential households for heat pump adoption across the UK.

Data Collected: Ambient collected 60,000 records of registered boilers across the UK from the official Gas Safe Register page. It extracted the boiler age and location to better understand potential demand to enable industry, central, and government partners to strategically target interventions, particularly for homes with aging boilers nearing the end of their lifespan.

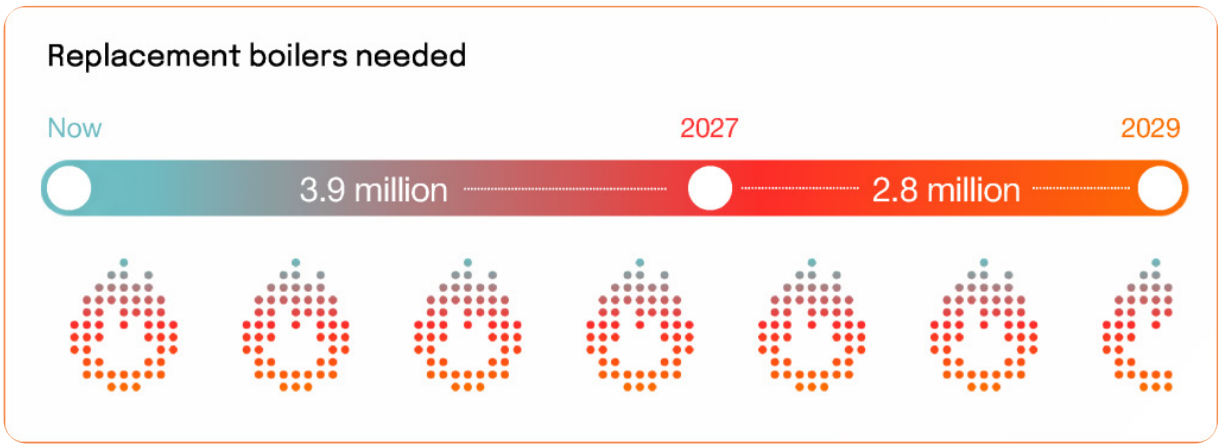
Highlights:

- 3.9M Gas boilers need replacement by 2027.
- Nearly 30% of boilers will need replacement by 2029

Impact: By utilizing Bright Data's Web Scraper IDE, Ambient's report insights highlighted the value of publicly available data in identifying high-potential households for heat pump adoption, guiding targeted campaigns, and making policy recommendations.

120,665

Page Loads



Source: 'How unlocking the Gas Safe Register data could accelerate the UK's clean heat economy', Ambient

New AutoMotive

Monthly Tracking Platform for Accelerating the Global Electric Vehicle Market

New AutoMotive is an independent transport research organization with a mission to support the switch to electric vehicles. This year, it released the Global EV Tracker, covering key car markets that account for around 85% of global car sales.

Goal: The organization gives unique insights into the transition away from fossil-fuelled transport, helping businesses and governments plan with an interactive market tracker and monthly reports.

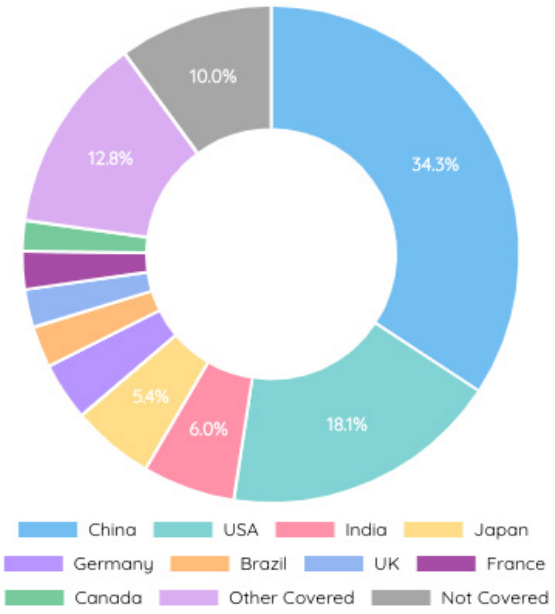
Data Collected: It collects electric cars market insights and sales reports from both governmental and other reliable sources around the world including Italy, the UK, Norway, and more. Once collected, the data is streamed to their Global EV Tracker and used for interactive dashboard and monthly reports covering the global market.

2,647,041
Page Loads

Highlights:

- Through the research, the team gathered insights from 40+ markets accounting for 85% of car sales.
- In the 12 months to October 2024, 10.8m battery-electric cars were sold, an increase of 1.1m, or more than 11%, compared to the 12 months to October 2023.

Impact: By utilizing Bright Data's Web Scraper IDE, NewAutomotive's Global Electric Vehicle Tracker can now share data-based insights on a monthly basis to help governments, manufacturers, and entrepreneurs around the world navigate and accelerate the transition to electric vehicles.



