Entrepreneurial Solutions to Online Disinformation: Seeking Scale, Trying for Profit

Any Schifferin
In 2016, the Nasdaq Educational Foundation awarded the Columbia University School of International and Public Affairs (SIPA) a multi-year grant to support initiatives at the intersection of digital entrepreneurship and public policy. Over the past three years, SIPA has undertaken new research, introduced new pedagogy, launched student venture competitions, and convened policy forums that have engaged scholars across Columbia University as well as entrepreneurs and leaders from both the public and private sectors. New research has covered three broad areas: Cities & Innovation; Digital Innovation & Entrepreneurial Solutions; and Emerging Global Digital Policy. Specific topics have included global education technology; cryptocurrencies and the new technologies of money; the urban innovation environment, with a focus on New York City; government measures to support the digital economy in Brazil, Shenzhen, China, and India; and entrepreneurship focused on addressing misinformation.

With special thanks to the
Nasdaq Educational Foundation for its support of
SIPA's Entrepreneurship and Policy Initiative.
Entrepreneurial Solutions to Online Disinformation: Seeking Scale, Trying for Profit

By Anya Schiffrin

Executive summary

As worries about dis/misinformation online and the contamination of public discourse continue, and international regulatory standards have yet to be developed, a number of private sector companies have jumped into the breach to develop ways to address the problem. Most use a combination of people and natural language processing to try and identify false information online. Nascent and not yet widespread, the firms hope to survive by finding commercial applications for their products or by being adopted by the social media platforms. Most of the 13 startups we looked at fall into two categories: “bootstrap” operations with small amounts of funding, or slightly larger enterprises that have raised about one-two million dollars of capital. Whether any of these firms will be able to scale is not clear.

Introduction

Many solutions are being tried to combat the alarming spread of mis/disinformation online. To name just a few: governments around the world are imposing regulations as to what can and cannot be said online, foundations are funding fact-checking initiatives and efforts to build trust in media. Journalists are trying to build relationships with their audiences and strengthen local news reporting so as to provide societies with credible and useful information. Journalists also track the people and governments who create and spread false news and propaganda. Facebook, Google and Twitter claim to be tackling the problem of removing dangerous/illegal content but are universally resented because they profit from the spread of mis/disinformation and have been slow to respond to warnings about the dangerous effects it is having on society. Facebook has blocked tools promoting transparency of advertising on the site. Facebook has consistently refused to release data to researchers seeking to measure the impact of exposure to disinformation.

In this fragmented universe of solutions, characterized by a lack of comprehensive platform regulation, a number of small private sector companies have proposed ways of solving the problem. The aim of this report is to understand what solutions they are developing and outline some characteristics of the sector and points in common. We also tackle the question of whether their proposed solutions will be effective in the absence of widespread adoption by the giant platforms.

We carried out more than 20 in-depth interviews between December 2018 and February 2019 and this report includes short profiles of 13 firms. We identified the companies by asking experts in the field for suggestions which we then cross referenced as well as by reading press reports. Most interviews were conducted on the phone but some were done in person. We asked the companies about their technology and how it works, their business models, annual revenues, and their plans to scale as well as their thoughts about government regulation.

---

1. Anya Schiffrin is the director of the Technology, Media, and Communications specialization at Columbia University’s School of International and Public Affairs.
Disinformation is false information spread deliberately to deceive. The English word disinformation is a loan translation of the Russian dezinformatsiya, derived from the title of a KGB black propaganda department.  

Claire Wardle, executive director of First Draft, has further broken down the definition into types of dis/mis information citing a number of scholars who have also tried to do so. They have provided a taxonomy of the different kinds of mis/disinformation (including, for example, satire and false context as well as imposter content and fabricated content) and also made taxonomies of who the different actors and targets are.

