Policy Paper

July 2022

AI Startups and the Fight Against Mis/Disinformation Online: An Update

Anya Schiffrin, Hiba Beg, Juan Carlos Eyzaguirre, Zachey Kliger, Tianyu Mao, Aditi Rukhaiyar, Kristen Saldarini, and Ojani Walthrust

G | M | F The German Marshall Fund of the United States
STRENGTHENING TRANSATLANTIC COOPERATION

Washington, DC Ankara Belgrade Berlin Brussels Bucharest Paris Warsaw

July 2022

Summary

The events following Russia's invasion of Ukraine have shown again the power of online mis/disinformation. As it continues to grow and spread, there are new and continuing attempts to address this problem. These include supply-side and demand-side fixes (including media-literacy programs, fact-checking and, in Europe, new regulations) but few of these have scaled. This paper looks at one kind of supply-side attempt to tackle the prevalence of online mis/disinformation: the market for tech-based solutions that use some form of artificial intelligence (AI) machine/deep learning for content moderation, media integrity, and verification.

This paper presents the findings of interviews of 20 niche firms that use AI to identify online mis/disinformation, many of which were previously surveyed for a 2019 paper on the role of AI startups in the fight against disinformation. These companies did not release their revenue figures but it seems that the market for their services is smaller than many entrepreneurs had originally hoped, and that Google and Facebook are not relying on such firms for help in identifying online mis/disinformation. The cost of the services provided by these startups and the desire to keep things in-house and protect their activities from outside scrutiny are part of why the tech giants do not rely on small startups for help with screening online mis/disinformation. This may be why funding for these startups does not seem to have grown significantly and so more than half of them are now focusing on the business-to-business market, selling mis/disinformation mitigation services to, for example, insurance companies, large public entities, and governments, among others. There also appears to be a limited market for business-to-consumer solutions for detecting mis/disinformation.

However, continuing advances in AI as well as forthcoming regulation by the European Union and the United Kingdom will continue to spur innovation, which may stimulate demand. University initiatives, academics, journalism organizations, and cybersecurity experts are all also trying to come up with ways to identify and control the spread of mis/disinformation.

Ultimately, despite its advances, technology alone will not solve the online mis/disinformation problem. Giant social media platforms have few financial incentives to crack down on this—quite the opposite, in fact. To push social media platforms to act against online mis/disinformation and illegal speech, regulation must deftly address the issue while preserving freedom of expression. There is a further problem in the form of the political polarization that has intensified in the United States and other parts of the world. Fixing this is likely to be well beyond the role of business and technologists.

Introduction

In an address at Stanford University in April 2022, former president Barack Obama said that "one of the biggest reasons for democracies weakening is the profound change that's taking place in how we communicate and consume information."[1] He pointed to the problem of disinformation and suggested that artificial intelligence (AI) would soon exacerbate the threat.

In many ways, Obama's speech summarized an emerging consensus about the problems in the information ecosystem. Interest in these problems and discussion of solutions have grown among scholars, activists, and legislators since 2016, when investigations revealed the role of mis/disinformation in the US presidential election as well as in the Brexit referendum in the United Kingdom. Over the course of the coronavirus pandemic, worries about the effect of vaccine and public health mis/disinformation have grown. And Russia's mis/disinformation campaign before and during its invasion of Ukraine once again revealed the high stakes of the problem and demonstrated the need for efforts to combat it, including comprehensive legislation governing social media platforms.[2]

The problem goes beyond politics, public health, and national security. The spread of mistruth makes it possible to finance phishing schemes, credit-card fraud, fundraising for fake charities, identity theft, and myriad other dark web activities. What may appear to be a political campaign may actually be a fund-raising scheme. False information and manipulated media are even prevalent on dating sites and TikTok, where appearances can be altered so that the final image differs substantially from the real one.

While many technology companies are committed to building trust in what is on their sites, including affirming the origin of content and ensuring that associated audio, video, and text are authentic, they continue to invest too little in addressing misinformation or deceptive media. As noted by Mounir Ibrahim, the founder of Truepic, a photo and video verification site, fixing online mis/disinformation is "either not part of their business model or antithetical to it."[3]

As online mis/disinformation continues to grow and spread, so have attempts to address the problem. In other publications, the authors have discussed the rise of fact-checking and regulatory fixes. This paper looks at the market for tech-based solutions, many of which use some form of artificial intelligence (AI) and machine/deep learning for content moderation, media integrity, and verification. Extending earlier research conducted for the German Marshall Fund in 2019,[4] this paper focuses on a selection of niche entrepreneurial firms using AI to identify online mis/disinformation.

While many technology companies are committed to building trust in what is on their sites, they continue to invest too little in addressing misinformation or deceptive media.

According to Justin Hendrix, founder and editor of Tech Policy Press, "the problem of online mis/disinformation is substantial and unsolvable. But there are nevertheless regulatory, reputational, and other commercial reasons to address it. This has created a market for a variety of solutions bought by governments and enterprises." The analysis in this paper suggests that, while government regulation is critical, the economic and political incentives for mis/disinformation are so powerful—and the complexities of addressing it so substantial—that there is little chance the problem can be meaningfully solved by the market. The firms profiled in this paper—which have emerged to address what appears to be a relatively narrow commercial opportunity—have a role to play

1 Tech Policy Press, "Transcript: Barack Obama Speech on Technology and Democracy," April 22, 2022.

2 US Embassy in Georgia, "Russia targets Ukraine with disinformation campaign," January 21, 2022.

3 See Truepic's website. Accessed on May 10, 2022.

4 Ellen P. Goodman and Anya Schiffrin, "AI Startups and the Fight Against Online Disinformation," German Marshall Fund of the United States, September 2019.

Defining Online Mis/Disinformation

In 2017, Claire Wardle published a widely cited taxonomy of online mis/disinformation.[a] This broke down the phenomenon into seven broad categories: satire and parody, misleading content, imposter content, fabricated content, false connection, false context, and manipulated content.

Another paper by Wardle and Hossein Derakhshan also includes a rubric of the different actors and targets, such as states targeting states, states targeting private actors, or corporate entities targeting consumers.[b] It also examines intent, defining the different categories of misleading or false information as:

- Misinformation—When false information is shared but no harm is intended.
- Disinformation—When false information is knowingly shared to cause harm.
- Mal-information—When genuine information is shared in the public sphere to cause harm, such as releasing information considered to be sensitive or private.

Other scholars have formulated their own definitions and taxonomies. Meanwhile, the major platforms generally examine behavior to identify whether content is false. Meta uses the term "coordinated inauthentic behavior" when people collaborate to mislead others about their identity, activities, or intentions.

[a] Claire Wardle, Fake News. It's Complicated, First Draft, February 16, 2017.

[b] Claire Wardle and Hossein Derakhshan, Information disorder: Toward an interdisciplinary framework for research and policy making, Council of Europe, 2017.

in stemming the tide. But, as is true of virtually all the initiatives tried since 2016 to combat mis/disinformation online, market growth has been slow and available financing limited.

