

THE MANY FACES FIGHTING DISINFORMATION



SAFEGUARDING
CIVIL SOCIETY'S
ROLE IN THE
RESPONSE TO
INFORMATION
DISORDERS

Funded by:



**FRIEDRICH NAUMANN
FOUNDATION** For Freedom.

European Dialogue

TABLE OF

CONTENTS

5

EDITOR'S NOTE

6

[VerificaRTVE](#)
META-DEBUNKING
DECENTRALISED
DISINFORMATION

8

[Debunk EU](#)
DISINFORMATION IS SMALL
WATER DROPS THAT OVER
TIME CAN HEW OUT A STONE

10

[Maldita.es](#)
JOURNALISM NOT TO BE
FOOLED

12

[Who Targets Me](#)
"WE USE TECHNOLOGY
TO TELL THE STORY OF
TECHNOLOGY"

14

[MEMO 98](#)
MONITORING ELECTIONS IN
THE DIGITAL AGE

16

[Citizen D](#)
COUNTERING
DISINFORMATION FROM A
TO Z

17

[WTFake!?](#)
THE FIRST CITIZENS' ANTI-
FAKE NEWS BRIGADE

18

[Bellingcat](#)
A CENTRAL NODE IN A
GROWING NETWORK

20

[Global Disinformation Index](#)
"OUR GOAL IS TO DEFUND
DISINFORMATION"

21

[Tracking Exposed](#)
KNOW YOUR ALGORITHM

22

[Lie Detectors](#)
NEWS LITERACY FOR ALL

24

[AlgoTransparency](#)
"ALGORITHMS NOT
DESIGNED TO HAVE QUALITY
INFORMATION IN THEIR
OBJECTIVE FUNCTIONS
WILL NATURALLY FAVOUR
DISINFORMATION"

26

[First Draft](#)
ADDRESSING THE
INFORMATION ECOSYSTEM
DISORDER

29

[Account Analysis](#)
"MACHINE LEARNING OFTEN
GETS IT WRONG, I WANTED
TO GO A STEP BACKWARDS"

30

2020, A TURNING POINT
IN OUR RESPONSE TO
DISINFORMATION?



KEY TRENDS



These are micro-entities. 43% of the initiatives surveyed count between 0 and 2 employees



They are community-reliant. 57% work with volunteers and 64% rely on crowdsourcing to some extent



Strained relationship with the private sector. Almost a third of the participants feel they are in opposition to the major platforms, and two thirds have a weak relationship with telecommunications actors

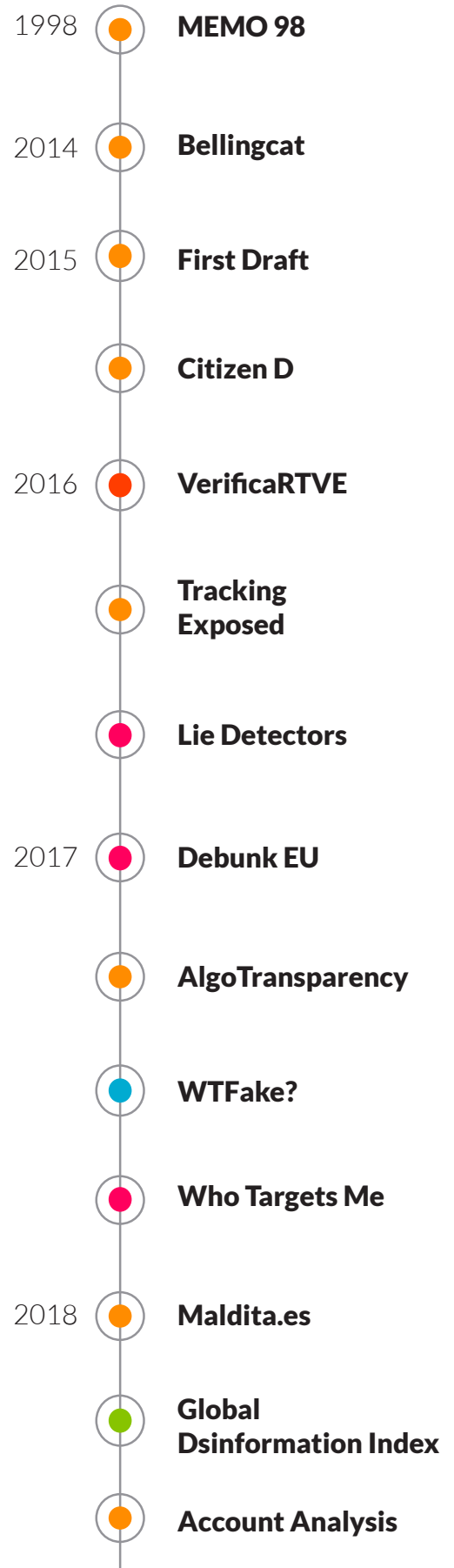


Sustainability. Only one of the initiatives surveyed said that their operation is fully sustainable.



Security. Cybersecurity is a concern for all actors interviewed. None feels that their operation is entirely secure.

TIMELINE





As we have noted at EU DisinfoLab, disinformation has many faces (manifestations, motives, and tools). It is only logical that the response to disinformation must have many faces as well. In this project, we seek to present a panorama of the different kinds of actors responding to disinformation today - from broadcast journalists to open source investigators to election observers to technology developers. In the report that follows, we interview 14 actors from across this emerging civil society ecosystem.

It is important to note that many of the actors interviewed do not see themselves as responding to disinformation per se. We use the term 'disinformation' as a shorthand for the many illnesses in our information ecosystem (what First Draft, whom we interview here, refers to as "information disorders"). The problem for us encompasses disinformation and misinformation, mistrust in journalism and the weaknesses or deficits in existing media, the business model behind clickbait and disinforming content, algorithmic targeting, and other opaque mechanisms in our digital information architecture.

This report is necessarily limited in scope. The interviews here represent a snapshot or cross-section of an expanding network of individuals and initiatives. Still, the findings we share are broadly representative of this space. We conducted qualitative, semi-structured interviews in order to understand how these initiatives were created and have evolved over time, and to examine in detail the difficulties these actors face in terms of sustainability, security, and impact. At the same time, these actors serve as examples of what we hope to see

more of. They exhibit new types of expertise - from digital forensics, to crowdsourced research, to journalism-classroom partnerships. Though they might seem unusual today in their techniques or organisational structures, we believe that they also point towards the future of civil society engagement against disinformation.