Some general observations about the companies we studied:

- The companies more or less fall into three categories a) detection and mapping b) curation c) certification.
- There is disagreement as to whether natural language processing will be able to identify all forms of mis/disinformation online. Therefore, most of the outlets use a combination of people and natural language processing to try and identify false information online so it can be taken down.
- Nascent and not yet widespread, these companies hope to survive by finding commercial applications for their products or by being adopted by the social media platforms. Brandwatch and Alto have existing businesses and this new area or new application of their technology is not critical for their business continuity.
- Most of the 13 startups we looked at fall into two categories: bootstrap operations with small amounts of startup funding or slightly larger firms that have raised about one-two million US dollars of funding. Some, like Brandwatch and Alto, are not startups. Alto has more than 100 employees and Brandwatch was founded in 2006. Brandwatch has 64.7 million dollars in funding.
- Mostly the firms did not set out to crack the problem of dis/misinformation online. Many were involved in other activities and came across mis/information and decided to do something about it, notes Alejandro Romero of Alto Analytics, a data analysis firm based in Madrid with presence in UK, US and Brazil. Many have day jobs and core businesses. Many have ties to intelligence and government, cyber security, fraud detection. Some are tiny (Eric Feinberg from GIPEC is just one person) some are large and well established.

“The all the companies working in this space were working on something else and the disinformation they saw had an impact on what they were doing. They saw a business opportunity and thought this could be an interesting learning. I’ve not seen a company that started up just to fight disinformation,” Romero said.

The startup executives we interviewed assume that widespread regulation of online hate speech and the platforms is not imminent and that consumer and corporate demand for their products will continue. They note that Facebook and its affiliates, such as YouTube, and Twitter have no incentive to change a business model that is based on generating outrage and engagement. In this scenario, without regulation the platforms will continue to allow, or even encourage, the circulation of falsehoods online.

The platforms could do more to fix the problem but have no incentive to do so.

These firms exist because of a lack of action by social media companies that have no financial incentive to fix the problem themselves and because government regulation addressing the problem has not yet been passed. Or the regulation that does exist, i.e. section 230 of the Communications Decency Act, protects the social media platforms from being liable for what they put on their platforms. Many of the companies we spoke to said that Facebook, YouTube, and Twitter have been extremely lax and irresponsible in allowing hate speech and disinformation to contaminate their networks. Many believe that, technically, it’s not a hard problem to fix but they note that there is no incentive for Facebook or Twitter or YouTube to do so. “If the platforms try to tackle the problem internally it will be a huge revenue loss for them so
they don’t have an incentive to do it,” predicted Sagar Kaul, one of the founders of Metafact, a fact-checking platform based in India.

“The online disinformation is a lot like the spam problem. And it could probably be solved the way we solved the spam problem. We solved spam with a combination of legal and technical measures. First, Congress passed a law (the CAN-SPAM Act) that imposed fines on the senders of spam. Faced with liability, the email industry then set up a “black list” of known spammers that they all agreed to block.

Similarly, if there was some legal or financial cost to the platforms, they would likely set up a “black list” of disinformation outlets to block. But they currently have no incentive to do so,” says leading technology journalist Julia Angwin. Mark Zuckerberg has called for global regulation but observers note that his comments are belied by the amount of time and money that Facebook and other tech giants spend lobbying to avoid regulation.

“The platforms are extremely good at making soothing noises. They are silver tongued and very good at what they do. Their response is ‘We are hiring 20,000 people. We are all over this problem, we have a community that will flag false news.’ Facebook’s point of view is ‘we got this. We acknowledge the problem. We are in the 12 step program. But they haven’t admitted the first step: There is a power beyond ourselves. Facebook has never acknowledged this problem is bigger than them,” said John Batelle.”

Some firms try to work with advertisers to pressure the platforms.

In the absence of regulation or financial incentives, there is no reason for the platforms to remove mis/disinformation online. Therefore many of the startups hope that reputational risk and naming and shaming will prove effective. They are working with advertisers to see if the advertisers can push the platforms to take more action.