Methodology

For this paper, 20 companies were surveyed through interviews to learn about their developing technologies, customers, views of the overall landscape, and expectations of the effects of current and potential regulations in Europe and the United States. The research also dug into the financial incentives for these solutions, the benefits and shortcomings of using these technologies to limit the spread of harmful content online, and the latest innovations in the field. The aim was to see whether the tech giants have turned to these firms for assistance in the fight against online mis/disinformation.

The use of AI and human content moderation can be seen as part of a spectrum of solutions to contain the flow of mis/disinformation as well as to shore up media integrity and verification. In the absence of overarching regulation, several measures have attempted to address the problem. For simplicity, Anya Schiffrin in 2017 divided the measures according to demand and supply.[5] Demand-side measures tend to address audiences, or the consumers of content. They include teaching media literacy in schools so that young people can distinguish between truth, opinion, and false or misleading information,[6] and building trust in journalism[7] so that audiences can be appropriately skeptical and think critically about the source of information in order to separate truth from falsehoods. Rating efforts such as the Journalism Trust

5 Anya Schiffrin, "How Europe fights fake news," Columbia Journalism Review, October 26, 2017.

6 Theodora Dame Adjin-Tettey, "Combating Fake News, Disinformation, and Misinformation: Experimental Evidence for Media Literacy Education," Cogent Arts & Humanities, 9:1, 2022.

7 See Journalism Trust Initiative's website. Accessed on May 10, 2022.

Initiative[8] and NewsGuard[9] strive to show audiences the look and feel of quality information. Supply-side measures aim to choke off the supply of mis/disinformation online by, in part, putting pressure on platforms to refuse its circulation.

The recent blocking of Russia's RT and Sputnik by major platforms suggests that supply-side measures will remain the most powerful method to slow or halt the spread of misinformation, and their usage will likely increase once the European Union and United Kingdom pass bills aimed at stemming the harm from online mis/disinformation.[10] These bills—including the EU's Digital Services Act—may require social media platforms to step up their use of AI to identify and act on mis/disinformation online.

The Situation in 2019 and in 2022

Many of the findings of the 2019 study on the role of AI in the fight against mis/disinformation are still germane today.[11]

Technology solutions alone cannot identify all forms of online mis/disinformation—humans are needed. Since 2019, AI applications have become more nuanced and sophisticated. But without human intervention, AI cannot identify all forms of online mis/disinformation. "[Many] disinformation sites look, sound, and feel like an authentic site but publish false claims. AI can help identify content that needs to be reviewed, but I don't think AI can work without a human in the loop," observed Matt Skibinski, general manager of the ratings website NewsGuard.[12] However, effective content moderation requires vast and costly skilled human labor for forensics, network analyses, and fact-checking and are thus unlikely to scale.

Tech giants have no economic incentive to solve the problem of online mis/disinformation—government regulation is needed to push them to do more. The business models of platforms such as Facebook, Google, YouTube, and Twitter are built on engaging content, irrespective of accuracy or intent.[13] Self-regulation and codes of conduct have helped but are not enough. Although regulations are difficult to enforce, the mere awareness of them may incentivize tech giants to increase removals or down-rank mis/disinformation on their sites. However, this may not be true in countries where illiberal leaders, such as India and Brazil, decline to regulate mis/disinformation or hate speech because they themselves use it on social media for political purposes. This applies in the United States too, where Republicans and the far-right benefit from the spread of conspiracy theories online.

Online mis/disinformation is not exclusively a technology problem—it is a by-product of broader political and economic systems, polarization, and lack of trust. It is also a matter for regulators who could, for example, require consumer protection and transparency or address the ease with which sites misrepresenting their backers or intentions can be set up. "Disinformation and misinformation have been approached as a technical issue. That's the agenda of the big tech players. But more and more, elements are not technical. They are political, economic and regulatory. This is well understood in the industry," said Alejandro Romero, chief operations officer and co-founder of Constella Intelligence, which monitors online mis/disinformation.[14]

What Is New?

The companies did not release their revenue figures but it appears there is less of a market for AI solutions that track and halt mis/disinformation campaigns than previously thought. Funding for the startups surveyed

8 Ibid.

9 See NewsGuard's website. Accessed on May 10, 2022.

10 Elizabeth Dwoskin and Cat Zakrzewski, "Facebook and TikTok ban Russian state media in Europe," Washington Post, February 28, 2022.

11 Goodman and Schiffrin, "AI Startups and the Fight Against Online Disinformation."

12 See Matt Skibinski on NewsGuard's website. Accessed on May 10, 2022.

13 House Committee on Energy and Commerce, Testimony of Tim Kendall, Accessed on May 10, 2022.

14 See Constella Intelligence's website. Accessed on May 10, 2022.

does not seem to have grown significantly. Information gathered by Crunchbase, and confirmed in interviews, suggests that only four startups in this area (Truepic, Zignal Labs, Blackbird, and Logically) have received more than $10 million in funding since 2019.[15]

There appears to be a limited market for business-to-consumer solutions for detecting mis/disinformation.

In search of reliable revenue streams, more than half of the companies surveyed are focusing on the business-to-business (B2B) market, selling mis/disinformation mitigation services to insurance companies, large public entities, and governments, among others. There appears to be a limited market for business-to-consumer (B2C) solutions for detecting mis/disinformation. The different business models and companies in this sector are discussed further below. Guyte McCord, chief operations officer of Graphika, provided an overview, saying: "We are yet to see a B2C scenario. There are consumer-facing applications (fake news detection, news source ratings, etc.), but they are sold through B2B."[16] Graphika uses AI to create detailed maps of social media landscapes to discover how online communities are formed and how information flows within large networks.[17]

Finally, AI is not the only technology that is effective. Content provenance and blockchain can help authenticate the accuracy or origin of information by watermarking particular pieces of content. News organizations in a number of countries are collaborating with companies on some of these initiatives. Whether these efforts can scale remains to be seen.

How AI Screens Online Mis/disinformation

AI is an easy-to-use technology that trains computers to perform specific analytical tasks based on repeated exposure to data. Success with AI-based tools thus hinges on a compilation of rich and large data sets. The firms surveyed generally tap AI for the following mis/disinformation detection tasks: content analysis with natural language processing (NLP) and pattern recognition with machine/deep learning.

Content Analysis with NLP

NLP is an AI technique that teaches computers to understand speech and the "intent sentiment" of text at a level of comprehension that approximates that of humans. NLP combines computational linguistics—a field that applies computer science to the analysis of language—with models of other AI subsets including machine learning and deep learning.[18]

Firms using NLP for mis/disinformation detection generally draw on one of two approaches.

One approach—which to date has achieved less success—involves training an algorithm to classify assertions as true or false by showing it large numbers of assertions that have been manually labeled as true or false. For NLP to accurately identify mis/disinformation using this method, consistent definitions of the type of speech need to be identified and sufficient data for training, validation, and testing is required. Unless the models are built with adequate, unbiased, and representative datasets, such as data from different platforms or geographic regions, results can be biased or misleading.

The other approach—more practical at present and more widely used—is to use AI to match text assertions with assertions in a fact-check database. With this method, the AI does not need to figure out what is true or not, but instead essentially performs a keyword search to match claims with fact-checks. This latter approach similarly requires a sizable database of fact-checks but does not require data to train the AI—the AI in this case is said to be "pre-trained."

