the bright initiative | by Bright Data

IMPACT REPORT 2025

Table of Contents

Letter from Bright Data CEO
Or Lenchner

Dear Partners and Friends,

2025 was a breakthrough year for AI, when bold ideas became reality. At Bright Data, we've spent years preparing for this moment. Today, our data solutions power some of the world's most innovative AI systems, models that now run at global scale and drive real business impact. From powering the data infrastructure of most leading AI labs to enabling millions of intelligent agents that interact with the live web daily, Bright Data is the data infrastructure layer behind the AI revolution.

Our mission is about ensuring that web data is accessible, transparent, and ethically sourced. As the line between physical and digital life continues to blur, we remain the connecting layer between today's static models and the real-time, ever-changing web. This is how we keep AI learning from the world as it is, not as it was.

Through The Bright Initiative, we continue to stand by our commitment to use data for good. From supporting NGOs tackling climate challenges, to helping academic researchers access high-quality data, to empowering startups developing responsible AI, we've opened our technology to changemakers worldwide. Their work reminds us that access to data drives progress far beyond commercial boundaries.

Our commitment to setting global standards for public web data and access continues through our work at the Alliance for Responsible Data Collection, to ensure innovation is built on fairness, transparency, and trust. It's not enough to lead technologically, we must lead responsibly.

AI will define the next decade and Bright Data will keep ensuring that it's built on a foundation of integrity, inclusivity, and truth.

Thank you for being part of this journey. The next era of data and AI is already here and we're proud to be shaping it together.

With appreciation,

Or Lenchner
CEO, Bright Data

Letter from The Bright Initiative General Manager
Dana Mazia

Dear Community,

As we enter 2026, I'm proud to present our latest Impact Report, highlighting our ongoing commitment to using public web data to create positive global change. We saw extraordinary growth in our global network, growing nearly 200%, with 76% coming from the non-profit sector and 23% from prestigious academic institutions.

This year, the Bright Academy expanded its reach to the next generation of data experts. Through our support of major events like CalHacks, the world's largest hackathon, we are bridging the gap between academic theory and real-world innovation for the next generation.

We continued to lead the conversation surrounding ethical public data collection and the establishment of cross-sector standards in an ever-changing and increasingly critical field in the age of AI. The Alliance for Responsible Data Collection (ARDC), co-founded by Bright Data, succeeded in expanding its coalition by onboarding 12 new leading partners, including major technology and AI companies such as Microsoft, Google, and OpenAI, demonstrating a commitment to open internet access in the tech industry.

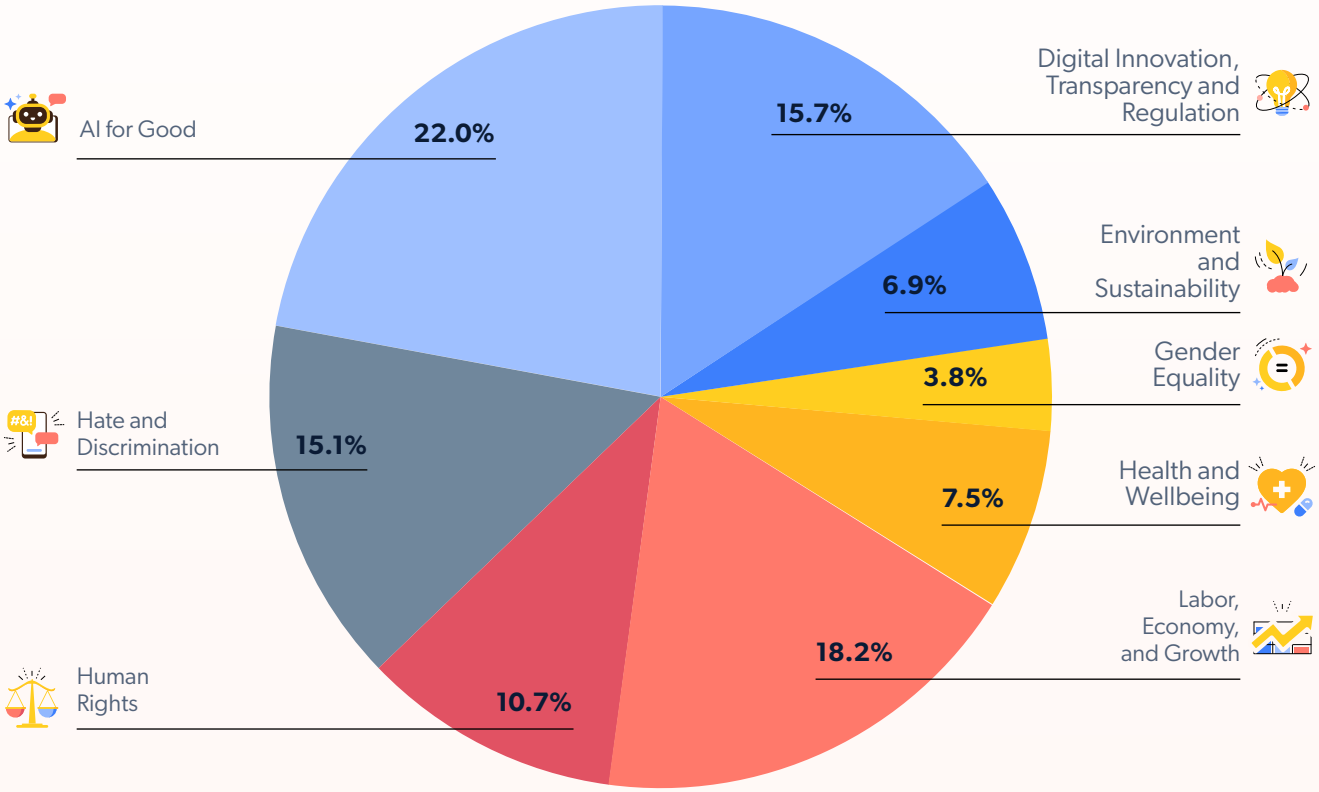
I invite you to explore the stories within this report, which showcases the profound impact our community has achieved together. Thank you for your continued partnership as we work to build a safer digital world.