Assuming that big and respectable brands will not want to run their advertisements next to unsavory content, a number of the startup founders we spoke to expect that reputable corporate advertisers will push for change and force platforms to crack down on hate speech online as well as dis/misinformation. These platforms and some other groups, like the United For News coalition, hope to persuade advertisers to push the platforms to do a better job of screening ads.

The idea of pressuring advertisers to pressure the platforms is an attractive one. Since most of the platform revenue comes from advertising it seems that they would be most susceptible to pressure from advertisers. Journalists also hope to promote direct buying of advertising rather than programmatic so that quality media can benefit from advertising revenue rather than see it go to random websites.

It is worth noting that advertisers would themselves be damaged by government bans on micro targeting or the spread of regulation aimed at protecting privacy and so pushing the platforms to take action would protect both sectors (platforms and brand advertisers) from regulations they don’t want. Groups like United for News also hope to protect the quality of journalism.

Regarding United For News, publisher John Battelle says: “I am a huge fan of the philosophy. Direct media buying supports what I think is the most important part of publishing—the direct relationship between publisher, audience and marketer. Programmatic advertising has stripped away the context of an experience with the audience. Advertisements go into unknown places, out of context. Direct advertising will re-establish the connection with the audience. The problem is that programmatic advertising is cheap for advertisers and cheap for publishers to implement, so it’s irresistible—especially for the advertisers, who control the ecosystem.”

What is their business model?

While some of the firms in this field have been around for years, others are nascent and small. Many got help from friends and family as well as small startup grants and are now busy trying to develop and launch their tools or scale. Metafact was started with funding from Hanyang University in partnership with Seoul Metropolitan Government while others, like the Global Disinformation Index, relied on foundations. A few, like Brandwatch, have been around for years and have a core activity of monitoring the web for corporate clients. Yet some, like Vett News, consider fighting disinformation to be a part of their core activity. NEVA is doing facial recognition and has an ongoing business
that will fly on its own. NewsGuard is a product for journalists to do their job better and is seeking funding from the platforms.

Some like Truepic, hope for commercial applications for their technology. Truepic verifies photographs taken with its technology and assigns them a unique number stored in the cloud. This will not only help governments and human rights groups and media organizations that need to verify, say, pictures of atrocities or human rights violations but could also be extremely useful for insurance companies that are vetting claims and rooting out fraud.

There may be a way for some of these firms to generate revenue from the technology they have developed but only a few are likely to scale.

“There are probably a handful of scalable ones that use machine learning and natural language processing. Extracting information from video and images is difficult and something you can profit from. Companies that figure out how to do that will make some money and the rest of the firms will have a hard time scaling,” said Justin Hendrix from NY Media Lab.

Disagreement as to how much screening of dis/misinformation can be done without people

There is some disagreement among the founders as to how much of the screening of mis/disinformation can be automated and how much can’t. Can Artificial Intelligence (AI) and natural language processing identify all or most of the bad stuff? The people we interviewed were split on this question. Some argue that people are an essential part of the process because so many of the sites are designed to mislead and look like real news sites. As a result, it’s almost impossible for a computer to recognize all the different characteristics of these kinds of sites.
Can this problem be solved by technology?
An interview with NYU Professor Danny Rogers

Danny Rogers’ background is in intelligence and the dark web. His day job is working on combatting fraud and identity theft using dark web intelligence. Developing the Global Disinformation Index is a labor of love which he hopes will be used by advertisers who want to avoid placing their advertisements next to disinformation, as well as by platforms and other tech companies to help de-platform disinformation efforts.

“I don’t think automated natural language processing is scalable. Computers are not going to be able to distinguish content that is designed to fool humans.”

Rogers distinguishes two kinds of disinformation and says these need to be analyzed differently. At the top of the food chain are “highly organized threat actors like state-run operations or commercial ones like Cambridge Analytica.” At the bottom are decentralized purveyors such as trolls, 4chan, click bait purveyors, the legendary Macedonian teenagers.