15 See Crunchbase's website. Accessed on May 10, 2022.

16 See Graphika's website. Accessed on May 10, 2022.

17 See Solutions on Graphika's website. Accessed on May 10, 2022.

18 Steven Johnson, "A.I. Is Mastering Language. Should We Trust What It Says?" New York Times Magazine, April 15, 2022.

Pattern Recognition with Machine/Deep Learning

Machine learning and deep learning—of which NLP is a subfield—involve training an algorithm on text and non-text data signals to imitate human learning, identify actor networks, and understand traffic patterns. A common example of machine learning is recommendation engines embedded in apps used by platforms to collect user data, feed inputs into their algorithms, and note user habits and preferences so that companies can better predict trends and user behavior.

An example of pattern recognition through machine learning was provided by Jennifer Granston, chief customer officer at San Francisco-based startup Zignal Labs:

> We don't label content as "true or false" or "harmful or not harmful." NLP and different sentiment models allow us to identify, for example, what accounts on Twitter behave as if they are using a high level of automation—or which accounts are likely to be bots, click farms or troll farms—the ones propagating bad content. [19]

Blackbird.AI, a New York-based startup, uses machine learning and other automation and AI technologies to uncover patterns of malicious behavior and harmful narratives. These patterns might indicate the nature of relationships between users and the content they share or identify the connection and shared beliefs of various online communities through what it calls a "coalition" signal. An example of an AI startup using deep learning is London-based Fabula AI. Founded in 2018, it pioneered the field of "geometric deep learning." Fabula AI maps geometric routes of how online content spreads on social networks through its deep learning algorithms. As a result, the detection of malicious information or actors does not require reading or understanding of content.

19 See Zignal Labs' website. Accessed on May 10, 2022.

The Business Landscape

The AI mis/disinformation market is wide and varied but with limited potential for smaller firms. As Tech Policy Press' Justin Hendrix put it:

> Some big players/whales in the business offer enterprise solutions. Then there are a very small number of well capitalized startups. Everyone else are guppies and minnows. The dream still seems to be that regulation may change the game. But is the real story that these massive, centralized platforms are closed and like to build their own solutions, so there simply isn't a well developed marketplace? And maybe there never will be?

But big tech has proved to be a tougher customer, in sharp contrast to the hopes expressed by the firms surveyed in 2019. At that time, many hoped to expand their customer base to big platforms. As Danielle Deibler, co-founder and CEO of Marvelous AI, put it:

> We would love to sell to Google and Facebook, but these large companies are trying to solve this problem themselves, and want to build [the tools] themselves. They don't want to be subject to public scrutiny for their algorithms. I don't see them paying lots of money for third parties.[20]

For example, rather than turning to the startups surveyed here, Meta tends to outsource[21] much of its content moderation to third parties such as accenture,[22] concentrix,[23] and TaskUs.[24] These companies frequently hire content moderators to make decisions about content removal and ranking. Meta is notorious for not itself hiring enough content moderators, with

20 See Marvelous AI's website. Accessed on May 10, 2022.

21 Adam Satariano and Mike Issac, "The Silent Partner Cleaning Up Facebook for $500 Million a Year," The New York Times, August 31, 2021.

22 See accenture's website. Accessed on May 10, 2022.

23 See concentrix's website. Accessed on May 10, 2022.

24 See TaskUs' website. Accessed on May 10, 2022.

those they do hire often based in the Philippines or India where wages are relatively low.[25]

Corporate secrecy also deters Meta and other platform heavyweights from hiring firms like those profiled here. Factmata's Antony Cousins said Facebook does not "want to work with third parties because they don't want people to see how bad the problem is." A notable exception is Kinzen,[26] whose co-founders Mark Little and Áine Kerr have worked with large platforms on fact-checking. Barred by a non-disclosure agreement from going into details, Little noted that companies like Twitter are now scaling up their use of AI, anticipating that online mis/disinformation and other threats will grow before events such as elections.

Low Growth, New Business Model

The 2019 paper on the role of startups mentioned Silicon Valley's monopsony and how hard it would be for entrepreneurs to scale up their businesses. This time, many of the startups reported that the lack of a growth path prompted them to change their strategy. While all noted the ubiquity of mis/disinformation online, which suggests a healthy market for their services, not all were able to grow their revenues. Some have shifted their customer base from the public to businesses. Others found a scant market for their services and have narrowed their focus. Nearly all sell services to companies that track how their brand is referred to online.

The services provided by the firms fall into the following, often overlapping, categories:

- Ensuring security and mapping for governments
- Combatting online extremism
- Monitoring brand safety—often for corporate clients
- Nonprofits and open source
- Automated fact-checking
- Improving the quality of user/reader engagement to grow target audiences

Security

Companies like Constella Intelligence analyze abnormal digital patterns and emerging digital risks such as mis/disinformation and online malign campaigns. They examine data across the full Internet—from surface digital communities and social networks to deep and dark web forums and breached data. These companies map techniques, tactics, and procedures to understand the sort of mis/disinformation being spread and by whom and how. To protect the integrity of authentic sites, companies, governments, intelligence agencies, nongovernmental organizations, and media companies may seek these services to understand potential digital risks to customers, constituents, revenues, product lines, or "real news" that comes from the unknown corners of the Internet.

But the source of mis/disinformation is becoming increasingly muddy. "More and more of the disinformation toolbox is being used by local actors," said Alejandro Romero, chief operations officer and co-founder of Constella Intelligence. It is increasingly difficult to distinguish between foreign versus local mis/disinformation. False or misleading content is ever more sophisticated and under the radar. Disinformation is now a persistent threat supported by well-organized actors that sell their services—from bots to tailored deepfake videos—in specialized deep and dark web forums, making these accessible to anyone.

Disinformation has also become more ubiquitous. "A Crime-As-A-Service model implies that you don't even need to fully understand the technology. Bad actors can rent a network of bots or a set of stolen identities to launch their malign campaigns," Romero added.

Brand Safety

Other firms garner revenue by selling brand safety—services that use AI to help identify and counter mis/

25 Satariano and Issac, "The Silent Partner," 2021.

26 See Kinzen's website. Accessed on May 10, 2022.

disinformation that may harm a firm's online image or reputation. One such firm is London-based Factmata whose chief executive officer, Antony Cousins, said that firms know that "getting involved in a conversation on social media is a great way to find out what people are saying about your product."[27] Factmata helps companies track discussion of their brand across multiple social media platforms through fully automated AI. Cousins noted: "We do not judge the content ourselves. No humans are involved. We do not put our biases onto the content."

Jay Pinho, of the brand safety division of Oracle,[28] said his office judges "millions of pages a day online and then categorizes on an automated basis what they're about so brands can make a decision about what they want to be near or far from." For example, advertisers want to keep well away from controversial content such as that involving obscenity, hate speech, terrorism, or military conflict.