Sincerely,

Dana Mazia
General Manager, The Bright Initiative

2025 at a Glance

New Projects: Areas of Focus Breakdown



600 Million+ data endpoints collected

> *The Bright Initiative is a* ***perfect fit for our research needs...*** *[providing] us with the tools to* ***bridge the gap between private companies and research institutes.*** *This partnership is exactly what we were looking for.*
>
> Amit Zac, Postdoctoral Researcher at ETH Zurich

2025 at a Glance

Policy

Bright Data led global ethical public data collection through its work with the Alliance for Responsible Data Collection (ARDC), collaborating with major leading partners including **Common Crawl, Electronic Frontier Foundation (EFF), Microsoft, Google, OpenAI** and more.

ARDC Alliance For Responsible Data Collection

> *Organisations such as the Alliance for Responsible Data Collection (ARDC) have* ***developed 'Technical Standards and Governance Guidance for Responsible Data Collection' based on input from a range of for-profit and non-profit entities*** *that crawl, scrape or rely upon scraped web data.*
>
> Financier Worldwide, "Data scraping, AI and the battle for the open web", October 2025

Partner Highlights

143M

Surgery Backlog Supported

Matched volunteer clinicians to facilities in 73 countries.

4,000+

Companies Analyzed & Verified

Scraped emissions data to validate climate pledges.

2x

AI Skills Demand

Informed national reviews of AI readiness in English education.

Data for Good

The Bright Initiative significantly expanded its collaborative network in 2025 by nearly 200%, welcoming a dynamic group of new partners

76% of our new partners came from non-profit

24% came from universities and public sector bodies

Academy

Empowered 5,000+ Future Tech Leaders

with hands-on access to real-world data and industry tools.

30+ Academy Events

Including a first-of-its-kind complete course at a leading university.

New Partners





Driving Policy and Regulation in Data and AI

ARDC Alliance For Responsible Data Collection

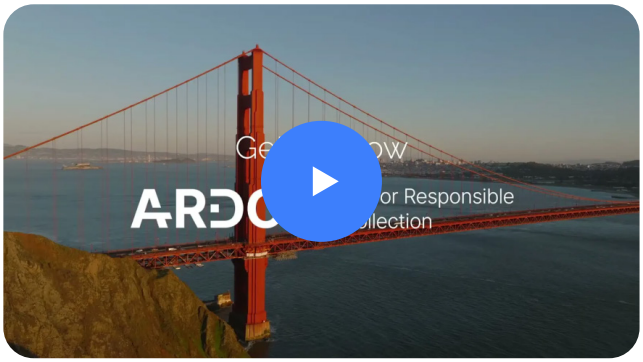
The Alliance for Responsible Data Collection (ARDC), co-founded by Bright Data, develops practical standards for responsible public web data collection, addressing legal uncertainty, operational risk, and growing tension between AI demand for data and publisher rights.

In 2025 ARDC helped shape global rules: ARDC's Technical Standards and Governance Guidelines were included in the AMAS Report on AI and Multimedia Authenticity Standards, convened by the ITU/World Standards Cooperation, positioning our approach within an international standards framework.

ARDC contributed input to the EU digital simplification package and EU AI Act implementation and participated in the Data for Policy Conference in the Netherlands, helping cement responsible public web data collection as a recognised, legitimate use case. ARDC members also engaged in IETF 124 in Montreal, bringing responsible scraping practices that respect website operators into emerging internet protocol standards.

2025 moved ARDC from a promising initiative to a recognised reference point in responsible data collection. As AI scales faster than the rules that govern its data, ARDC convenes industry, NGOs, and researchers to define clear, enforceable standards that protect an open, innovative web.

> *Organisations like ARDC provide alternative solutions to the challenges of surging demand for data: rather than blocking access and innovation, they provide* ***practical guidance for organisations committed to responsible scraping and crawling.***
>
> Jo Levy, The Norton Law Firm



ARDC San Fransisco Community Workshop
"Shaping Standards to Preserve Public Web Access"

2025 Partner Spotlights

AI for Good

Labor, Economy, and Growth

Digital Innovation, Transparency and Regulation

Environment and Sustainability

Gender Equality

Health and Wellbeing

Human Rights

Hate and Discrimination

Ashoka

Ashoka's Data Revolution

Ashoka, a global network of social entrepreneurs, is dedicated to creating systemic change. To better support their community, the Ashoka AI Lab saw an opportunity. They developed an AI-powered tool named "Rapid" to automatically track the activities, impact, and recognition of their 4,000 fellows.

Goal: To gain deeper insights into their social impact ecosystem by developing a semantic search tool that could automatically track the activities, impact, and recognition of their fellows in near real-time.

Data Collected: Public data from the websites and social media channels (LinkedIn, X, Instagram, and Facebook) of Ashoka's fellows, updated on a weekly basis.

Highlights:

- By automating the collection process with Bright Data's Web Scraper APIs, Ashoka transformed its data gathering from an impractical manual task into a streamlined, accurate, and powerful engine for insight.
- The project quickly scaled from an initial test group of 400 fellows to 700, and is now on track to encompass all 4,000 fellows within months.