Despite the critical work they do, many of these actors feel alone. They struggle to make their voices heard by policymakers and by the tech companies at the heart of our information ecosystem. Their long-term security and sustainability are not assured. In order to effectively counter disinformation challenges, we need an ambitious framework to sustain and further develop this civil society network. Following these interviews, we make several recommendations which we feel are necessary to safeguard a resilient, decentralized civil society ecosystem.



DESPITE THE CRITICAL WORK THEY DO, MANY OF THESE ACTORS FEEL ALONE. THEY STRUGGLE TO MAKE THEIR VOICES HEARD BY POLICYMAKERS AND BY THE TECH COMPANIES AT THE HEART OF OUR INFORMATION ECOSYSTEM.

Last, it should be noted that this project is a view from the inside. These discussions were framed by our own experience at EU DisinfoLab, as a relatively young, small NGO, finding our way in this new environment. We consider the people featured here to be our colleagues, allies, and friends. We admire them, we learn from them, and we are continually grateful for their work and their collaboration.

META-DEBUNKING

DECENTRALISED DISINFORMATION



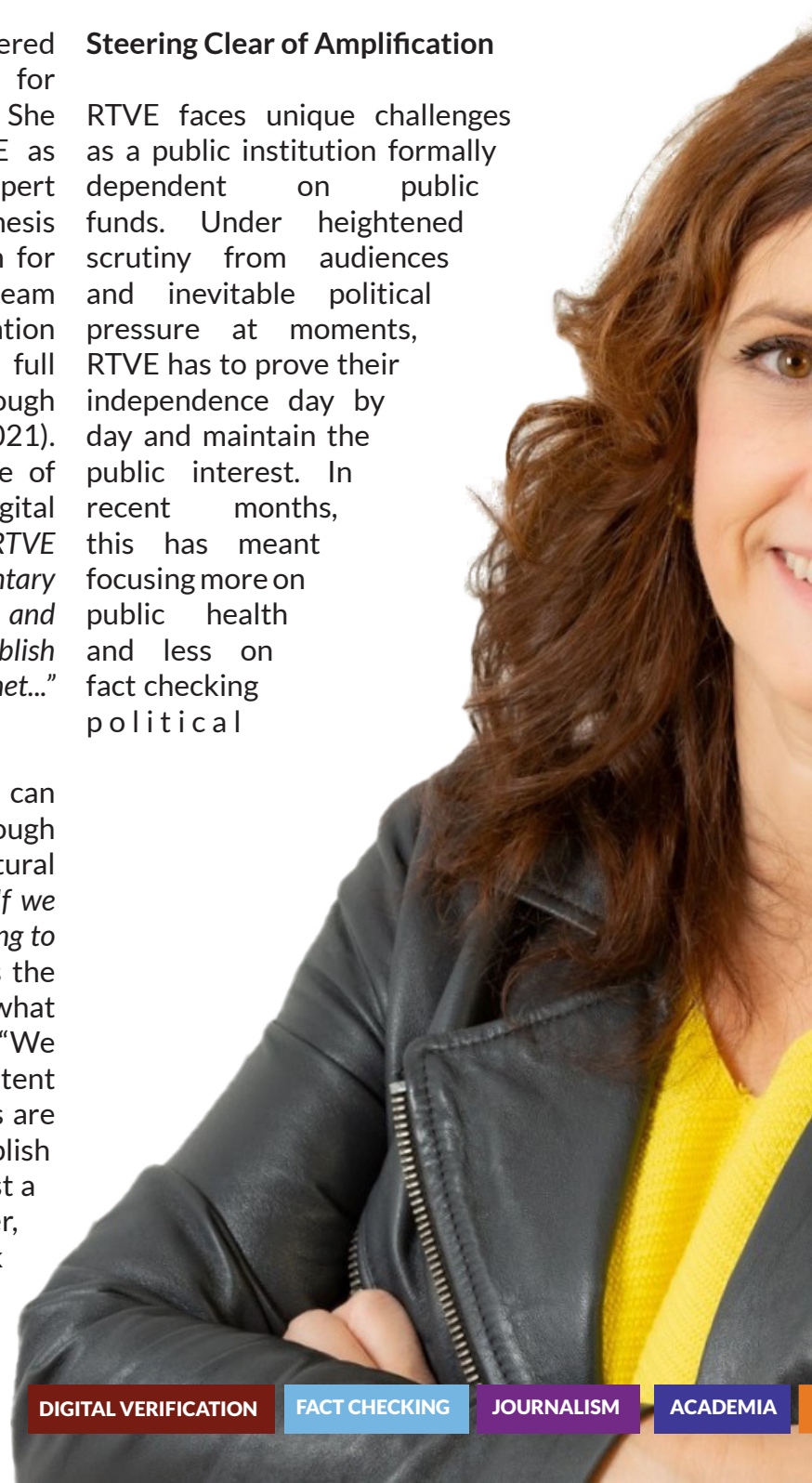
The purpose of Verifica RTVE is to change the culture of the institution.

Dr. Myriam Redondo has pioneered digital verification workshops for Spanish journalists since 2012. She came into contact with RTVE as an external trainer in 2016. An early expert in the field, she released her doctoral thesis on “Internet as a source of information for international journalism” in 2006. The team she is currently part of, the digital verification team Verifica RTVE, includes only 2 full time and two half time employees (though they will be adding new members in 2021). Their purpose is to change the culture of the institution as a whole by teaching digital verification. *“This is transversal across RTVE and that is the success. It involves documentary experts and archive experts, designers and journalists, all positions in the team. We publish in multiple formats, radio television, internet...”* she explains.