Hoping to increase advertising revenue from companies that care about where their brand is seen online, news organizations are reminding advertisers that they provide accurate and high-quality information. To further their goal of getting more advertising revenues, many news organizations[29] have joined coalitions to support quality advertising such as United for News.[30]

Working with News Outlets—Another Path to Profit

While news organizations are using the brand-safety argument to obtain more advertising revenue, several startups are trying to sell services to newsrooms, including fact-checking, evaluating the origin of content, improving the tenor and quality of online discussions, and monitoring safety threats to journalists. However, news organizations are skeptical customers and many prefer to develop such products in-house so they can keep a tight grip on quality, safety, and ethical standards. Paul Glader, founder and chief executive officer of Vett News, said:

> It's hard to sell anything into the [media] industry right now. It will often only try new things if convinced beyond a shadow of a doubt that it will make the news publisher more money. So we plan to win more business with convincing data.[31]

Automated Fact-Checking Systems for News Outlets

These services verify written or spoken statements, numbers, and claims. They strive to build audience trust in the hope this will produce more revenue from audiences who value accurate information. For that reason, news organizations are collaborating with tech companies on such products.

Content Provenance

Some news organizations have become involved with tech companies to authenticate information and images. One example is the Adobe-led Content Authenticity Initiative, which has an open-source method of verifying content.[32] It is also helping to establish standards for the field through the Coalition for Content Provenance and Authenticity.[33]

Crowd-sourced Claims Checks

Netherlands-based nwzer takes a different approach with publishers, including newsrooms.[34] The company encourages an audience-driven approach to verify the accuracy of content, similar to that explored in James Surowiecki's *The Wisdom of Crowds.* It uses an NLP-based algorithm for readers to self-moderate and

27 See Factmata's website. Accessed on May 10, 2022.

28 See Oracle's website. Accessed on May 10, 2022.

29 Rebecca Frank, "Flaws in Ad Tech Contribute to False Perceptions of Brand Safety, Ad Blocking, and Disinformation," Medium, January 14, 2021.

30 See United for News on Internews. Accessed on May 10, 2022.

31 See Vett News' website. Accessed on May 10, 2022.

32 See Content Authenticity Initiative's website. Accessed on May 10, 2022.

33 See Overview on Coalition for Content Provenance and Authenticity's website. Accessed on May 10, 2022.

34 See nwzer on EU-Startups' website. Accessed on May 10, 2022.

self-regulate against mis/disinformation in its online comments sections. "You have to make sure that the crowd self-corrects [against mis/disinformation]," explained Karim Maassen, the company's founder and chief executive officer. Founded in 2017 and funded by Google News Initiative, nwzer says it earns revenue and is profitable.[35]

Memetica works to identify threats against journalists or other public figures for newsrooms and private security clients.[36] "Existing platforms and law enforcement are still catching up with what it means to be an average person at the center of a disinformation campaign," explained Ben Decker, its founder and chief executive officer.

Non-profit Efforts: Universities and Open Source

Along with private companies, universities have become involved in efforts to rein in mis/disinformation on the Internet. Some efforts that are funded by foundations include a Bill & Melinda Gates Foundation $250,000 grant to Harvard University in 2018 "to understand the scale and nature of the mis- and disinformation problem and [to] determine how to effectively debunk health-related and other falsehoods traveling on social media platforms,"[37] and the Center for Security and Emerging Technology at Georgetown University, which does public policy non-partisan research.[38]

There are also many academics researching AI and misinformation, such as Sarah Oates at the Philip Merrill College of Journalism at the University of Maryland,[39] and Katherine McKeown[40] at the Data Science Institute at Columbia University.[41]

Other universities are also incubating AI start-ups to tackle the mis/disinformation problem. Columbia Technology Ventures'[42] Vidrovr[43] creates technologies to analyze video, which is often used to spread mis/disinformation. Shih-Fu Chang, the interim dean at Columbia University's Fu Foundation School of Engineering and Applied Science,[44] is chief technical advisor at Vidrovr.

For-profit companies also work with universities. Open AI makes its Application Programming Interface available to help others "train" datasets with human input.[45] Marvelous AI's StoryArc analyzes narratives to track and quantify mis/dis information.[46] The data provided by StoryArc helped a research team from the University of Maryland's Philip Merrill College of Journalism track character and identity attacks on Twitter targeting female candidates during the 2020 US presidential primaries.[47]

Conclusion: Regulation Will Create Innovation

Despite advances, technology alone will not solve the online mis/disinformation problem. Giant social media platforms have few financial incentives to crack down on this—quite the opposite, in fact. To push social media platforms to act against online mis/disinformation and illegal speech, regulations must deftly

35 Ibid.

36 See Memetica's website. Accessed on May 10, 2022.

37 Bill & Melinda Gates Foundation, Committed grants – Harvard University, August 2018.

38 See Center for Security and Emerging Technology's website. Accessed on May 10, 2022.

39 See Sarah Oates' page at the Philip Merrill College of Journalism, University of Maryland. Accessed on May 10, 2022.

40 See Kathleen R. McKeown's page at the Data Science Institute, Columbia University. Accessed on May 10, 2022.

41 See Data Science Institute's website. Accessed on May 10, 2022.

42 See Technology Ventures' website. Accessed on May 10, 2022.

43 See Vidrovr's website. Accessed on May 10, 2022.

44 See Shih-Fu Chang's page at Fu Foundation School of Engineering and Applied Science, Columbia University. Accessed on May 10, 2022.

45 See Open AI's website. Accessed on May 10, 2022.

46 See Marvelous AI's website. Accessed on May 10, 2022.

47 Marvelous AI, Marvelous AI Teams Up with Scholar Sarah Oates to Track How Twitter Commentary Disadvantages Female Candidates in the 2020 U.S. Primaries, October 22, 2019.

address the issue while preserving freedom of expression. Marvelous AI's Danielle Deibler said:

> We see ourselves as part of an ecosystem. One group is not enough to fight misinformation. You need policy and regulation and you need the social media companies and journalists to not spread and propagate [untruths or deceptive claims]. You also need people to help keep the government and journalists in check. Hopefully public sector companies and academics can do it.

Most effective, she added, would be a federal privacy law rather than a patchwork of state regulations.

There has been progress. The EU's Digital Services Act (DSA), which was agreed in April 2022,[48] and the United Kingdom's proposed Online Safety Bill[49] require platforms to conduct risk assessments and share plans with regulators to address potential harms caused by illegal content. In Europe this can mean several forms of speech, including hate speech or incitement.

The DSA focuses on risks to society, while the UK bill focuses on risks to individuals. Germany's NetzDG law, passed in 2017 and modified in 2021, was an inspiration for the DSA and includes fines for tech giants that have a pattern of knowingly disseminating illegal speech.[50] French regulators say the DSA is similar to banking regulation because rather than supervising every transaction, it requires companies to build systems to mitigate risk.

Due to the United States trailing the United Kingdom and the EU in regulating technology platforms and the expansive view of the First Amendment upheld by US courts in recent years, Europe may well set the standard for other countries. What this means for the future of the niche firms profiled here remains to be seen. Nonetheless, regulation is likely to continue to evolve and laws about online harm will spur demand for the types of services described in this paper as well as new opportunities for innovation. With regulation coming sooner in the EU than in the United States, there may be short-term business opportunities for European companies trying to identify potentially harmful mis/disinformation online. Regulation is likely to continue to evolve and laws about online harm will spur demand for the types of services described in this paper as well as new opportunities for innovation.