Impact: With dynamic and accessible data, Ashoka can now foster greater connection and collaboration between fellows, amplifying their collective impact.

databricks X Virtue foundation

The Virtue Foundation x Databricks for Good

VF Match - AI in Action for Healthcare Delivery

143 million people in low and low-middle income countries are waiting for surgery. The Virtue Foundation partnered with Databricks for Good to solve a critical inefficiency: connecting volunteer healthcare professionals with facilities that need their specific expertise across 73 countries.

Goal: Create an AI-powered platform that intelligently matches volunteer doctors with vetted opportunities based on: Michael Burke, Technical Lead at Databricks.

- Hospital location and accessibility
- Medical capabilities and equipment
- Specialist availability and skills gaps
- Resource optimization for maximum impact

*Bright Data makes this global meta-analysis possible. We would not have been able to build our own scrapers to achieve this **scale**.*

Michael B., Technical Lead

▶ Video Testimonial Link

The challenge: Bright Data served as the foundational data infrastructure through two key products:

The SERP API enabled their team to search the web across 73 countries for medical facility information from government health sites, directories, NGO databases, and local sources.

The Web Crawler API enabled information extraction using LLMs and generative AI for key information about healthcare infrastructure.

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

Carnegie Mellon University & UC SANTA BARBARA

Carnegie Mellon University & University of California, Santa Barbara

GenAI Signals & Labor Market Shifts

In the rapidly evolving world of AI, a team of researchers from Carnegie Mellon University and the University of California, Santa Barbara, asked a groundbreaking question: Could the buzz around Large Language Models (LLMs) online actually predict the future of work? They saw that traditional methods only showed a snapshot of the current job market, not where it was headed.

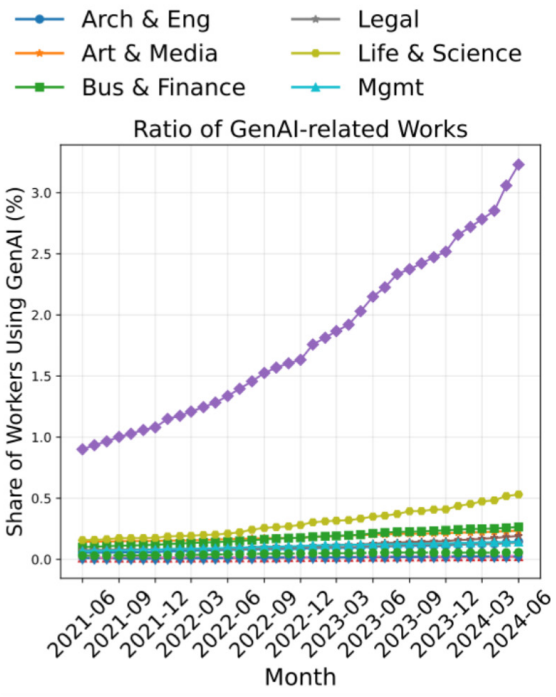
Goal: To determine if online discussions about LLMs could serve as a reliable early indicator for future labor market shifts, providing a predictive tool for workers, organizations, and policymakers.

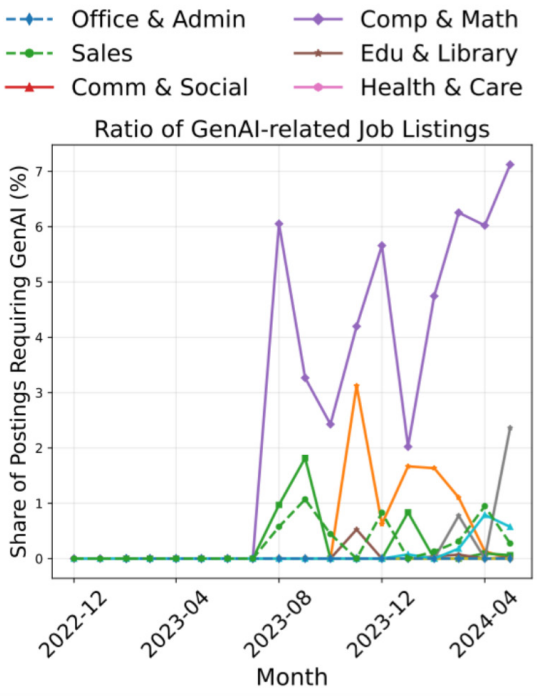
Data Collected: A comprehensive dataset including discussions about LLMs from news media, over 200,000 U.S. job postings from Indeed, and over 4 million public LinkedIn user profiles.

Highlights:

- The study proved that spikes in online discourse lead to labor market changes by 1 to 7 months, acting as a predictive bellwether for shifts in job postings, hiring, and unemployment duration.The project quickly scaled from an initial test group.
- Workers transitioning into GenAI-related roles tend to have shorter job tenures (4–12 months) compared to their peers, signaling a more dynamic and fluid career path.
- The strongest predictive links were found in knowledge-intensive fields like Computer & Math, Arts & Media, and Education, showing where the impact of GenAI is most immediately felt.

Impact: This research provides a real-time framework that complements traditional labor statistics, offering workers crucial intelligence for career planning and enabling policymakers to respond proactively to workforce disruptions.