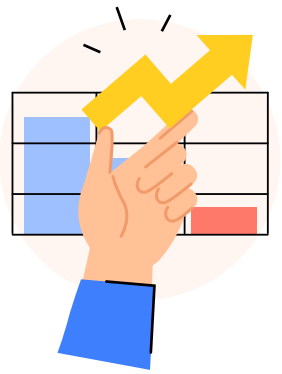
Capturing over 85% of the global car market, NewAutomotive's Global EV Tracker

2024 Partner Spotlights

Focus Area:

Labor, Economy, and Growth

Addressing poverty, industrial growth, and educational advancement within communities.



University of Warwick, Dr. Nikhil Datta

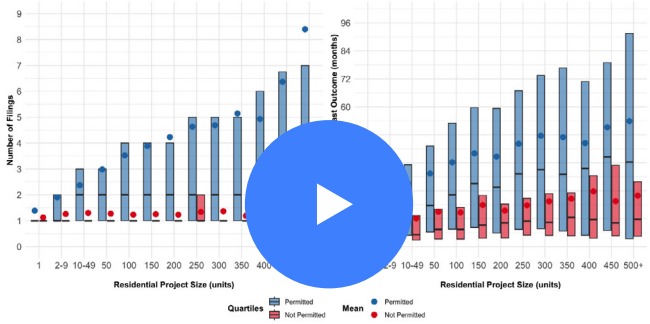
Monthly Tracking Platform for Accelerating the Global Electric Vehicle Market

Dr. Nikhil Datta, Assistant Professor at the University of Warwick, has undertaken a crucial research project to address inefficiencies in the UK's planning system, which hinders the housing supply in the face of rising demand. The research, 'Five Facts on the UK Planning System', examines 18 million planning applications spanning from 2000 to 2023, revealing significant roadblocks in approval processes for large-scale housing developments. This research builds a comprehensive dataset aimed at informing housing policies and facilitating targeted interventions to increase housing availability in the UK.

Severe housing shortages in the UK are largely driven by bottlenecks within the planning system. Local authorities regularly miss government housing targets as large-scale developments, essential for addressing rising demand, face bureaucratic delays, infrastructure dependencies, and community opposition. The research required accessing vast amounts of public data hosted on websites across local councils in England. However, maintaining access across multiple platforms posed a challenge, as some sites blocked requests or restricted data extraction after only a few queries. To overcome this obstacle, Nikhil partnered with The Bright Initiative to utilize Bright Data's proxies, which ensured uninterrupted access to valuable publicly available planning data.

Dr. Datta's research has already provided critical insights into the underlying issues of planning restrictions in the UK. Key findings indicate that while larger projects have lower refusal rates they frequently face longer bureaucratic delays, disrupting the housing supply. Utilizing Bright Data tools, the project successfully standardized and cleaned datasets across local councils, detailing approval and rejection statistics, timelines, and decision-making trends. The study aims to empower local authorities with actionable statistics via a publicly available database and a policy brief to guide interventions. This work represents a significant step toward enabling accurate, data-driven decision-making to address the housing affordability crisis in the UK.

Policy-relevant research



Barriers to Housing Supply by Project Size

- Tackling the lack of new and affordable housing is a policy priority in the UK
- Application-level planning data help us go beyond commonly-used aggregate statistics to study barriers to new housing

Rise IL

Driving Innovation and Economic Resilience

Rise IL (formerly SNPI) is an independent nonprofit think tank dedicated to fostering a competitive, sustainable, and innovation-driven economy in Israel. With a vision of positioning Israel as a model nation where innovation powers economic prosperity and societal impact, Rise IL develops and advocates for effective policies to address the country's complex local and global challenges. Its work is centered on three key areas: strengthening Israel's high-tech ecosystem to enhance economic resilience, expanding access to innovation across diverse populations and regions, and leveraging AI technologies to improve the quality of life for all Israelis.

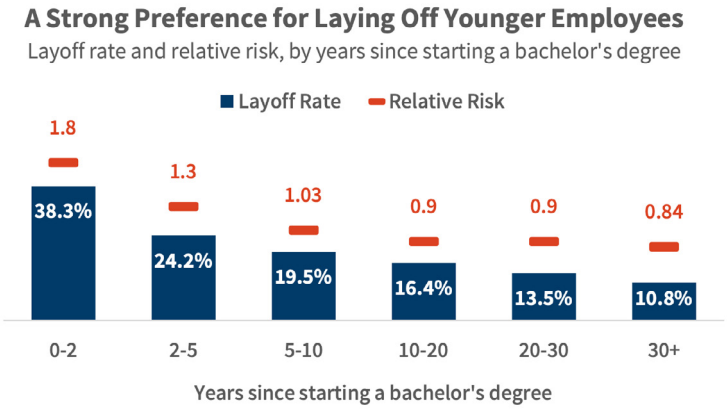
To support its research and advocacy, Rise IL relies on public web data for in-depth analysis and reporting. Through a collaboration with The Bright Initiative, it collects LinkedIn and Crunchbase datasets, which provide crucial insights into the Israeli workforce and ecosystem. This partnership has enabled Rise IL to enhance the impact of its policy recommendations through the production and wide distribution of four comprehensive and data-driven reports.