Regulation is likely to continue to evolve and laws about online harm will spur demand for the types of services described in this paper as well as new opportunities for innovation.

It is also important to note that the types of technologies developed to assess content at scale can be employed quite differently by authoritarian regimes that seek not to create guardrails that preserve free expression, but rather to contain or limit it. Firms in this field must be mindful of the environments in which they operate, which can change suddenly. There are no easy answers when it comes to governing human expression, only trade-offs.

48 European Commission, The Digital Services Act: ensuring a safe and accountable online environment, accessed on May 10, 2022.

49 UK Department for Digital, Culture, Media & Sport, Government, Online Safety Bill: factsheet, updated on April 19, 2022.

50 Center for Democracy & Technology, Overview of the NetzDG Network Enforcement Law, July 17, 2017.

Appendix A. Company Profiles

ActiveFence

Rachael Levy, director of geopolitical risk: "ActiveFence is a leading tool stack for trust and safety teams. Anyone can throw AI at a problem. It's not going to solve it. Hordes of content moderators are also not going to be a robust solution. Trust and safety teams need to be agile, efficient, and accurate—what ActiveFence enables them to be—by uniquely combining technology and human expertise to analyze campaigns at scale. Our analysts are subject-matter experts in particular fields, geographies, and languages."

Overview: An AI-based software platform used by trust and safety teams worldwide to reduce harm and keep users safe on online platforms. The platform focuses on a range of online harm, unwanted content, and malicious behavior, including disinformation, fraud, hate speech, terror, nudity, and that which could endanger the safety of children. Its advanced AI and team of experts continuously collect, analyze, and contextualize data.

Funding: Raised $100 million to date. Backed by Silicon Valley investors CRV and Norwest.

Staff Size: 280

Launch Date: 2018

Future Plans: ActiveFence "proactively search[es] the darkest corners of the web for bad actors to understand the sources of malicious content," Noam Schwartz, co-founder and chief executive officer, told TechCrunch in 2021.[1]

AverPoint

Shouvik Banerjee, founder and chief executive officer: "We focus on the demand for higher quality information and empower individuals with healthy news habits and critical thinking skills. This is just as important as the other approaches, which focus more on information supply: machine learning for content moderation, fact-checking, regulations, and standards."

Overview: An app and browser extension that helps individuals and communities build healthy news habits and media literacy skills. AverPoint measures and analyzes a user's news consumption, and then nudges them to increase their source, topic, and geographic diversity. AverPoint's credibility layer lets readers interact with articles to ask questions, request reviews, and evaluate evidence.

Funding: Undisclosed. No prior public announcements.

Staff Size: 5

Launch Date: 2016

Future Plans: Charge US consumers for access to paywalled content from news partners. Currently, readers measure their reading automatically through the browser extension and manually through the mobile app. Over time, AverPoint plans to integrate its services directly into publisher websites and apps so it works more ubiquitously in the background.

Blackbird.AI

Naushad UzZaman, co-founder and chief technology officer: "Opening up data with de-identification (removal of users' personal information) to the research community and externally [expands awareness of] the spread of mis/disinformation on big platforms. For example, because Twitter allows the collection of data using their Application Programming Interface, people know how much disinformation is being spread on Twitter. However, similar research cannot be conducted on any of Meta's platforms such as Facebook and Instagram, while a large amount of harmful content spreads on these platforms."

Overview: A Software-as-a-Service platform using AI to spot, predict, and examine emergent threats and provide risk intelligence by identifying mis/disinformation. The platform uses a combination of five core signals related to content, context analysis, and pattern recognition to surface threats: manipulation, deception, narratives, networks, and coalition.

1 Ingrid Lunden, "ActiveFence comes out of the shadows with $100M in funding and tech that detects online harm, now valued at $500M+," TechCrunch+, July 27, 2021.

Funding: A first round of funding in 2014 from a Silicon Valley incubator. Blackbird.AI raised $10 million in Series A[2] in 2021, led by Dorilton Ventures. Other investors include Generation Ventures, Trousdale Ventures, StartFast Ventures, NetX, and Richard Clarke, former chief counter-terrorism advisor for the National Security Council.[3]

Staff Size: 30

Launch Date: 2017

Future Plans: The next iteration of the platform will include image and video analysis. The startup aims to analyze any data on the web regardless of the sector and "to stop harm, before it gains traction so that the online space is used to elevate society," said Naushad UzZaman.

Constella Intelligence

Jonathan Nelson, digital intelligence specialist: "We pivoted to the analysis of the dark web: How can we connect the dots between what is happening on the surface and in the dark web such as in illegal markets?"

Overview: A digital threat protection software that taps proprietary data, technology, and human expertise. Services include executive cyber protection, brand protection, threat intelligence, fraud protection, and geopolitical intelligence monitoring. Its clients include companies on the FTSE 100 and Fortune 500, the UK government, EU institutions, and global tier one banks. Software licenses allow clients to use the tools to understand key trends.

Funding: Received more than $60 million funding: €12.5 million ($13.8 million) in Series A in 2016, $18 million in Series B in 2018, and $30 million in Series C in 2020. Major investors include Adara Ventures, Benhamou Global Ventures, C5 Capital and Forge-Point Capital.[4] Constella does not disclose whether it is profitable. It earns revenue from third-party licensing.

Staff Size: 166

Launch Date: 2020

Future Plans: In September 2021, Constella launched its Dome Platform, which provides Executive Cyber Protection.[5] This automated platform can be integrated with companies' existing IT and security infrastructure, expanding protection to any number of employees. According to chief executive officer Kailash Ambwani, "This platform [allows] companies the opportunity to monitor diverse digital sources. So, limiting its digital threat monitoring services to a few executives or employees will no longer be necessary."

Cyabra

Dan Brahmy, co-founder and chief executive officer: "I don't think Big Tech will develop solutions in-house because they lack the will and financial incentive. And existing regulations are not forcing them to do anything substantial. [Big Tech] will not acquire companies like ours unless they have to. And even if they have to, they might acquire companies to silence them."

Overview: A Software-as-a-Service platform that conducts narrative and pattern analysis to measure mis/disinformation. Cyabra's customers are primarily in the financial services industry with others in the consumer brands, media, and public sectors. "Our most interesting market is the private sector: large companies with brand or reputational issues. This is why financial services, and advertising and data analytics agencies like TBWA[6] are interesting to us," said Dan Brahmy.

Funding: Received $1.2 million in the pre-seed round in 2018 and $5.6 million in Series A in 2021, the latter led by OurCrowd.[7] Other investors include Peter Thiel's Founders Fund, Harpoon Ventures, Alabaster, Accomplice, Red Shepherd Ventures, Summus Z, TAU Ventures and Capital Y management. Among angel

2 Series A, Series B, and Series C are external funding rounds that may follow seed funding.

3 Sai Venkatesh, "Blackbird.AI raises $10M in Series A from Dorilton Ventures and others," SaaS Industry, September 22, 2021.