TONY BLAIR INSTITUTE FOR GLOBAL CHANGE

Tony Blair Institute

Generation Ready: Building the Foundations for AI-Proficient Education In England's Schools

The Tony Blair Institute for Global Change (TBI) works with political leaders around the world to drive change. It is a not-for-profit organisation that provides expert advice on strategy, policy and delivery, unlocking the power of technology across all three. Its mission is to support leaders to build more open, inclusive and prosperous countries for people everywhere. TBI provides expertise in several sectors, including digital transformation, economic development, climate and energy, health and life sciences, and geopolitics and security, and works with a wide range of partners, including governments, bilateral and multilateral institutions, private corporations, academic institutions, foundations, and philanthropists who share its commitment and ambition.

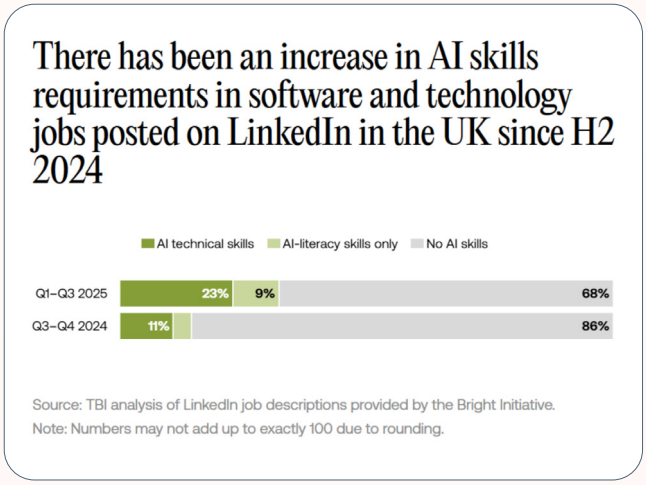
Goal: AI tools have become pervasively used in work and life, yet the English educational system struggles to adapt their curriculum and equip students with the skills they need to succeed in a fast-changing world. The stakes are high: economic resilience, national competitiveness, and social mobility all hinge on a coherent response.

Data Collected: TBI's analysis of Linkedin data provided by The Bright Initiative by Bright Data, there has been a 69% increase in demand for AI-literacy skills in professional occupations since the latter half of 2024.

Highlights:

- Only 20% of state secondary teachers reporting pupils are instructed how to use AI in their schools and just 10% report incorporating AI into subject teaching.
- Key enablers of this change, like teacher training and infrastructure, are also missing: 43% of teachers rate their confidence with using AI at just three out of ten, and only 68% of schools have dependable WiFi. Furthermore, 34% of parents report that their child lacks continuous access to a device at home for online schoolwork.

The challenge of preparing young people for a world being reshaped by AI is intensified by an AI-Technical-Skills Gap; LinkedIn data, provided by The Bright Initiative, shows AI skills requirements in software and technology jobs have more than doubled between the second half of 2024 and the first half of 2025. This underscores the critical need for comprehensive reforms to prepare students for an AI-driven future.



Impact:

The report recommends significant improvements in educational infrastructure to ensure pupils' AI readiness in England. Recommendations informed content of sessions about AI readiness in the British education system ahead of the release of national reviews on education.

JCOTS
JOINT COMMISSION ON TECHNOLOGY & SCIENCE

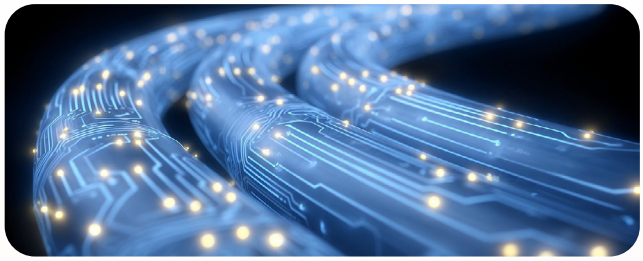
The Virginia Joint Commission on Technology and Science

Broadband Affordability & Adoption

The Virginia Joint Commission on Technology and Science (JCOTS), a legislative agency for the Commonwealth, recognized a growing problem in Virginia: the digital divide. In collaboration with researchers from the University of California, Santa Barbara (UCSB) and UC Berkeley, JCOTs conducted a statewide evaluation.

Goal: To establish a clear, defensible benchmark for internet affordability, providing a foundation for targeted legislative action across the Commonwealth.

Data Collected: Researchers collected and analyzed 62,000 address-level samples across approximately 900 census block groups in ten different localities.



Highlights:

- Ensuring access for 93% of all Virginians, a $30/month internet plan serves as the federally-aligned affordability benchmark for broadband.
- Plans priced at just $50/month become unaffordable for nearly half of all households in Virginia.
- The report offers targeted policy recommendations, including mandating greater visibility for low-cost plans and establishing a statewide standard of 100 Mbps for $30/month.

Impact: This research provided Virginia's leaders with a clear, defensible path forward to take decisive action on the digital divide. By establishing a concrete affordability benchmark, the findings have already become a cornerstone for new legislative proposals and state-level broadband programs.

LWV LEAGUE OF WOMEN VOTERS X

League of Women Voters x National Conference on Citizenship

Strengthening Civic Resilience and Fighting Misinformation in U.S. elections

2025 saw the continuation of the global rise of opposition to democracy, misinformation campaigns and rise in political polarization.

The League of Women Voters (LWV), a US-based nonpartisan, grassroots organization working to protect and expand voting rights and ensure all are represented in American democracy, sought to combat this by combating election-related misinformation and disinformation through a targeted program of research-informed education for their members and volunteers.

Goal: Transform social media data into clear, practical resources for volunteer education and engagement.

Data Collected: 40,000+ social media accounts' post data, engaged in online discourse about elections, voting, and democratic governance.

Highlights:

- State-specific rumor reports for LWV chapters in Tennessee, Wisconsin, and Georgia were developed, providing localized context and tailored guidance for voter education and communications.