Among Rise IL's most notable studies is 'In-Depth Analysis of Layoffs in the Israeli High-Tech Sector', which examines critical issues such as who faces the greatest risk of job loss, how long it takes for individuals to find new employment, and where displaced workers transition next. Its research also explores Israel's artificial intelligence landscape, the decline in new Israeli startups, and the overall state of the high-tech industry. As a pivotal force in Israel's innovation ecosystem, Rise IL's work is widely recognized for its significant contributions to policy development and its role in shaping the future of the Israeli economy.

3,675,764

Records

Graph Source: 'An In-Depth Analysis of Layoffs in Israeli High-Tech', by RISE IL



Impact Through Education and Innovation: The Bright Academy

The Bright Academy, the educational arm of The Bright Initiative, transforms the knowledge and expertise of Bright Data into engaging, hands-on educational workshops. Designed to empower participants with practical skills, these sessions emphasize the potential of public web data in solving real-world challenges in business and for social good. Through classroom settings, both physical and digital, across universities worldwide, it has helped bridge the gap between theoretical knowledge and practical application, offering tailored sessions to equip participants with actionable insights.

A Focus on Real-World Impact with Hands-On Experience: Hackathons

This year, we took practical application a step further by putting data directly into participants' hands, showcasing its power to drive meaningful solutions. Beyond the classroom, our hackathon initiatives spanned student and industry-led events, ensuring a broad reach and significant impact. These hackathons fostered innovation and collaboration, enabling participants to tackle pressing global challenges through data-driven solutions.

> *I had the pleasure of collaborating with The Bright Initiative for the Future of Data Hackathon, where they made remarkable contributions to our participants. The Bright Initiative provided a comprehensive dataset of social media posts on AI, gender, race, equality, and environmental topics—**empowering participants to create meaningful projects centered on social impact.***
>
> Srihari G. Das, MidWestCon, The Future of Data Hackathon

Highlights:

- **AI for Democracy:** A six-month challenge dedicated to developing innovative solutions to combat misinformation and disinformation.
- **Snowflake Hackathon:** We supported two Snowflake hackathons: one at the annual summit, addressing homelessness in the U.S., and the second with the India community tackling education and employment challenges.
- **TED AI in San Francisco:** A competition focused on developing AI-powered solutions to address the United Nations' Sustainable Development Goals, targeting some of the planet's most pressing challenges.
- **Forum Michal Sela Hackathon:** An event dedicated to developing life-saving technologies and solutions to prevent violence against women.

Credit: Mor Arkadir

AI Ramp

Data is a foundational element of every AI breakthrough but is a difficult resource for early-stage AI startups to access. Without data, innovation stalls before it even begins. Recognizing this challenge, The Bright Initiative created the AI Ramp Program, a bridge between ideas at the beginning of their development and the data needed to bring them to life.

By equipping these startups with the resources they need at a pivotal stage, the AI Ramp Program is addressing a critical data access gap. Several emerging early-stage Israeli AI startups have joined the program, leveraging Bright Data's scraping technologies and datasets to advance their work.

Spotlight on Some AI Ramp Startups:

Brinker AI:

Brinker is an end-to-end misinformation threat management platform that serves the public sector and major enterprises. It combats malicious and social threats using proprietary narrative analysis, AI-enabled detection, and automated OSINT investigations. The SaaS platform offers a suite of mitigation tools at the press of a button, including pre-legal action, media publication, content removal, and counter-narratives.

Visit Website

Elements:

Elements is a platform that helps companies measure and reduce carbon emissions from employee activities like business travel, commuting, and working from home. Using AI provides clear and accurate data, with practical suggestions to cut emissions and save costs. This data builds trust with clients, encourages them to take action, and makes Elements a valuable tool for businesses aiming to meet sustainability goals while staying efficient and cost-effective.

Visit Website

Palta:

Palta is the world's first AI-driven marketplace designed to revolutionize inclusive fashion. Palta is tailored for people with disabilities, offering an unparalleled shopping experience by prioritizing accessibility, comfort, and inclusivity. The platform is not just about clothing, it's about empowerment. Through advanced data insights and curated product collections, Palta bridges the gap between brands and the disabled community, ensuring fashion is accessible to everyone.

Visit Website

the bright initiative | by Bright Data

See you in 2025!