For Myriam, the problem of false news can only be addressed transversally, through journalistic capacity building and a cultural shift. *“We as journalists cannot do it all. If we fight fake to fake, we’ll get tired. It’s like trying to swim in a vast sea.”* Instead she suggests the need to meta-debunk, particularly for what she calls distributed disinformation: *“We tend to analyse a fake, we take the content and debunk it, but, at least in Spain, liars are sophisticated in their activity. They publish content but it doesn’t include a lie, it’s just a suggestion, then a second liar goes farther, and a third one farther. You have to debunk the whole chain, the idea behind it”.*

Steering Clear of Amplification

RTVE faces unique challenges as a public institution formally dependent on public funds. Under heightened scrutiny from audiences and inevitable political pressure at moments, RTVE has to prove their independence day by day and maintain the public interest. In recent months, this has meant focusing more on public health and less on fact checking political





statements, which Myriam perceives very often to be “noise”. *“Politicians from extremist parties are tempted to use our services to amplify a topic. When we verify a topic we enter their agenda.”*

The team is also strategic in their method of debunking to avoid sharing content more widely than is necessary; they try to respond to queries from citizens in the same channels where they are posed (directly in a WhatsApp message, for instance) and they try to

reply to personal questions privately. Of course, this takes massive human resources. Not only is this kind of monitoring time consuming, but it is often not possible in closed messaging spaces. *“We cannot clearly*

see what is happening. We are partially watching what happens in each but don’t have the whole vision”. Myriam explains the need for more tools, in particular tools that provide network analysis and track trends across platforms. She also needs the ability to parse more carefully between countries, to avoid unnecessary transnational amplification through fact checking.

Myriam wants more collaboration between journalists and specialized institutions. She also sees a need for guaranteeing diversity in the growing industry of digital verification; political fact checking is a clear example of an area where a multiplicity of voices is needed, rather than a monopoly. *“I envision a world in which journalists are doing our job, but in which we need [digital verification] organisations for deeper analysis on a given trend of topic. Also, a world in which citizens receive a more robust education on media literacy and critical thought,”* she concludes.



IT’S IMPORTANT TO DEBUNK FAKE BY FAKE BUT ALSO TO META-DEBUNK TRENDS AND INTENTIONS

ABOUT RTVE



RTVE is the Spanish Radio and Television Corporation, founded in 1973. It is also the first Spanish national media that began training its staff on UGC (user generated content) and digital verification techniques, at a time when political fact checking was the dominant trend in the country. This was before the election of Donald Trump, “the event that changed everything”.



DISINFORMATION IS
SMALL WATER DROPS
that over time can hew out
A STONE

- DISINFORMATION ANALYSIS
- MONITORING & REPORTING
- TECHNOLOGY DEVELOPMENT
- STRATEGIC COMMUNICATION
- LARGE-SCALE MEDIA LITERACY CAMPAIGNS
- FACT CHECKING
- COMMUNITY TRAINING



 **DebunkEU.org**
Disinformation analysis center

 **Viktoras Daukšas**

 **Lithuania**

 **Lithuania, Latvia,
Estonia, Poland, US,
North Macedonia**

Debunk EU defines itself as an independent technological think tank and analysis center for disinformation.

For Debunk EU, building a response to match the scale of the disinformation problem has meant merging automation and artificial intelligence with human analysis and dedicated volunteers (a network of ‘elves’). “How do you find the needle in the haystack, the needle being disinformation cases that have the biggest impact? If you do it manually, you’ll never see the bigger picture”, explains Viktoras. Debunk EU has developed an AI-based analytics tool which spots and identifies topics of interest in online articles in real time. This means that from 1 million pieces of content they receive each month, their analysts can focus on the most harmful ones (10 – 15 000 content pieces). Long term reporting and analysis are also at the core of their approach: disinformation analysts in the four countries provide thematic reports on topics and trends, which are then shared with a wide range of stakeholders. Debunk EU has applied process automation across their reporting and analysis activities, allowing them to produce around 10 reports per month.

A scalable approach

Founded in 2017, the project emerged in the Baltic countries and was supported by DELFI and the Google News Initiative. Debunk EU’s initial partnership with colleagues in Latvia and Estonia was inspired by their shared history and disinformation threat constantly coming out of Russia since the fall of the Soviet Union. In order to meet this challenge, they

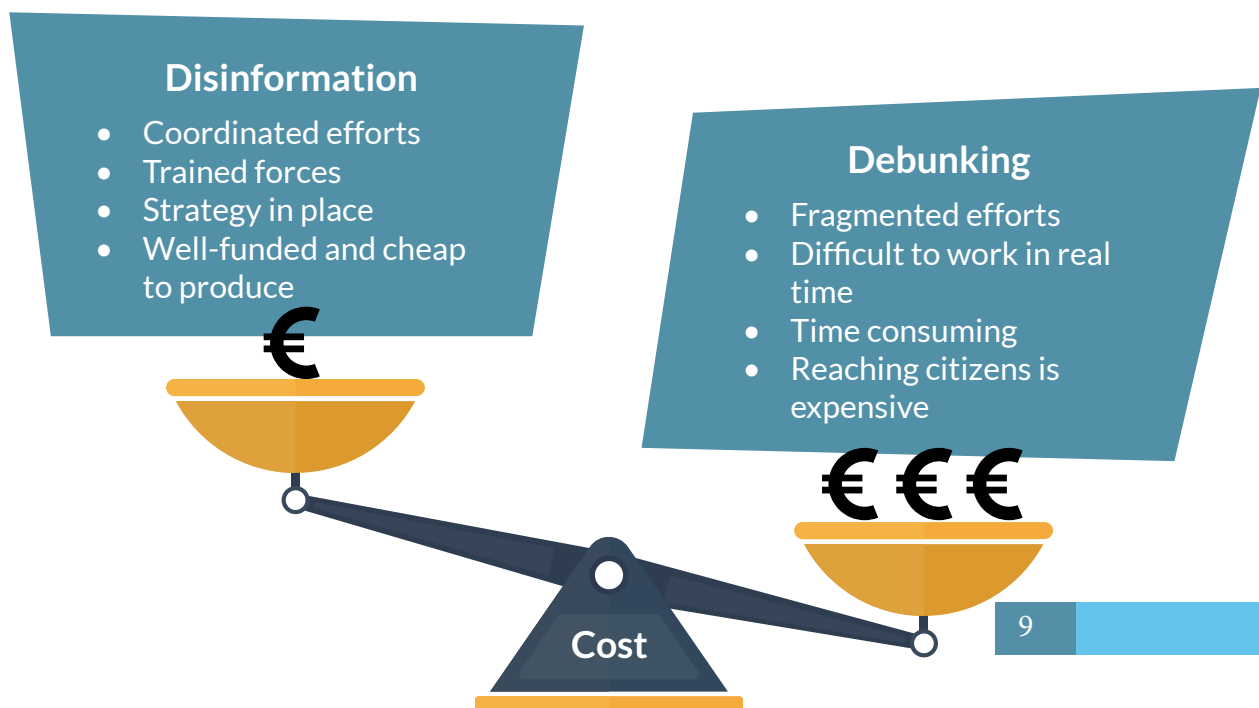
have drawn on their expertise in research and analysis, accumulated over many years. Currently they see potential for working with Eastern partnership countries as well. “The technology is scalable, and we can work with local partners”, say Viktoras. The fact that the Debunk EU team in Poland was able to produce their first report within four weeks of its establishment in Warsaw proves that the system is not only scalable, but also able to achieve significant results in a short time.