1500 members reached nationwide	**84%** of LWV members feel more informed
92% More confident addressing election-related rumors	**62%** Applied briefing content to League work

SCIENCE FEEDBACK

Science Feedback

Structural Indicators to Monitor Online Disinformation Scientifically

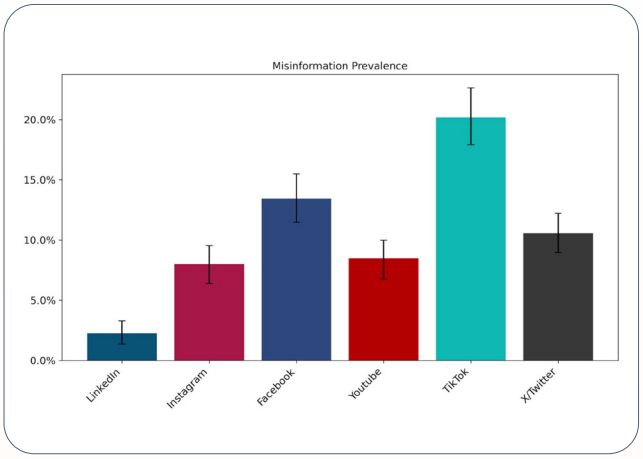
Science Feedback is a not-for-profit organization whose mission is to improve the credibility of science related information online, in the media and social media.

Goal: To provide independent, external measurement of key Structural Indicators and assess whether platforms respect users' rights to be informed truthfully and not manipulated and comply with the EU framework.

Data Collected: Over 2 million posts across social media in France, Poland, Slovakia, and Spain.

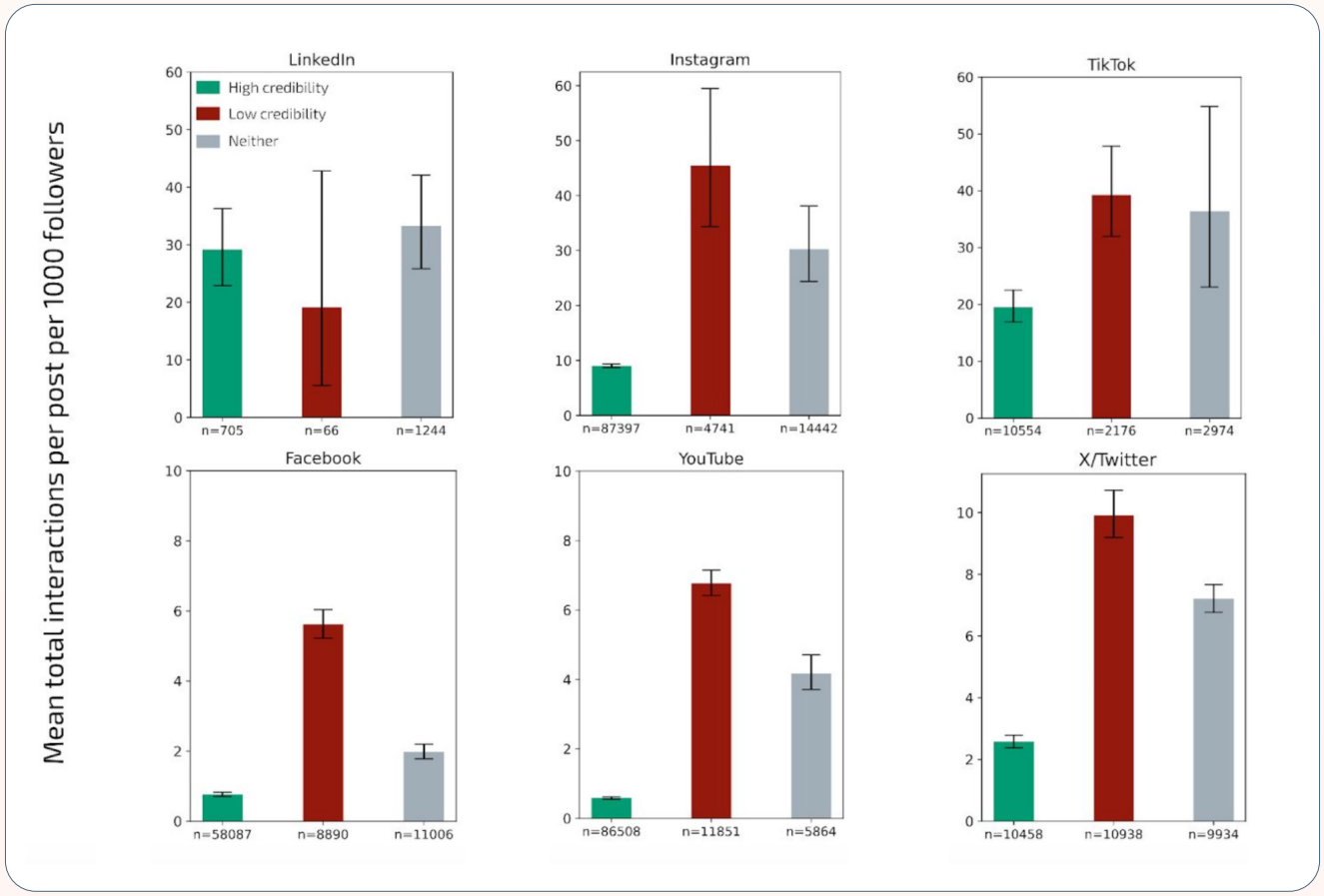
Highlights:

- TikTok: 20% of posts on public-interest topics contain mis/disinformation.