One is high quality and comes from fully fledged media operations such as Breitbart or Russia Today. “It’s more pernicious because at least 60% of RT is high quality journalism and it looks different from the low quality sites put up quickly to get eyeballs,” Rogers said.

“High quality sites will be largely impossible to differentiate. A computer will have a really hard time computing the difference between Breitbart and CNN. Breitbart is very nuanced and you have to look at it from a journalistic perspective not computational. The junky misinformation has lots of hallmarks such as spelling errors, recycled material, it’s often presented on a Word Press template. All of this has signatures that you can identify.”

One unresolved question is how to define the problem so as to get bi-partisan consensus. “Using the word ‘disinformation’ makes Conservatives wary of a liberal plot to silence the media. But it’s not two equal sides: it’s irrationality versus enlightenment thinking,” Rogers said.

“Facebook’s job is to get people to click on links. They don’t want to combat this. They run affiliate marketing conferences teaching people how to do this. They got away with it for years until the 2016 elections. Facebook is siphoning off all the ad revenues of these clicks. Twitter has no incentive from a business model perspective to kick off the bots because their stock price is directly tied to their user count. The bots make them look bigger and more popular than they are.”

Others more, optimistically, believe that detecting and even suppressing false/disinformation can be automated nearly completely using AI and natural language processing but all admit that there is some uncertainty as to how this can be done as well as limits on how effective it will be.

They also note that tech solutions will take seven-ten years to implement and so aren’t any faster than waiting for comprehensive government regulations. “Private sector solutions are all seven to ten years down the road. I don’t see one coming up in one to two years,” said Joe Ellis from Vidrovr.

Will it even work?

Another problem is that much of the mis/disinformation is put on small fake sites that are new and change constantly so it may be impossible. Organizations like Mediabiasfactcheck.com can’t keep up with the new sites constantly appearing, says Romero. Further, these small sites inject their false information into Facebook and Twitter where it gets circulated further.

The human factor also comes into play as it’s hard to see how different tech solutions can be applied all over the world in places with different values. “How do you handle a global platform operating in different places
with different values? Even if the technology existed, the application in different contexts would be a nightmare,” says Romero.

According to Marie Frenay, a member of the office of the European Commission’s Vice-President and Commissioner for Digital Single Market Andrus Ansip, “There are very promising research projects and startups which explore the potential of AI to detect disinformation, identify patterns. We need to continue investing in this area. At the same time, it is also clear that human expertise is needed. I see technologies as tools that can assist disinformation experts in their work. It is about complementarity. As disinformation is conducted more and more subtly and covertly, making it harder to detect and attribute, we need the best of human brains and machines to address it.”

**Paradox of rating systems**

Others worry about whether the rating systems can be gamed, politicized or corrupted in the same way Moody’s was before the 1997 financial crisis. At that time, Moody’s business model came under fire because, as well as providing ratings, they also sold services to countries wanting to improve their ratings. They were also notorious for giving high ratings to countries that collapsed shortly after. The problems with the sovereign rating agency model can be applied to the rating agencies in the disinformation/news sector.

Julia Angwin said, “A better model would be peer accreditation, where journalists band together to enforce a set of standards on their industry and only include outlets that meet the criteria.” Danny Rogers is working on developing international standards. This idea of an industry standard is also being attempted by the Paris-based NGO Reporters Without Borders which is working with the European Standards Authority and an international coalition of journalists in order to come up with a list of credible media outlets that follow internationally agreed-upon standards. The list has already been criticized for including RT, the Russian television network.

Indeed, making such lists transparent opens up the rating entities to criticism. It is ironic that, given the desire for more transparency online, the entities that are rating news sites are not trusted by people if they know who is behind them.