4 Constella Intelligence, Dealroom.co, accessed on May 10, 2022.

5 See Constella Dome on Constella Intelligence's website. Accessed on May 10, 2022.

6 See TBWA's website. Accessed on May 10, 2022.

7 Kate Park, "Cyabra gets $5.6M Series A to launch news disinformation detection analysis tools," TechCrunch+, October 26, 2021.

investors are former global co-general manager of Samsung Pay Will Graylin and former chief product officer of Tinder, Brian Norgard. Cyabra expects additional funding in 2022 of low-mid eight figures.

Staff Size: 25

Launch Date: 2018

Future Plans: To expand online reach beyond written online content into speech, including that in podcasts, online videos and streaming.

Graphika

John Kelly, chief executive officer, and Guyte McCord, chief operations officer: "No magical algorithmic toolset will solve online mis/disinformation. We hope to deliver more human-driven solutions around it."

Overview: Graphika's platform identifies and tracks the formation of communities online, and maps how influence, narratives and information flow within large-scale networks. This allows it to map structural relationships among social media actors, and segment these complex networks based on patterns it observes in relationships.

Funding: Received over $8 million in total: $3.4 million in 2014 and $4.9 million in Series A in 2017.[8] Major investors include Lavrock Ventures, Social Media Enterprises and First In.

Staff: 50

Launch Date: 2013

Future plans: Graphika now sees its primary market as B2B. This year, it will deliver an important technology product which will allow businesses to have a subscription service for important topics being tracked, mapped, and analyzed by Graphika. This actionable intelligence will be important for companies impacted by topics and communities that matter to them the most.

The Factual

Arjun Moorthy, co-founder and chief executive officer: "Think of AI as a complement to human thinking, not a replacement. While computers can tell if a specific fact is true or false, tying facts together and understanding the news requires tremendous context and history, which is where humans excel. Hence AI can help identify all the facts and make it easier for humans to reach their own conclusions."

Overview: An AI-enabled news platform that finds unbiased news, pushing content consumers to expand their trustworthy news sources. The Factual analyzes the credibility of more than 10,000 news articles daily based on the quality of sources, tone of writing, author expertise, and historical site scores, surfacing the stories based on these factors across the political spectrum in a daily newsletter, app, and website.

Funding: Received $1 million from angel and venture capital investments to maintain operations. Investors include HubSpot co-founders Brian Hallian and Dharmesh Shah, former chief executive officer of Lola, board member of Repsly Mike Volpe, and former president of Pinterest Tim Kendall. The Factual also received investments from Defy Ventures and Matrix Partners. The Factual is planning a seed round investment in 2022.

Staff Size: 10

Launch Date: 2019

Future Plans: To validate broad demand in the market for quality journalism with 100,000 subscribers or more.

Kinzen

Mark Little, founder and chief executive officer: "Two things [around information ecology] have emerged that have not been properly teased out. First, insufficient defense of the protection of free speech through content moderation. The only way to protect free speech is to stop its weaponization. Where are the articulate voices defending good moderation as crucial to the survival of free speech and democratic values? Secondly, the issue of provenance. Quality information is not moving fast enough. How do we get it to move faster? How can we accelerate good information [amid] an information gap?"

8 "Graphika Valuation & Funding," Pitchbook, accessed on May 10, 2022.

Overview: A proprietary content understanding engine that helps content moderators and policymakers stay ahead of information threats such as misinformation and hate speech through a blend of human judgment and artificial intelligence. Kinzen's technology detects and scores information risk in text, audio, and video content.

Funding: Raised a total €3.45 million including €1.65 million in a seed round in 2020 led by Danish start-up accelerator FST Growth.

Staff Size: 40

Launch Date: 2017

Future Plans: Scaling its analysis and data services to all major global languages (to 20 from 12 in 2022), and further investment in machine learning models optimized for the detection of information risks. In the long term, Kinzen hopes to support user-controlled moderation services in partnership with global technology companies.

Marvelous AI

Danielle Deibler, co-founder and chief executive officer: "When it comes to building a model, it's human powered. Humans create the labels, and flesh out [the] nuances of narratives. You need enough examples [of these narratives] to build enough models. But you can do a lot with a very small amount of data. The narrative pipeline requires human care and feeding. The model takes into account emotional characteristics like anger, sadness, and joy."

Overview: An early-stage startup founded by tech industry veterans that is building an augmented analytics platform to generate actionable insights about online narratives. Marvelous AI combines human intervention with natural language processing, computational linguistics, and machine learning to detect mis/disinformation and harmful narratives.

Funding: Mostly funded by friends, family, and venture funds. The investment covers roughly one year of its operations.

Staff Size: Founders Danielle Deibler (chief executive officer), Christopher Walker (chief operations officer), and Olya Gurevich (chief scientist).

Launch Date: 2018

Future Plans: To provide power and tools to continue uncovering mis/disinformation. It does not seek to be acquired. It seeks a way to tamp down mis/disinformation or, at a minimum, halt its flow.

Memetica

Ben Decker, founder and chief executive officer: "From an investor point-of-view, [the market] is oversaturated; from a business growth perspective, the bubble will burst somewhere, somehow, and probably consolidate to a few key players."

"Our main differentiator from those who provide threat monitoring services is our overemphasis on the human and our white-glove customer service," explained Decker. "We're like the luddites of the industry. We hardly use AI or machine learning. We ingest raw data; real human analysts evaluate and query it for relevant items. [This] allows us to catch what more robust AI systems aren't catching due to lack of context."

Overview: A digital investigations consultancy that provides intelligence and risk advisory services to media companies, civil society organizations, whistleblowers, and other public figures facing threats from coordinated harassment, disinformation campaigns, and violent extremism. Memetica does this through a combination of open-source and human intelligence gathering practices to collect raw data from fringe communities on sites like Telegram, 4chan, and others, as well as mainstream platforms like Twitter and Facebook. Analysts uncover, investigate, and flag threats of harassment campaigns and real-world violence. Working with several R&D partners, Memetica taps its team of experts and advanced tools to improve digital protection for public safety.

Funding: Launched in 2019 with a small research grant from Jigsaw, a technology incubator created by Google that supports technology solutions to combat

disinformation, censorship, toxicity, and violent extremism online.[9] Between 2019 and 2021, Memetica increased its revenue by 250 percent, growing from $160,000 in its first year to $570,000 in 2021.

Staff Size: 5

Launch Date: 2019

Future Plans: Memetica will continue developing tools and products to provide threat intelligence and risk mitigation services to media companies, civil society organizations, and other public-facing entities. Focus industries for expansion include healthcare, entertainment, and sports. The company also plans to fortify existing and future R&D partnerships by working with major research labs to improve industry standards for curbing imminent harms and reducing long-term structural vulnerabilities to election integrity and public safety.

NewsGuard

Matt Skibinski, general manager: "Every platform has said that they were one algorithm away from solving the problem. [For us], it wasn't just an algorithm problem. It was a media literacy problem and a journalism problem."