- Facebook (13%) and X (11%). YouTube and Instagram have a prevalence of mis/disinformation of about 8%.
- LinkedIn has the lowest prevalence at just 2%, showing platforms can design systems that don't reward falsehoods.
- When abusive and borderline content is included, TikTok (34%) and X/Twitter (32%) perform worst.

Impact: Under the Digital Services Act, platforms must prove they're mitigating systemic risks, including disinformation. This study offers the clearest independent benchmark yet to track compliance with the Code of Conduct.



Universiteit Leiden

Leiden University, Vincent Brussee

Reading China Better

The "Reading China Better" project, developed by Vincent Brussee at Leiden University, Netherlands, and backed by the Dutch Research Council, is an academic initiative dedicated to creating the most comprehensive platform for analyzing Chinese law and policy outside of the People's Republic of China.

Goal: To understand and analyze official Chinese documents to better grasp the country's governance, regulatory trends, and international posture.

Data Collected: he 'Reading China Better' team scrapes millions of publicly available governmental documents on Chinese law and policy and adds them into a comprehensive, searchable platform.

Highlights:

- By integrating Bright Data's technology, they dramatically scaled its data collection capabilities and gathered information from over 140 unique sources.
- Reading China Better currently provides full-text access to over 2 million government documents from national, provincial, and municipal levels in China.

Impact: The project has resulted in a permanent, accessible archive of critical Chinese sources, promoting global transparency and ensuring that vital governmental information remains accessible to the public.

> *Bright Data has been **invaluable to obtain resources that are otherwise unavailable**, as over fifty percent of government websites are inaccessible from outside China. It has also tremendously **sped up the data pipeline** by making connection to these websites faster and more robust.*
>
> Vincent Brussee, Leiden University

Global Witness

Global Witness

Holding "Mega-Polluters" Accountable

Global Witness uses cutting-edge investigations to expose the financial and political systems enabling industries to fuel the climate crisis.

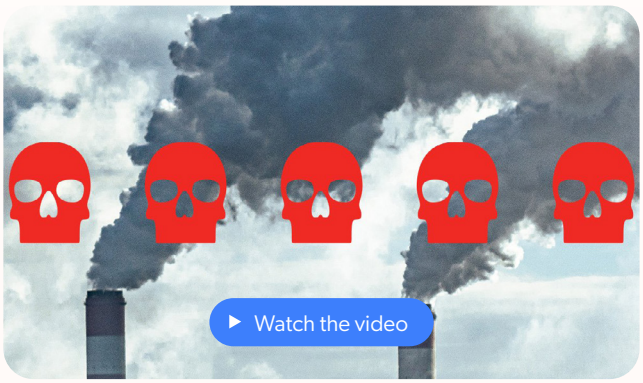
Goal: Investigate the EU's Corporate Sustainability Due Diligence Directive (CSDDD), which mandates that large corporations address environmental and human rights harms across their value chains.

Data Collected: Global Witness utilized Bright Data's SERP API to automate their data collection pipeline. By scraping emissions data from over 4,000 companies, they provided their Large Language Model (LLM) with a consistent stream of reliable public web data.

Highlights:

- **2 Billion Tons of CO2:** The annual emissions produced by companies currently covered by the CSDDD.
- Additionally, these companies were responsible for a further **23 billion tons of emissions** from their full supply chains.

Impact: By verifying corporate climate pledges against self-reported emissions data, Global Witness provided policymakers with the facts needed to keep the CSDDD robust and ensure corporate liability.



Company	HQ Country	Scope 1+2	Scope 3
Arcelormittal SA	Luxembourg	103m	8.4m
Exxon Mobil Corp	**USA**	100m	730m
Tata Steel Limited	India	82m	17m
Heidelberg Materials AG	**Germany**	68m	16.2m
Sasol Limited	South Africa	61m	35.6m
Shell PLC	UK	58m	1.08bn
RWE AG	**Germany**	54m	21.77m
Totalenergies SE	**France**	45m	417m
Lukoil PJSC	Russia	41m	no public data
Phillips 66	USA	40m	365m
Lausitz Energie Verwaltungs GmbH	Germany	38m	no public data
Linde PLC	Ireland	37m	24.37m
Air Liquide SA	**France**	36m	23.24m
British Petroleum (BP) PLC	UK	35m	322m
AP Moller-Maersk A/S	Denmark	34m	49.23m

Source: LSEG Data Library and public company reporting, European Commission, Reuters

Global Witness

C4ADS

C4ADS

Uncovering the Networks Enabling the Distant Water Squid Fleet

C4ADS is a nonprofit that produces evidence-based analysis to dismantle the illicit cross-border networks that threaten global security, economic development, and environmental stability. A recent report exposes the support networks enabling illicit distant-water squid fishing operations, an industry marked by human rights abuses and illegal, unreported, and unregulated (IUU) fishing.

Goal: Investigate the onshore and offshore enablers that aid illicit distant-water squid fishing operations in evading regulatory and law enforcement agencies.

Data Collected: C4ADS collected publicly available corporate, vessel ownership, government reported transparency records, and port data.

Highlights:

- **42%** of at-sea refueling was controlled by two vessels tied to UN sanctions evasion.
- **59%** of regulatory violations were generated by one previously penalized port agent.

Impact: This report reveals a web of actors enabling the obfuscation of persistent human rights and IUU fishing abuses in the distant-water squid fleet. While these actors provide necessary services to legal operations, some can also help illicit activities remain undetected.

The report offers actionable policy recommendations for stakeholders to increase accountability through stronger regulatory mechanisms, including beneficial ownership disclosure, due diligence on vessel risk profiles, sanctions on enabling entities, and restrictions on illicit operations moving between jurisdictions.