Developing automated solutions is costly and requires in-house technical expertise, which is rare in the disinformation space, and generally unheard of among NGOs of this size. Debunk EU is growing rapidly, and now sustainability is

the main question. As a tech-based organisation whose infrastructure is regularly attacked, they must invest heavily in cybersecurity and monitoring of their digital ecosystem. Even the most agile project management requires fuel to run on. While they haven’t found the optimal financial model yet, Viktoras says that they will be testing out new models in 2021.



IF YOU ARE A BUSINESS, YOU CAN BE FUNDED BY ANYONE, BUT IF YOU ARE FIGHTING DISINFORMATION, YOU CAN WORK ONLY WITH CREDIBLE AND TRANSPARENT DONORS



NOT TO BE FOOLED

Maldita is a Spanish non-profit fact-checking and data journalism platform working to monitor and counter digital discourse, to fight disinformation, and to promote media literacy.

Two Spanish broadcast journalists, Clara Jiménez Cruz and Julio Montes, began the organisation in 2014. It has since grown to 20 full time employees and between 30 and 35 operating staff. The team primarily debunks stories brought to them by their volunteer community and recirculates these debunks as widely as possible. Their approach is deeply community centered, driven by a tiered volunteer participation model: as of July 2020, this counted about 40,000 General Malditos or recipients of the newsletter, 1,000 Ambassadors or financial supporters, and 2,000 active Malditos and Malditas, individuals who contribute 'superpowers' or subject area expertise to the organisation's fact checking efforts.

Since their founding and particularly during the pandemic, the mis and disinformation problem has changed in scale and scope. *"We've gone from 1.5 million unique visitors per month to sometimes 8.5 during the pandemic, which is good and bad"* Carlos testifies. *"We have been hiring during the pandemic, expanding our operation because the situation called for it. The audience was more engaged, but small organisations like ours are fragile. The scale of the operation was sometimes overwhelmed."*

Maldita has been experimenting with artificial intelligence (AI) in the form of a

chatbot, which currently has an answering rate of 8 hours. Carlos is confident that the AI will improve further as the community diversifies. Growing and diversifying the community is top of the agenda, bringing in what he describes as *"the many people who care about disinformation but who aren't following every second of the news cycle."* Maldita hopes to reach these wider audiences through new partnerships, for example with the popular tabloid *20 minutos*. Meanwhile they're exploring other verticals, for example through a science public engagement project, a platform focused on gender-related hoaxes, and a browser extension to promote transparency in public and private institutions. Recently, they are moving more into the public policy space to weigh in on major policy issues too big to ignore.

Maldita brings clear benefits to many social media platforms. In a sense, Carlos explains, they "pay the bill" for fact checking services that platforms claim to provide.

Though Carlos specifies that different platforms have different practices, in general, these relationships are far from reciprocal. At the same time, platforms are not sufficiently forthcoming with the data that is necessary for actors like Maldita to assess the effectiveness of their debunking. Still, "the task at hand is too important to

forgo” in Carlos’ words, and the community is in agreement. Maldita has just received foundation status in Spain, which allows them to preserve their reputation and independence - a long and expensive endeavor that was made possible through crowdfunding contributions.

M MALDITA.ES
PROCESANDO PARA QUE NO TE LA CUALCAN

 Carlos Hernández

 Spain

 Spain, North and South America



WE NEED TO FIND WAYS TO MAKE SURE WE AREN'T WORKING FOR FREE. WE HAVE TO ENSURE OUR FUTURE

WHY “MALDITA”?

Maldita translates to ‘damned’. The name is a reference to Maldita’s first fact checking initiative, Maldita Hemeroteca (Damned Archive), which is a project to confront politicians with past statements they themselves made that contradict their current views.



644 22 93 19

SERVICIO AUTOMÁTICO DE WHATSAPP

"WE USE TECHNOLOGY TO TELL THE STORY OF TECHNOLOGY"



CONSUMER LITERACY

RESEARCH

TECHNOLOGY DEVELOPMENT

POLICY/EXPERT ADVISORY ROLE

ADVOCACY

Sam Jeffers co-founded Who Targets Me in 2017, having worked previously on political campaigns which had favoured grassroots, bottom-up tactics. He'd witnessed the ecosystem shift around 2015, when parties and candidates began to buy social media ads to target small groups of voters with large quantities of tailored messaging.

Who Targets Me is interested in "practical transparency for political campaigning", in Sam's words, and most focused on transparency for people running for office.

"Others might be looking for other things such as fake profiles or state influence efforts. We're looking at the very top of the tree because we think that sets the norms and examples

around which the rest of politics and democracy work."

Who Targets Me is working in an uncrowded space, in part due to the technical challenge, and in part due to an increasingly chilly research climate around the large social media

platforms. They are a small operation and all of their funding is project based. Sam and his co-founder, who has another fulltime job, take on capacity and particular skills as needed, which gives them a certain lightness compared to traditional NGOs. *"We're a bunch of flexible people who have interesting and innovative ideas, which we try to execute cheaply and quickly and simply."* Despite their size, they are committed to staying *"at the forefront of thinking on political ads and regulation"*. They bring a more balanced voice to the policy space, where new entrants are often eager to simply *"see everything banned"*.

A lack of creativity in the policy response implies a general need to better bridge the product-policy divide, Sam thinks. *"Two years ago we could have been experimenting with labeling*



WHILE WE DO COMPLAIN ABOUT THE PLATFORMS (A LOT), SOMETIMES THEY'RE THE ONLY ONES DOING ANYTHING



in an independent way and looking through a more imaginative portfolio of responses. Alternatives are possible, but no-one is really doing the work of designing what better services might actually look like.” He also notes the need for longer term thinking. “We need to think about election years and cycles rather than the few months and weeks before elections. People’s opinions are

shaped over a longer period of time. We should be looking for longer and using the aftermath of elections to push forward with reforms.”

CONSUMER AD LITERACY

Who Targets Me provides a free browser extension to help people understand more about the paid media they are exposed to on Facebook. The software shows a library of all political ads sent to them and data on who is showing them the most ads, and also helps them understand the mechanics behind the tailoring of that content. Sam describes it as a consumer ad literacy tool. “We’re trying to exemplify the transparency that we want the platforms to provide.”