Overview: A browser extension tool that rates websites to equip users with context on the sources they encounter. NewsGuard has rated 7,500 domains in the United States, the United Kingdom, France, Italy, and Canada, and it has a subscriber base of 100,000 users, with wider distribution through partnerships with Microsoft and the American Federation of Teachers. The tool's overall rating is based on nine criteria that determine journalistic credibility. These include false claims published regularly by the source, funders, and any use of deceptive headlines. NewsGuard contacts publishers so they can rectify errors.

Funding: Raised $6 million in a seed round in 2018 led by Publicis Groupe, an ad agency holding company.[10] Other investors include Cox Investment Holdings, the John S. & James L. Knight Foundation, Blue Haven Initiative, and journalists Steven Brill and Gordon Crovitz.

Staff Size: 38

Launch Date: 2018

Future Plans: In February 2022, NewsGuard partnered with the Joint Research Centre of the European Commission.[11] Its data will support the JRC's research in source reliability, narrative detection, and the spread of these narratives in different languages.

Nobias

Tania Ahuja, founder and chief executive officer: "Trust is at an all-time low and social media has played a major role in this decline. Our goal is to give young investors a tool that provides the missing intelligence they need to start to trust their own critical thinking. We don't combat fake news and misinformation directly, but indirectly by providing our users with signals of credibility and bias in articles and among authors they see online, even before clicking or opening a website. We hope these signals will force our users to slow down and think critically about the information they read and share online."

Overview: A B2C data-driven software that uses NLP to identify bias in online political, financial, and health articles with the goal of promoting responsible and inclusive technology to protect consumers from misleading or deceptive online content. Nobias does not disclose its customer base.

Funding: Its source of revenue is the finance application in which users see stock insights like ratings from Nobias-rated analysts and read articles from popular finance authors rated with Nobias insights including sentiment of the article and the credibility of the authors based on the accuracy of their recommendations over the past three years.[12]

Staff Size: 12

Launch Date: 2017

9 See Jigsaw's website. Accessed on May 10, 2022.

10 "Finsmes, NewsGuard Raises $6M in Funding," March 6, 2018.

11 NewsGuard, "NewsGuard partners with the Joint Research Centre of the European Commission," February 10, 2022.

12 See Nobias' website. Accessed on May 10, 2022.

Future Plans: Will run the business based on a freemium model, with basic features remaining ad-free.

nwzer

Karim Maassen, founder: "In The Netherlands during Easter, shoppers at grocery stores can guess how many candies are in a bowl. The closest guess wins. But research shows that all the guesses averaged out are the right answer. Bring that to 'the wisdom of crowds' and apply it to fact checking content for readers with a common goal. Readers will quickly validate any informal commenting or any piece of opinionated text [and the nwzer algorithm will learn from that activity]."

Overview: A user-generated news agency whose algorithm, which is built on NLP, makes possible self-moderating, -correcting, and -evaluating by readers. Working in the background on the behalf of publishers, its algorithm learns from readers' comments and separates valuable from less valuable information to support fact checking.

Funding: Started with a €40,000 grant from the Google News Initiative in 2017, a partnership between Google and publishers in Europe to advance "the practice of quality journalism," strengthen and evolve "publisher business models," and cultivate a "global news community."[13] The company now operates exclusively on earned revenue and is profitable.

Staff Size: 10

Launch Date: 2016

Future Plans: To develop its algorithm into a platform for citizen journalists. It plans to use its technology—which allows users to write about, interact with, and annotate news content in real-time—to combine individual consumer interactions with digital content and, ultimately, index "what the crowd is saying" about current events.

Truepic

Mounir Ibrahim, vice president of public affairs and impact: "COVID-19 accelerated a pre-existing trend of the digitization of everything we do. Who you're voting for, dating, what you're buying; everything starts with a video or a picture. Trust technology will need to play a role in a lot of verticals. Furthermore, government awareness…of trust technologies is growing [to monitor] digital content online. For more than consumer protection; for national security. Not trusting what you see and hear online is not only a fraud issue, but also a national security issue."

Overview: A photo and video authentication technology that allows businesses, non-profit organizations, nongovernmental organizations, and citizen journalists around the world to verify their photos and videos.

Funding: Raised more than $1 million in seed rounds between 2016 and 2017, $8 million in Series A in 2018, and $26 million in Series B in 2021, led by Microsoft's Venture Fund, Adobe, Sony Innovation Fund, Hearst Ventures, and individuals from Stone Point Capital.[14]

Staff Size: 53

Launch Date: 2015

Future Plans: In addition to its TruepicVision B2B vision platform,[15] Truepic is building a consumer-facing Software Development Kit, called TruepicLens.[16] Available for iOS and Android users, consumers will be able to integrate Truepic's highly specialized camera technology into existing applications to verify the authenticity of photos.

Vett News

Paul Glader, founder and chief executive officer: "Many citizens don't want help solving misinformation, and to pay for such technology to help them do so in a B2C product. So, we pivoted to a B2B product designed to help newsrooms strengthen their relationship and communication with the public. This kind of NewsTech product offers hope for a better future with quality information."

13 See Google News Initiative's website. Accessed on May 10, 2022.

14 Truepic, "Truepic Raises $26 Million Series B Financing Led by M12—Microsoft's Venture Fund to Scale World's Most Secure Camera Technology," Cision PR Newswire, September 14, 2021.

15 See TruepicVision on Truepic's website. Accessed on May 10, 2022.

16 See TruepicLens on Truepic's website. Accessed on May 10, 2022.

Overview: A workflow automation system that enables readers to click a button on any article and fill out a form to report typos, factual errors, issues of bias, or context. The reader gets an immediate "thank you" note and the editor gets an automated notification to handle the reader request in a dashboard that allows the editor to manage credibility issues quickly and personally.[17]

Funding: Received $75,000 in 2019 from the Knight Foundation, and $10,000 from NYC Media Lab.

Staff Size: 5

Launch Date: 2017

Future Plans: Hopes to expand its paying customer base and raise angel and corporate funding. Current pricing is $50 per month for independent press outlets, $500 per month for community papers, and $1000 per month for national publications.

Zignal Labs

Jennifer Granston, chief customer officer and head of insights: "We don't label content as 'true or false,' or 'harmful or not harmful.' Zignal's technology enables users to detect and mitigate narrative-borne threats and capitalize on narrative-borne opportunities—as they emerge in real time. For example, NLP and machine learning allow us to identify which accounts on Twitter behave as if they are using a high level of automation and sharing content in inauthentic ways."

"At Zignal, we are constantly reviewing media and social platforms to see the conversations that are happening and identify emerging narratives. We look at online sources like Twitter, Reddit, forums, and Sina Weibo, among others, as well as traditional media like print and broadcast. We see a lot on Twitter, 4chan, and Reddit, but the tactics and methods really vary based on how you look at the data. The data doesn't follow a pattern, so it's key to be able to look at the entire landscape and see narratives and channels as they pop up."