Wharton Impact Investing Research Lab

Uncovering the Gender Gap in University-Led Venture Funding

It's a persistent puzzle in the world of entrepreneurship: why do female founders still receive significantly less venture capital funding, even when they have the same top-tier education and resources as their male peers? A team of researchers, Tyler Wry and Christopher Bruno from The Wharton School and Tiffany Yau from the USC Marshall School of Business, decided to tackle this question head-on.

Goal: To empirically test whether elite university ecosystems act as "equalizers" by analyzing if women-led ventures founded by their students and alumni attract venture capital funding at rates comparable to those led by men.

Data Collected: Public LinkedIn profile data from students and recent alumni of the top 85 U.S. universities, resulting in a comprehensive dataset of over 26,000 ventures.

Highlights:

- Male entrepreneurs were found to be over 70% more likely to attract VC funding than their female counterparts.
- Approximately 1 in 15 male-led ventures received investment, compared to only 1 in 25 women-led ventures.

Impact: The findings suggest the problem isn't a lack of access to university resources. Instead, they point to deep-seated systemic biases within the broader venture capital landscape that universities alone cannot overcome.

University of Minnesota

University of Minnesota

Employer-Paid Maternity Leave and Women's Hiring Outcomes in India's IT Sector

Researchers examined the effects of a new labor law in India extending employer-paid, job-protected maternity leave from 12 to 26 weeks for companies with 10+ employees. Using millions of applications from a major Indian e-recruitment platform, they analyzed how the reform changed female applicants' chances of receiving interview invitations, especially at firms covered by the law based on company size.

Goal:

- Measure the causal impact of extended employer-paid maternity leave on women's interview callback rates in India's IT sector.
- Compare outcomes at firms subject to the law (10+ employees) versus those not covered.

Data Collected:

- 4.1 million job applications from a large Indian e-recruitment platform.
- Applicant gender and interview invitation outcomes.
- Firm coverage determined by company size and profitability.
- Bright Data's LinkedIn Dataset to identify which firms were subject to the new law.

Highlights:

- After implementation, women were about 22% less likely to be invited for interviews at low-profitability companies affected by the law.
- There was no observed change in the number of applications submitted by women post-legislation, indicating reduced callbacks were not driven by lower female interest or participation.

Impact: Using Bright Data's LinkedIn Dataset, the researchers identified which firms were covered by the new employer-paid maternity leave law. The study finds that employer-paid maternity can unintentionally reduce women's hiring, especially at tighter-margin companies, highlighting the need for complementary measures like anti-discrimination enforcement, cost-sharing, and shared parental leave.

Rooted

Rooted Research Collective

Tackling Nutrition Misinformation

Rooted Research Collective is a UK-based research consultancy dedicated to making sense of complex issues across food, health, sustainability, and public policy. With a strong focus on evidence-based analysis and public impact, Rooted Research specializes in uncovering misinformation and translating data into actionable insights. Their team, with over four decades of combined experience, is known for producing trusted, nuanced research that supports healthier societies and more informed public discourse.

Goal: To understand how misleading dietary narratives spread on social media, who is behind them, and what tactics they use to build trust and influence.

Data Collected: Using Bright Data's Instagram dataset and web scraper, along with other data sources, the team was able to extract and analyze public Instagram posts tied to nutrition-related hashtags and keywords.

Highlights:

- The team identified 53 super-spreader accounts with over 24 million followers and grouped them into three personas - The Doc, The Rebel, and The Hustler - to show how misinformation is tailored. By mapping tactics like fear, joy-mongering, and "sprinkling," they showed misinformation spreads through emotionally resonant narratives mixing personal stories, pseudoscience, and commercial incentives.

Impact: Research also revealed that **96% of these influencers had direct financial incentives tied to their content**, further complicating the landscape of trust and authority in online health communication.

Bar-Ilan University
אוניברסיטת בר-אילן

Bar-Ilan University, Elad Yom-Tov

Modernizing Disease Surveillance

When a traveler falls ill abroad, the trail often goes cold for public health officials back home. Seeing this critical gap, Professor Elad Yom-Tov of Bar Ilan University, alongside a team from University College London and a UK public health institute, had a revolutionary idea. What if they could find the warning signs of an outbreak not in clinics, but hidden within the online reviews left by vacationers?

Goal: To create a rapid detection system for gastrointestinal disease hotspots by analyzing online hotel reviews, enabling public health agencies to intervene more quickly and effectively to prevent the spread of illness.

Data Collected: Over 300,000 public reviews for 3,350 hotels across 20 popular vacation destinations.

Highlights:

- The research found that while rare, some vacation destinations have 3 times as many mentions of food poisoning in hotel reviews as the average.
- At the worst-offending hotels, over 3% of reviews cited food poisoning.

By identifying high-risk locations, the project provides actionable insights for public health agencies, local authorities, and travelers, ultimately aiming to improve hotel sanitation and keep people safe.

Love Justice

Uncovering Trafficking Networks with Data

Love Justice International (LJI) is an NGO dedicated to fighting human trafficking. Their primary strategy involves a method called transit monitoring, where trained staff identify and assist potential victims at transit hubs like borders and bus stations.

Goal: To enhance their preventative efforts by automating the monitoring of web platforms to identify and intercept individuals being lured into trafficking situations.

Data Collected: Job advertisements from web platforms that contain language and offers used to lure potential victims into trafficking schemes.