“WE WANT TO UNDERSTAND WHAT THE BIG ACTORS ARE DOING, WHO’S FUNDING THEM, AND HOW THAT HAS AN IMPACT ON DEMOCRACY AND ON PEOPLES’ UNDERSTANDING OF ISSUES.”



MONITORING ELECTIONS IN THE **DIGITAL AGE**

MEMO 98 is a monitoring organisation consisting of media and election experts. What began in 1998 as a project to monitor the Slovak media prior to the parliamentary elections developed into a permanent organisation that has conducted media and election monitoring across the world.

MEDIA MONITORING

ELECTION OBSERVATION

SOCIAL MEDIA MONITORING

JOURNALISM

CAPACITY BUILDING / TRAINING

POLICY / EXPERT ADVISORY ROLE



ELECTIONS TOOLKIT

In November, Rast'ó published a toolkit "How to monitor media coverage of elections". The toolkit was developed within the context of the Council of Europe project "Supporting the transparency, inclusiveness, and integrity of electoral practice in Ukraine", implemented within the framework of the Council of Europe Action Plan for Ukraine 2018-2022.

Currently MEMO has a core team of 7 full time employees, but they work closely with local partners in a number of countries. MEMO's methodology of media monitoring focuses on content



**WE AIM TO HIGHLIGHT
THE WORK OF LOCAL
PARTNERS, SERVING
AS A SORT OF QUALITY
GUARANTOR**

and aims above all to evaluate political and social diversity in media reporting. In recent years, they've adapted their methodology from traditional media to account for the principles of the social web, but they maintain the focus on

both spheres. *“This combination of traditional and social media monitoring is very important,”* Rast’o explains.

Their monitoring is designed to provide in-depth feedback on pluralism and diversity in media reporting, including coverage of particular themes (integration of minorities, corruption etc.). MEMO does not only focus on disinformation. They have *“a more holistic and general approach, assessing both the positive and negative impacts of social media platforms on election integrity”* in Rast’o’s words. They study three things: the actors (from both traditional media and social media), the messages and narratives (how they are used by parties to make claims and to polarize) and the messaging (how the message is amplified).

Much of MEMO’s work includes capacity building and training, enhancing the media monitoring activities of local partners. They also train journalists, NGOs, regulators, and other members of the media. For Rast’o, it is critical to support these local actors who will

be the ones left after international election observers leave. *“It doesn’t end at the end of an election cycle. It is usually the start of another one.”*

Where MEMO struggles, along with others in this space, is in fully understanding the mechanics of amplification online. *“We focus less on inauthentic behavior - bots and trolls who amplify content. But from a different perspective, this is a critical part. This can make marginal voices more visible.”* The tools currently available for social media analysis (for instance, Facebook’s CrowdTangle) have been a game changer for MEMO, but the data is incomplete without access to private pages and closed messaging spaces. While entering these kinds of spaces raises ethical questions for researchers, there are legitimate design and transparency issues that hinder election integrity monitoring.



PLATFORMS SHOULD BE GEOGRAPHICALLY BLIND WHEN IT COMES TO APPLYING THEIR RULES. WE SHOULD NOT ASSUME THAT EVERY ELECTION IS LIKE THE ONE IN THE US



CITIZEN D'S ONGOING CAMPAIGNS

Citizen D is currently tracking two governmental ad campaigns which are being funded with taxpayer money. The web and television ads are directly related to government parties through various non-governmental organisations with strong government ties. Citizen D is relying mainly on FOIA requests to gain insight into the government's opaque process of cost tracking and selection of the ad placements.

INVESTIGATIVE REPORTING

MEDIA LITERACY

CAMPAIGNING



THE TERM 'FAKE NEWS' HAS IN OUR OPINION MORPHED INTO A CATCH-ALL PHRASE THAT DOES MORE HARM THAN GOOD

COUNTERING MISINFORMATION FROM A TO Z

Based in Ljubljana, Slovenia, Citizen D is a nonprofit whose core mission is the promotion of human and digital rights.

Despite counting only two full time employees, Citizen D's activities are various and in-depth. Their actions range from privacy rights monitoring and training, to investigating practices within the advertising industry and holding discussions on the purpose of mass media in a democratic society. They offer media literacy and digital privacy trainings, and they raise awareness through campaigns. One example is a national anti-hate campaign drawing attention to the Slovenian public funding of Hungarian propaganda outlets.

While Citizen D aims to encourage active citizenship and democratic participation, they take legal action on their own. *"This myth around an active citizen that will react and gather sufficient information from the media to make their own case, that's not the reality we're seeing,"* Domen explains. To help lift the onus from citizens, they've adopted an approach of working from A to Z. *"Our job doesn't end when we file a report. From that we analyze the problems, we define which decision makers are responsible, and we pursue that."* Frustrated by the limited functioning of the legislation in some areas, for example around false advertising, they have also taken to proposing legislative

recommendations.

In the Slovenian political context - which Domen has written about - Citizen D faces a double battle. In addition to the deeper problem of political propaganda funded by public money, they have to respond to a prevailing narrative around so-called 'fake news'. They have developed a targeted approach, following the money between all relevant public services and actors (advertisers, public funds, political parties, etc.). This posture can make financing complicated. Citizen D tries to maintain a dispersed model of funding, developing alternative sources of revenue that include commercial projects - in line with their mission and under a code of conduct. Meanwhile, the lack of outside, international media pressure on Slovenia means that local issues are often ignored. *"I'd have an easier time saying there was a problem with Russian propaganda. Nobody is focused on Orbán and Vishegrad,"* says Domen.



THE FIRST CITIZENS'

ANTI-FAKE NEWS BRIGADE

“You could say the problem I’m trying to solve is fake news, but, deeper than that, the problem is mistrust of journalists.”