An article by Will Oremus in *Slate* in January 2019, “Just Trust Us,” criticized NewsGuard for daring to make decisions as to what is trustworthy and what is not, pointing out the surprisingly positive rating for FoxNews.com and extremely negative rating for Al Jazeera and speculating that in the future, “the ratings authorities [could] become too powerful.”

“Making decisions about what misinformation to suppress or promote almost has to be done anonymously because if people know who is behind the effort they may not trust it. It’s strange but people almost seem to trust Facebook more than they would trust another group. The second you know who is behind the effort people will start arguing about whether the group is qualified to pass judgement,” Reg Chua, chief operating officer, editorial, of Thomson Reuters said in an interview. “It’s not that I am in favor of secret cabals deciding what we read; but more that all having a group vet information does is move the debate about who to trust upstream—from the news source to the vetters. Newsguard co-founder, Gordon Crovitz, points out that research on this question shows the opposite: Consumers are willing to trust journalists to rate other journalists so long as they operate in a transparent and fully disclosed manner and are not willing to trust Silicon Valley’s non-disclosed algorithms.”

**Fixing symptoms not causes**

One criticism of the companies trying to use AI to look at mis/disinformation is that they ignore its underlying causes and sources. For Alejandro Romero, “the social networks are the last building block,” in a process that begins with entities that find vulnerabilities in society and then target them, stoking fears about subjects such as immigration in order to affect the integrity of elections and political decision making. While the media is being targeted it is governments that are being affected.

**What do the startups think about regulation?**

In our interviews, we found mixed views as to whether government regulation would be a good idea. Some entrepreneurs such as Joe Ellis and Gordon Kro Ivitz from Newsguard believe in free market approaches. Mark Little from Nevalabs said, “I am afraid of regulation that doesn’t solve the problem but will make the perception of elite control of the media worse.”
Ellis also seemed wary of regulation, “The disinformation question is really hard. I don’t know how to solve it. The best way to try to solve it is to give as much power to actual consumer as possible so that search involves user intentions.”

Others are open to the idea of regulation but have not looked closely at the details.

Bringing back Section 230 would fix the problem straight away said Eric Feinberg and Julia Angwin. It would restore Intermediary liability, which was implemented in Germany at the start of 2019 and opposed by internet’s rights groups.

Other ideas include:

- removal of programmatic ads network
- legislation for more human fact checkers
- Twitter promotes ads that pay them. Instead it could promote fact-checked content
- Strong privacy regulations would help prevent exploitation, identity theft, micro targeting

“A ban on microtargeting would work. We strongly recommend European Commission to regulate Facebook, regulate algorithms so that they don’t micro target based on creating small information bubbles that fragment, atomize and divide our societies,” suggests Frantisek Vrabel from Semantic Visions in the Czech Republic.

As all noted, the threat of regulation will be an intrinsic part of getting the platforms to do more self-policing. “There is a carrot and stick approach. Big tech’s incentives are that advertisers are pushing that but the stick is even more powerful and governments are pushing for that,” said Oren Levi from Adverifai.

Conclusion

In this world of fragmentation, the small startups using AI and natural language processing are another niche to watch. It’s clear that the entrepreneurs we interviewed see a business opportunity in using people and natural language processing to identify and possibly remove mis/disinformation online. However, the limitations are all too apparent. The first is that AI is not subtle enough to identify all of the myriad forms of dis/information currently contaminating the information ecosystem. Second, even if it were possible to use the technology at scale, there is little evidence that Facebook, Google and Twitter would use it—one of many reasons why regulation of the platforms is essential. Third, these tech-based solutions don’t address the larger economic, social and political reasons that dis/misinformation spread.

“Platforms will not take on active defense of truth telling institutions [and] even if the platforms wanted to fix the problem, they can only have an important but limited impact in the disinformation landscape. They are a contributor to a massively organized disinformation system. But the digital ecosystem is broken and the possibilities of gaming the system are endless,” says Romero.
FAST FACTS

Brandwatch—monitors the web and sells the information to companies who care about their brand. Merged with Crimson Hexagon in 2018 and will continue to be called Brandwatch.