Highlights:

- Using Bright Data's APIs, LJI automated its data collection, pulling in 1,000 to 2,000 new job advertisements every night.
- This influx of data was instrumental in developing a graph analysis tool that exposes the hidden architecture of trafficking networks.
- The results were immediate, allowing LJI's team to intercept more than 20 potential victims within the first couple of weeks.

Impact: This new data-driven approach yielded immediate and powerful results. The organization is now building a system using AI and LLMs to score job advertisements based on "red flags" allowing their teams to focus on the most critical cases.

AMCHA Initiative

Combating Campus Antisemitism

The AMCHA Initiative is a nonprofit organization dedicated to a critical mission: documenting and combating antisemitism on U.S. college campuses. Since 2015, their public database has served as an essential resource for students, researchers, policymakers, and journalists seeking to understand the reality of antisemitism in higher education.

Goal: To automate and scale their data collection process to accurately document the rising volume of antisemitic incidents on campuses and enhance their ability to analyze the collected data.

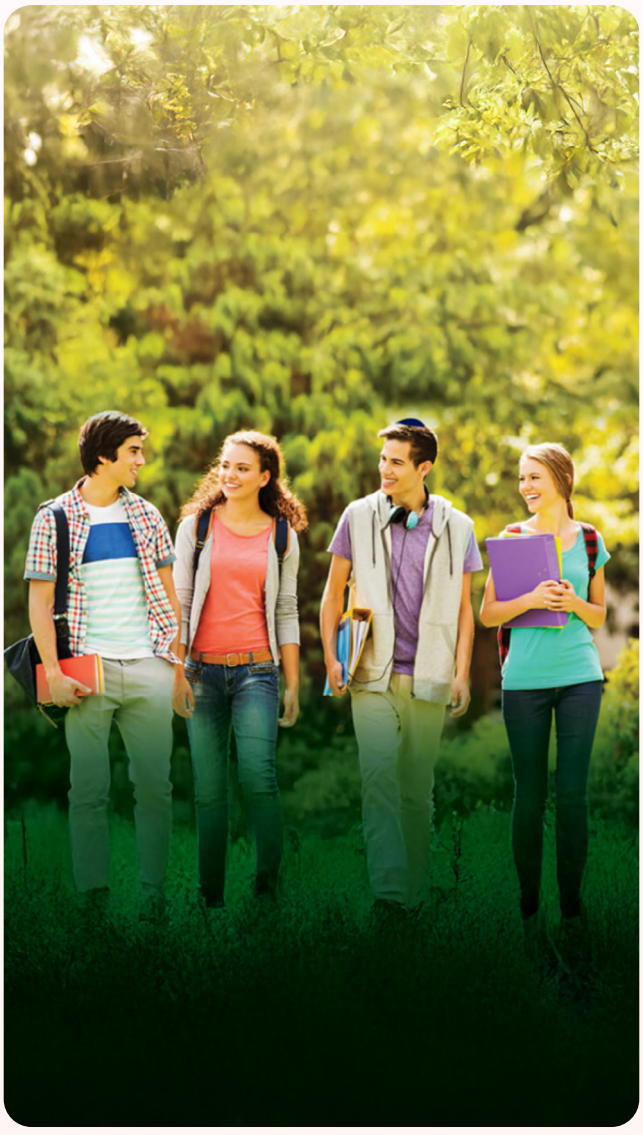
Data Collected: Information from campus newspapers, social media accounts (specifically Instagram posts), and public statements made by student and faculty groups.

Highlights:

- AMCHA dramatically scaled its monitoring, logging thousands of incidents in a short period, a massive leap from their typical 800 incidents per year.
- Using Bright Data's Web Scraper API, AMCHA automated the collection of social media posts, a key channel for campus activity.
- The new system enables advanced text and image analysis, allowing for rapid keyword searching and identification of explicit content.
- The database has grown to 11,000 entries, providing crucial insights and solidifying its role as an indispensable tool for ensuring safety and accountability on campus.

Impact: This partnership has significantly enhanced AMCHA's capacity to document antisemitism and support the broader community's efforts to ensure accountability and safety on college campuses.

> *Bright Data's tools and support have helped us quickly **strengthen our data pipeline** with high-quality data, enabling us to **efficiently analyze content** and **expand our public database** with greater speed and impact. Their partnership has been invaluable to our ambitious public-service work.*
>
> Tammi Rossman-Benjamin, Director, AMCHA Initiative



Impact Through Education and Innovation: The Bright Academy

In 2025, the Bright Initiative continued to leverage public web data and hands-on learning to drive meaningful, real-world impact. By providing high-quality data access and expert mentorship, we empowered the next generation of innovators to turn their ideas into reality.

Empowering Student Innovation:

We served as a bridge between data and action, supporting thousands of students through high-stakes competitive environments:

- Supported **3,000+ students** at the **World's Largest Hackathon**, CalHacks at the University of California, Berkeley.
- Powered over **10 additional hackathons** globally, providing the technology for rapid prototyping and problem-solving.

Scaling Academic Excellence:

We expanded our capacity to deliver industry expertise directly to the classroom, moving beyond guest lectures into core curriculum development:

- Conducted more than **15 Academy sessions** at top-tier universities worldwide.
- Developed and launched our **first full-semester university course** at Tel Aviv University, setting a new standard for data-driven education.

*Having access to real-world datasets **enabled our students to apply their skills to pressing questions in innovative ways**, while the partnership itself was collegial and highly responsive.*

*This collaboration not only strengthened the impact of the event but also **gave participants a deeper sense of what rigorous, data-driven research can achieve.***

Gunther Jikeli, Indiana University Datathon

the bright initiative | by Bright Data

SEE YOU IN 2026!