JOURNALISM

ACTIVISM

Aude is a journalist by training. She is also president of Fake Off, a media literacy association that works with young people in classrooms, which she founded together with a small group of journalists in the wake of the 2015 Paris attacks. Behind the crisis of media literacy and ‘fake news’, Aude identifies a lack of dialogue and a growing mistrust between journalists and people. Following the election of Donald Trump, she began to feel that it wasn’t sufficient to intervene in classrooms, and decided to take directly to YouTube. *“The motivation behind everything is anger, I see the impunity, the indecency of manipulating the public, it makes me so angry.”*

La redac’ WTFake (‘the citizen journal’) tracks and debunks conspiracy theories, primarily found on YouTube, with the support and admiration of an online following. In a way, Aude’s activities mirror those of her enemies - like the conspiracy theorist Jean-Jacques Crevecoeur - in a multimedia chase that plays out across the social web. The investigations progress over several days on a Discord chat, and culminate in a revelation streamed live over Twitch. A handful of dedicated followers

have become discussion moderators. Others contribute in different ways, like developing the logo. Her colleague Sylvain Louvet helps with editorial and video production. Aude now has 1411 people on her Discord, and the feedback from many of her followers suggests she is achieving the impact she hoped for, regularly receiving comments like *“usually I don’t like journalists, but I like you!”*

“C’est le bazar”

Organisationally, things are messy, Aude admits. She set out on this alone, without a financial model, and without the kinds of contacts that many need in this space to survive. A few crowdfunding links have not yielded much yet. *“I’m just a little investigative journalist, but now I find myself having to reflect like a business person”* she explains. While the energy of her followers motivates her to continue, animating the community also limits her ability to strategize and grow. To devote time to fundraising would mean abandoning the community, and both activities take away from time spent on investigations. Meanwhile, there are trolls to be wary of. Aude has already changed her phone number after a conspiracy theorist doxxed her online. *“Given the people I want to investigate, I’ll probably see more of this,”* she reflects.





A CENTRAL NODE IN A GROWING NETWORK

Bellingcat emerged in July of 2014 from the online community that had formed around Eliot's blog, in particular around his work tracking the downing of flight MH17.

OSINT / DIGITAL FORENSICS

JOURNALISM

JUSTICE & ACCOUNTABILITY

MEDIA & REPORTING

TRAINING

I launched Bellingcat wanting to give people a place to publish articles about what they were doing using open source, and also to create resources for people to learn how to do it themselves." Today, Bellingcat is a fully fledged organisation with 18 core staff across research, business and administration, a management team, a supervisory board, and "with proper policies" to quote Eliot. It has recently been registered in the Netherlands as a Public Benefit Corporation. Though it is no longer a volunteer dominated organisation, Bellingcat has retained the volunteer community as a key element: a network of open source investigators spread the use of open source methods while further developing tools and methodologies, maintaining a focus on justice and accountability.

Identify, verify, amplify

Bellingcat is not in the business of fact checking so much as "fact finding". Still, disinformation is inherent to their activities, in that online open source investigation has a strong

verification component. "You're often dealing with debunking one side through open sources," says Eliot. Their work revolves around three steps. First, they identify information online related to a topic of interest or importance (subjects include corruption, corporate misconduct, racial equality, far right movements, etc., but it is important that interests be led by the team). Next they verify that information using open source techniques, drawing strategically on volunteers and on an engaged social media community who can help with precise aspects like geolocation and identification. Finally they amplify that story, through a report, video, or podcast, or even as courtroom evidence. Bellingcat is increasingly exploring the role of open source evidence in legal processes; they have worked already with the ICC and the United Nations.

Much of what Bellingcat does could be considered capacity building. They offer training to journalists and fact checkers as well as to activists, NGOs and lawyers. These trainings

also provide a source of revenue (currently 30 percent of their funding). Eliot sees Bellingcat as “a central node in the network that makes up the online open source investigation community”,

a universe which spans human rights organisations, major media outlets, and individual

Twitter users.

Collaboration is key.

“Often when we come across a project

or something to investigate, and

if it goes beyond the scope of

what Bellingcat does or if it is

something that can

benefit from a

collaboration, we’ll build coalitions of groups to work on topics.” They recognize this can be strange for media organisations who are accustomed to scoops and exclusives, so they’re strategic in bringing together non-competitive media and in building trust.

Though many organisations look to Bellingcat as an example, Eliot readily admits that “we’re figuring this stuff out as we go”. Beyond OSINT, the team is focused on the editorial and production side; Bellingcat is training their researchers in journalistic writing and building up a production company (which would be another source of income). They’re also leaning into their volunteer community (through a volunteer app and a Patreon), and trying to increase their impact with international governance stakeholders, for instance through developing standards for open source evidence in the area of justice and accountability.



BLACK LIVES MATTER

Along with Forensic Architecture, Bellingcat geolocated and verified over a thousand incidents of police violence during the Black Lives Matter protests in the United States. The project analyses them according to multiple categories, and presents the data in an interactive cartographic platform.

“OUR GOAL IS TO DEFUND DISINFORMATION”

The Global Disinformation Index is responding to the growth of digital disinformation sustained through advertising revenue.

BRAND SAFETY

TECHNOLOGY DEVELOPMENT

RESEARCH & INTELLIGENCE ANALYSIS

POLICY / EXPERT ADVISORY ROLE

The Global Disinformation Index is a not-for-profit organisation founded in 2018. Registered in the United Kingdom, they have a virtual team of 15 experts based around the world. *“We focus on the advertising networks that place ads on websites regardless of whether the advertiser knows or wants their ads to end up there,”* Clare explains. GDI’s

core contribution is to provide the advertising industry with disinformation risk ratings to assess whether or not ads should be placed on certain sites, but they support other stakeholders as well, including through their media market risk rating reports.



GDI views its effort as aligned with “Brand Safety” efforts, except that the organisation is focused on disinformation, an area that was previously ignored. They brought with them a degree of insights and intelligence that was previously lacking. *“Brand safety usually involves keyword blocking, static lists... Until GDI came along, there has never been a box for highly disinformative, toxic, adversarial narratives,”* says Clare. GDI makes use of a comprehensive framework, both human and AI powered (humans can’t assess the whole internet at *“colossal speed and scale”*, while AI isn’t nuanced enough to differentiate high production, high traffic media outlets that publish disinformation). This combined, innovative

approach provides a disinformation risk score for websites which advertisers can use in real time.

GDI has already seen impact, for example, in Google’s decision to defund all Coronavirus conspiracies. But the change isn’t happening fast enough, and it isn’t systemic. *“They’ve [Google] known about anti-vax conspiracies for years, but the virus was 7 or 8 months old before they made that decision. What about all the other anti-science conspiracy theories out there affecting our ability to think critically? Flat earth conspiracies, climate change denial...”* Clare reflects. Moreover, other tech companies providing ad services must take the same actions for a whole-of-industry response, not just Google.