Background: “We work with the language of attribution rather defining what is true or false. We are not an encyclopedia or research institute. We can’t tell you what the GDP of Peru is but we can we can give you tools to determine where a story came from and where it’s likely to go. We give people a more structured away to understand sources and make intelligent decisions. That’s how it should be. The world shouldn’t abandon posture of neutrality towards information. I am not comfortable with saying what is true and what is not,” said Paul Siegel.

“Fear, uncertainty, doubt opens up pocket books better than consumer insights do.”

Brandwatch: “A true story looks the same as a false story. Both are a collection of sentences.” “There is a distinction between language and truth that is hard to make.” –Paul Siegel

What is it: UK-based company that monitors the web and sells clients data on the online public perception of their brand.

How funded: Privately held, VC funded as well as private clients.

Staff size: Over 550

Launch date: 2006

Future plans: Sees a market for selling services to government, government contractors and enterprises that are targeted by disinformation.

FAST FACTS

Factmata: “I think legislation will drive the platforms to do things and is needed urgently. It’s a shame that it has come to regulation and that they haven’t taken it on properly.” –Dhruv Ghulati

What is it: Software company that provides AI tools that detect specific types of disinformation. For businesses, they monitor their brand online to eliminate undesirable information about them.

How funded: Has raised $1.8m from seed funding.

Staff size: ~10

Launch date: 2017

Future plans: They are developing browser extensions and an app for individuals.

FAST FACTS

Global Disinformation Index: “Right now we have the most brand unsafe environment in the history of advertising. It’s the Wild West. Platforms have no incentive to actually secure themselves. We’re trying to catalyze grass roots support and get the advertising buyers to have a say.” –Danny Rogers

What is it: A UK non-profit trying to make an “AI-powered classifier which can identify junk domains automatically” and would then work with programmatic ad networks so that they have a “dynamic blacklist of sites thereby choking off funding for disinformation networks.”

“We want the Global Disinformation Index to be the ones to take on the risk. We have no skin in the game and can provide transparent, neutral ratings that platforms and the brand safety community can use.”
“The goal is to have a couple of products. One a self-updating blocklist of junky open web sites that are worth blocking in the ad exchange. This can be used by ad tech community to allow them to block ad buy on junky sites. No one company wants to take a stand or say ‘this is good or bad’. So we want to be neutral and transparent and be the risk-absorbing entity,” Rogers said.

**Funding:** $50,000 grant from Knight Foundation. Now has $1 million in seed money. Other funders include USAID and Luminate.

**Staff size:** Three co-founders.

**Launch date:** 2018

**Future plans:** Is a member of the Reporters Without Borders “technical advisory committee” and working with the European Standards Organization to get consensus-based standards developed for media outlets who opt-in. Eventually, this could lead to certification of outlets that meet certain standards of transparency and other criteria.

### FAST FACTS

**NewsGuard:**
- **What is it:** Ratings system that assigns red or green ratings called “nutrition labels” to thousands of news sites around the world. Can be used as a browser extension.
- **How funded:** Private investors including the founders. Charges the digital platforms to grant their users access to the nutrition labels instead of charging the actual publications being rated.
- **Staff size:** Stands at 28, currently expanding.
- **Launch date:** 2018

**Vett News:** “People would like a tool that when they are reading an article would allow them to understand if it’s fake or not, good or not, opinion or not, validation or confirmation” –Paul Glader

**What is it:** Like NewsGuard, they currently provide a Chrome extension that rates news sites based on their reliability; green for trustworthy and red for unreliable.