Overview: A Software-as-a-Service-based tool whose Narrative Intelligence Cloud analyzes billions of digital stories in real time to help customers discover and manage the narratives that can help or harm them. Originally conceived as a tool for political campaigns accustomed to media "war rooms", it is now used by the world's largest companies and public sector organizations to identify emerging risks and opportunities through Zignal's NLP and machine-learning algorithms, while providing insight into how to contend with the narratives that matter. Zignal serves customers around the world, including Expedia, Synchrony, Prudential, and The Public Good Projects.

Funding: Raised total of $74.9 million in funding over six rounds since 2012, with first-round funding in January 2012 from Mena Venture Investments. Series A funding of $4.2 million followed in January 2013. Among other investors are North Atlantic Capital and Blum Capital Partners, as well as individual investors Andy Ballard, chief executive officer of Wiser Solutions,[18] Jim Hornthal, chairman of M34 Capital,[19] and Mitchell Cohen of Trilogy Search Partners.[20] Zignal Labs' last round of debt financing totaled $20 million from Alignment Credit in January 2019.[21]

Staff Size: 100+

Launch Date: 2011

Future Plans: In June 2021, Zignal Labs introduced Emerging Narratives, an AI tool to help organizations understand narrative risk online.[22] "The important thing is putting technology into the hands of people who can do something positive with it," said Jennifer Granston. "And using data not just to look in the rear-view mirror, but to inform strategy."

17 Homepage, Vett News, accessed on May 10, 2022.

18 See Wiser Solutions' website. Accessed on May 10, 2022.

19 See M34 Capital's website. Accessed on May 10, 2022.

20 See Trilogy Search Partners' website. Accessed on May 10, 2022.

21 See Alignment Credit's website. Accessed on May 10, 2022.

22 See Zignal Emerging Narratives on Zignal Labs' website. Accessed on May 10, 2022.

Appendix B. Interviewees

Tania Ahuja, founder and CEO, Nobias, Google Meet interview by Aditi Rukhaiyar, February 16, 2022.
Shouvik Banerjee, founder and CEO, Averpoint, WhatsApp interview by Aditi Rukhaiyar and Juan Carlos Eyzaguirre, February 28, 2022.
Dan Brahmy, co-founder and CEO, Cyabra, Zoom interview by Juan Carlos Eyzaguirre, February 11, 2022.
Antony Cousins, CEO, FactMata, Zoom interview by Anya Schiffrin, February 10, 2022
Ben Decker, founder and CEO, Memetica, Zoom interview by Kristen Saldarini, February 4, 2022.
Danielle Deibler, co-founder and CEO, Marvelous AI, Zoom interview by Anya Schiffrin and Ojani Walthrust, January 28, 2022.
Camille Francois, global director of trust and safety, Niantic, Zoom interview by team, February 16, 2022.
Paul Glader, founder and CEO, Vett News, Zoom interview by Anya Schiffrin and Ojani Walthrust, February 4, 2022.
Jennifer Granston, Chief customer officer and head of insights, Zignal Labs, Zoom interview by Zachey Kliger, February 23, 2022.
Mounir Ibrahim, vice president of public affairs and impact, Truepic, Zoom interview by team, February 2, 2022.
Sagar Kaul, founder, Logically, Zoom interview by team, March 23, 2022.
John Kelly, CEO and **Guyte McCord**, COO, Graphika, Zoom interview by Aditi Rukhaiyar, March 3, 2022.
Rachael Levy, senior marketing manager and information operations lead and **Zohar Cohen**, vice president and head of delivery, ActiveFence, Zoom interview by Zachey Kliger, February 23, 2022.
Mark Little, founder and CEO, Kinzen, Zoom interview by team, March 2, 2022.
Karim Maassen, founder, nwzer, Zoom interview by Kristen Saldarini, February 7, 2022.
Arjun Moorthy, co-founder and CEO, The Factual, WhatsApp interview by Hiba Beg, January 31, 2022.
Jay Pinho, senior manager of product management in brand safety, Oracle, Zoom interview by team, February 23, 2022. (Pinho left the company in March 2022.)
Alejandro Romero, COO and Jonathan Nelson, digital intelligence specialist, Constella Intelligence, Telephone interview by Anya Schiffrin, February 10, 2022.
Matt Skibinski, general manager, NewsGuard, Zoom interview by team, February 9, 2022.
Naushad UzZaman, co-founder and CTO, Blackbird. AI, Zoom interview by Tianyu Mao, January 31, 2022.

As a non-partisan and independent research institution, The German Marshall Fund of the United States is committed to research integrity and transparency. This work represents solely the opinion of the author(s) and any opinion expressed herein should not be taken to represent an official position of the institution to which the author is affiliated.

About the Author(s)

Anya Schiffrin is the director of the technology and media specialization at Columbia University's School of International and Public Affairs and a senior lecturer in global media, innovation and human rights. She writes on topics related to journalism sustainability, impact, and online disinformation. Her most recent book is the edited collection Media Capture: How Money, Digital Platforms and Governments Control the News (Columbia University Press 2021).

Hiba Beg has worked for over five years as a multimedia journalist in India and is currently completing her MPA from Columbia University SIPA.

Juan Carlos Eyzaguirre is completing his MPA from Columbia University SIPA. He previously worked on educational policy for the Chilean Government.

Zachey Kliger is a program associate at the American Academy of Arts & Sciences and earned a MPA from Columbia University SIPA.

Tianyu Mao has an MPA from SIPA 2022. She specialized in Urban Policy with Quantitative Analysis.

Aditi Rukhaiyar is completing her Master of international affairs, at Columbia University SIPA and previously worked on advocacy and development communication in the public and private sector.

Kristen Saldarini has seven years of professional experience in strategic communication and is completing her MPA in social policy and management from Columbia University SIPA.

Ojani-Pierre Ruphin Walthrust received his MPA degree from Columbia SIPA in May 2022 and currently interns with the Department of State in the US embassy in Zagreb.

About GMF

The German Marshall Fund of the United States (GMF) is a non-partisan policy organization committed to the idea that the United States and Europe are stronger together. GMF champions the principles of democracy, human rights, and international cooperation, which have served as the bedrock of peace and prosperity since the end of the Second World War, but are under increasing strain. GMF works on issues critical to transatlantic interests in the 21st century, including the future of democracy, security and defense, geopolitics and the rise of China, and technology and innovation. By drawing on and fostering a community of people with diverse life experiences and political perspectives, GMF pursues its mission by driving the policy debate through cutting-edge analysis and convening, fortifying civil society, and cultivating the next generation of leaders on both sides of the Atlantic. Founded in 1972 through a gift from Germany as a tribute to the Marshall Plan, GMF is headquartered in Washington, DC, with offices in Berlin, Brussels, Ankara, Belgrade, Bucharest, Paris, and Warsaw.

Acknowledgments

The authors would like to thank Justin Hendrix for his support and comments on drafts, Noah Giansiracusa and Julia Angwin for reading drafts, and Michael Cowan and Carolyn Whelan for their edits. A team of students at Columbia University's School of International and Public Affairs researched this topic and their work was the foundation for this paper: Ryan Pan, Christina Cataldo, Anna Spitz, Mihir Mulloth, and Kasturi Girme. Francesca Edgerton also contributed research.

Cover photo credit: vs148 | Shutterstock