A young organisation, GDI needs to ensure their own sustainability, primarily by building out new commercial products and services to fund its not-for-profit work. Looking forward, they hope to solidify *“a coalition of the willing”* against lucrative disinformation, aligning with the corporate responsibility agenda. They also want to join forces with other sectors to expose other distortions that provide a funding lifeline to disinformation. *“Payment systems, merchandising, e-commerce... there’s a whole ecosystem and that is not well enough known,”* Clare concludes.



KNOW YOUR ALGORITHM



Tracking Exposed is a non-profit, free software project that aims to analyze evidence of algorithmic personalization. It was founded in 2016 by Claudio Agosti, a self-taught hacker and developer, currently researcher at the University of Amsterdam.

RESEARCH

ACADEMIA

DATA ANALYSIS

TECHNOLOGY DEVELOPMENT

ADVOCACY

USER EMPOWERMENT

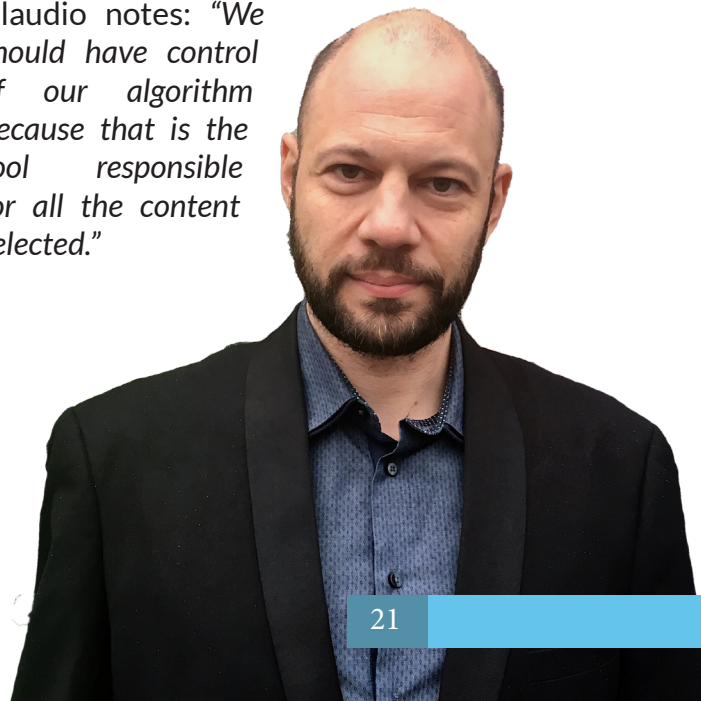
In 2018, the University of Amsterdam's Department of Media Studies expanded on Claudio's work, creating the project ALEX - Algorithms Exposed, Investigating Automated Personalization and Filtering for Research and Activism. Through his work, Claudio empowers end users to understand how aggressively the algorithm is mediating information for them. The project has been applied to four platforms so far. *"For Facebook and YouTube, it's about information quality. For Amazon and Pornhub, it's about explaining how algorithms exist in other places and can still have an impact on you."*

Claudio mostly works with researchers in the Netherlands and in Italy, though he has also carried out work related to Argentina. The academic label is important for the legitimacy of the project, and also for finding collaborators and growing the project. The work on Pornhub was carried out by a researcher he met while teaching at the Digital Methods Summer School, who wanted to expose heteronormativity on the platform.

Tracking Exposed gathers data through crowdsourcing, but getting access to wider audiences and groups is challenging. To the extent that outreach depends on marketing and visibility, this can exceed the capacity of a researcher. For a small programming project like Claudio's, recruitment requires compromises and calculations: investment in

onboarding, potential damage and security risks.

Claudio is working to make the tool simpler, with more visual results. He would like users to be able to play with and control their own algorithms. He's also interested in pursuing strategic litigation using his findings as evidence of platform rights violations. He has done election-related monitoring in the past, and he has a project coming up to work with a polling company. Elections are emerging increasingly a business case, he observes; offering this high expertise service could be a form of sustainability for Tracking Exposed. However, advertising composes a small portion of our informative experience, and so we should remain focused rather on the role of our personalized algorithms, Claudio notes: *"We should have control of our algorithm because that is the tool responsible for all the content selected."*



NEWS LITERACY FOR ALL



Juliane von Reppert-Bismarck

Belgium

Belgium, Austria, Germany

Juliane founded Lie Detectors to respond to two sides of a problem she identified in the information ecosystem.

JOURNALISM

EDUCATION

MEDIA LITERACY

POLICY / EXPERT ADVISORY ROLE

On the one hand, she was dismayed by increasing distrust in professional journalism, increasing polarisation, and the blurring of facts and opinions - a problem to which children are particularly vulnerable. And on the other hand, she noticed “the succumbing of journalism to the promise of click bait”, the response of an industry in crisis forced to take short cuts.