**How funded:** Boot strap, not yet ready to raise funding.

**Staff size:** About five.

**Launch date:** 2017

**Future plans:** Get a browser into schools and libraries and the ad tech market and be bought by Facebook or Twitter.

**Fast facts**

**Vidrovr:** Instead of passively receiving information, Ellis hopes that Vidrovr can help audiences as well as video producers become more active about what they see.

“You are at the behest of what the algorithm chooses to show you. Radical transparency will build trust so audiences can know who you are and what you have done. We think that will help combat misinformation and help companies monetize their video, allowing them to be transparent and provide intent into user search,” Joe Ellis who took leave from his Columbia PHD program to run Vidrovr.

“When you start a company, you have a vision for building it and that can get derailed by running out of money and going out of business or you get gobbled up by a large company that thinks what you are doing is really interesting. Google is very good at, building search solutions and use machine learning technology to make content available, and has a lot of talent. Can a startup do better than the big companies? usually not.”
What is it: Software that helps companies and news outlets index, annotate and search their videos.

How funded: Got tech start-up money aimed at student projects and then VC money, raised $1.25 million

Clients: Noting that they can’t disclose most of their customers, Ellis says Vidrovr “work[s] with some of the largest broadcasters in the US.”

Staff size: 7

Launch date: 2016

Future plans: Switch video viewing from social sites to an OTT (personal), more mobile based platform.

FAST FACTS

GIPEC: Eric Feinberg was in advertising technology. Has a patent on what he calls “anticipatory intelligence.” He is spitting mad after he began finding unsavory content online including ISIS posts calling for attacks on US troops.

How it works: his software looks for words like “caliphate”, beginning with hashtags and then trails it through the web. “My systems dig through all accounts using it.” Now Feinberg has a faux account and so the algorithms pushing pro-Jihadi content to him as well as to ISIS sympathizers.

Feinberg notes that he is “not going after the top, it’s the peer-to-peer, the sympathizers…You’ve got ISIS, radical jihadists getting radicalized in Indonesia, Bangladesh, Philippines….Facebook’s algorithm has connected the world for radical jihad.”

What is it: Monitors where clients’ ads are going online and generally helps its clients protect their brands online. Combines their research and reports with entities such as the Digital Citizens Alliance. (Digital Citizens Alliance is a nonprofit 501(c)(6) organization. Its IRS form doesn’t indicate its donors, simply stating that says DCA’s revenue comes from “program services.”

How funded: Bootstrap, looking for capital.

Staff size: Unknown

Launch date: 2015

Future plans: Hopes to be licensed, funded and to work with social media companies to reduce extremist content on their platforms

FAST FACTS

Metafact: “MetaFact is creating disinformation defense solutions for newsrooms, brands and organizations. By leveraging next-gen technology like advanced AI, and by analyzing pattern and bucketing data sets, Metafact helps newsrooms to understand if a certain discourse around a particular topic is genuine, or is a targeted campaign trail orchestrated to change public opinion or inflict financial damage.

Profiling bots that spread false claims is of paramount importance. Profiling human-run bot-like accounts is tougher, yet achievable with a claims-first approach. By being able to detect a claim as soon as it’s uploaded online, our tool is able to track the interaction of bot accounts and influencers long before any other tool is able to detect it as a threat. By using our claims-first approach we can proactively detect, monitor, and defend brands from disinformation attacks before they gain momentum and inflict financial losses.” –Sagar Kaul

What is it: A company that builds AI-based disinformation detection and defense solutions.

How funded: Bootstrap, friends and family. Obtained a grant for $20,000 from Hanyang University in South Korea. Kaul said “We also just completed an accelerator program in Ireland through the National Digital Research Centre (NDRC) sponsored by the Irish government and Enterprise Ireland. Metafact was the first startup selected from outside of Ireland and got 30,000 euros. We are working on IBM Watson platform since we were selected for IBM Global Entrepreneur program that provides IBM Cloud credits.”

Staff size: 7

Launch date: Prototype was launched in April 2018. MVP for the tool will be launched by the end of July 2019.

Future plans: Launching the tool for media houses, businesses and corporations. Developing an AI enabled media literacy app for kids: “Media literacy has to play an important part. We’re developing a mobile app that limits what kids can see on their phones and at the same time help them understand about echo chambering and how disinformation is spread. AI will be an integral part of the app. It will understand the needs of the kids and based on that the app can give recommendations for reading material.
FAST FACTS

AdVerif.ai: “Terrorism and violence are traditional[ly what companies are trying to eliminate]. Our focus is more fake news [that is] more challenging for technology to detect.” –Or Levi

What is it: Developing tools that use natural language processing to see if something is fake or suspect. Creating a blacklist for publishers and advertisers who want to protect their brand from being associated with fake content, and an API to help screen fake contact.

How funded: Bootstrap for now, hopes to be bought some day.

Staff size: One founder

Launch date: Company began in 2017, tool to launch soon.

FAST FACTS

Truepic: “We want to make sure anyone in the world with a smart phone has the ability to capture an image and prove its contents are real.” –Mounir Ibrahim, SIPA alum, who worked at USUN

What is it: Image verification technology. Truepic has several products. Users can download the free app and whenever they take a picture the system will log the time, date, location and pixilation and assign it an encrypted code that will be stored in the cloud. Truepic also has developed a remote inspections platform—known as Truepic Vision—for enterprise clients in Insurance, Banking and Lending.

How funded: Not profitable but generating revenue, raised $8m in 2018

Staff size: 30

Launch date: 2014

Future plans: Have commercial applications

FAST FACTS

Alto Data Analytics: “Alto was not created to research disinformation. We just found it.” –Alejandro Romero

What is it: The company provides software that monitors, retrieves and presents data that allow businesses to understand customer insights and offers help from a team of experts.

How funded: Sales

Staff size: Over 100 people.

Launch date: 2012

FAST FACTS

Semantic Visions: “We don’t focus on analysis of online social networks but we focus on online news. In our experience the disinformation and propaganda start on news sites and blogs” –Frantisek Vrabel

What is it: A large, speedy database and a web mining system that are used for risk assessment and monitoring. Has roots in the defense industry and uses open source intelligence.

How funded: Work for corporate clients (real-time risk detection solution integrated into SAP Ariba business commerce platform risk) pays for the work on disinformation. No government funding or funding from the platforms. Recently won a $250,000 grant from the US Department of State to help fund the development of cutting-edge new technology to combat disinformation online. The grant is going to come from the US Department of State’s Global Engagement Center.

Staff size: 25

Launch date: 2005
Acknowledgements

Supported by a grant from the Nasdaq Educational Foundation on behalf of the School of International & Public Affairs (SIPA) Entrepreneurship & Policy Initiative

About the SIPA Entrepreneurship & Policy Initiative

Columbia University’s School of International and Public Affairs (SIPA) Entrepreneurship & Policy Initiative engages scholars, entrepreneurships, and leaders from the public and private sectors to advance understanding of how to promote innovation, entrepreneurship, and social entrepreneurship. In 2016, the Nasdaq Educational Foundation awarded SIPA a multi-year grant to support initiatives in entrepreneurship and public policy. Since 2016, the Entrepreneurship & Policy Initiative has been a thought leader on topics including the Internet of Things (IoT), global education technology, cryptocurrencies and the new technologies of money, and the urban innovation environment.

This paper was made possible with support from the Nasdaq Educational Foundation. Thanks to John Batelle, Justin Hendrix and Raju Narisetti for introductions to the firms we profiled. I am also grateful for the research provided by Chloe Oldham and edits suggested by Hollie Russon Gilman. Above all, thanks to our interviewees for their time and to Dean Merit Janow and the Committee that allocated the research funding.
End Notes


2. Controlling the Assault of Non-Solicited Pornography and Marketing, signed by President Bush in 2003.