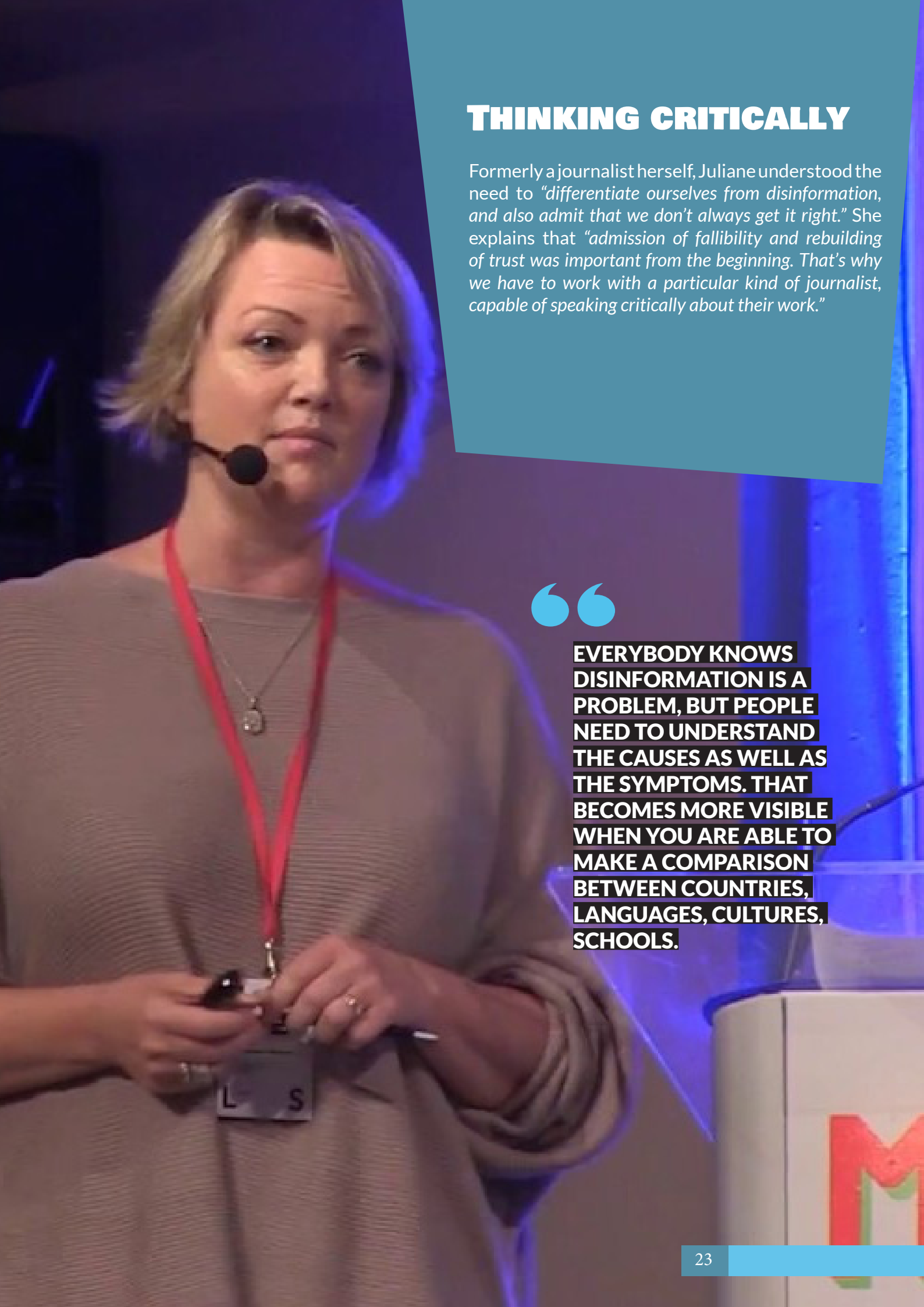
Lie Detectors brings the two sides together, placing working journalists into schools to deliver dynamic trainings. The organisation currently has a core team of 14 staff and coordinators, along with a network of 200 journalists and a growing ecosystem of educators and schools in Belgium, Germany and Austria. The approach is circular: Lie Detectors trains journalists, who then train children and educators in the classroom; meanwhile that classroom experience provides feedback for journalists and newsrooms, as well as insights for Lie Detectors’ policy development and advocacy work.

Sessions are free and journalists participate as volunteers with compensation, which means that trainings are delivered in diverse school environments and from journalists with diverse experiences. (Incidentally,

it is possible because Lie Detectors has that rarest of gifts in this ecosystem: long term flexible funding). From a research and methodological perspective, this diversity is crucial. The organisation also works closely with questionnaires and feedback forms to identify developments in mis and disinformation from the perspective of the children. Juliane explains that conspiracy theories, which appear to be a growing challenge for students, will require particular policy recommendations and particular training for educators: “*Conspiracy theories are the ultimate deep fake, because they are layers upon layers upon layers that you have to unravel.*”

“This is not something we should be feeling proprietorial about”

The service Lie Detectors is providing is in demand, both on the side of schools and from journalists (they currently have a wait list). “*This is a method and a structure. It’s so simple, it makes perfect sense... We want to amplify and facilitate organisations doing similar work.*” They increasingly hold seminars that enable teachers to deliver these trainings themselves.



THINKING CRITICALLY

Formerly a journalist herself, Juliane understood the need to *“differentiate ourselves from disinformation, and also admit that we don’t always get it right.”* She explains that *“admission of fallibility and rebuilding of trust was important from the beginning. That’s why we have to work with a particular kind of journalist, capable of speaking critically about their work.”*



EVERYBODY KNOWS DISINFORMATION IS A PROBLEM, BUT PEOPLE NEED TO UNDERSTAND THE CAUSES AS WELL AS THE SYMPTOMS. THAT BECOMES MORE VISIBLE WHEN YOU ARE ABLE TO MAKE A COMPARISON BETWEEN COUNTRIES, LANGUAGES, CULTURES, SCHOOLS.

“ALGORITHMS NOT DESIGNED TO HAVE **QUALITY INFORMATION** IN THEIR OBJECTIVE FUNCTIONS WILL NATURALLY FAVOUR **DISINFORMATION**”

AlgoTransparency



Guillaume Chaslot



France



Global

RESEARCH

ARTIFICIAL INTELLIGENCE

PROGRAMMING

ADVOCACY

Guillaume Chaslot is a computer programmer with a PhD in artificial intelligence, the founder of the consulting firm IntuitiveAI and of the nonprofit AlgoTransparency. He is currently a Mozilla Fellow. During his three years at Google, he had worked beside YouTube engineers on their recommendation system. He observed how the algorithm, which optimized for watch time, had some dangerous side effects.

Guillaume particularly became aware of the “snowball effect” boosting conspiracy theories on the platform. *“I realized the algorithm was promoting disinformation, very often more disinformation than truth. I looked at topics which were clearly disinformation, like flat earth theories, and realized that the algorithm was producing more flat earth than round earth videos.”* Guillaume had identified a fundamental design flaw playing out at a massive scale. *“We can only do so much fact checking. If the algorithm decides 70% of the views, it’s a losing battle.”*

Unable to convince his colleagues to intervene - even to recognize

the problem - he left the company and began AlgoTransparency in 2017, with the objective of exposing what content recommendation algorithms are showing people on different platforms. They currently look at Facebook, Google Autocomplete, Twitter Trending, and of course YouTube, where they monitor over 800 top information channels. The project runs on small grants and is now supported by Guillaume’s Mozilla fellowship.



At the moment, AlgoTransparency works primarily with journalists. The functioning of the tool is a bit complicated, and it takes time to explain how it works. Guillaume would like the tool to be useful

to more people and broader in what it surveys. In general, he still sees a huge need for transparency into algorithms. "People either don't know how to do it because they weren't insiders, or are afraid because they don't want to face big tech companies, or discouraged because there isn't a real business in it" he says. Access to data is still an issue, and it is insufficient for platforms to be able to decide what information they share publicly. What's needed is external pressure.

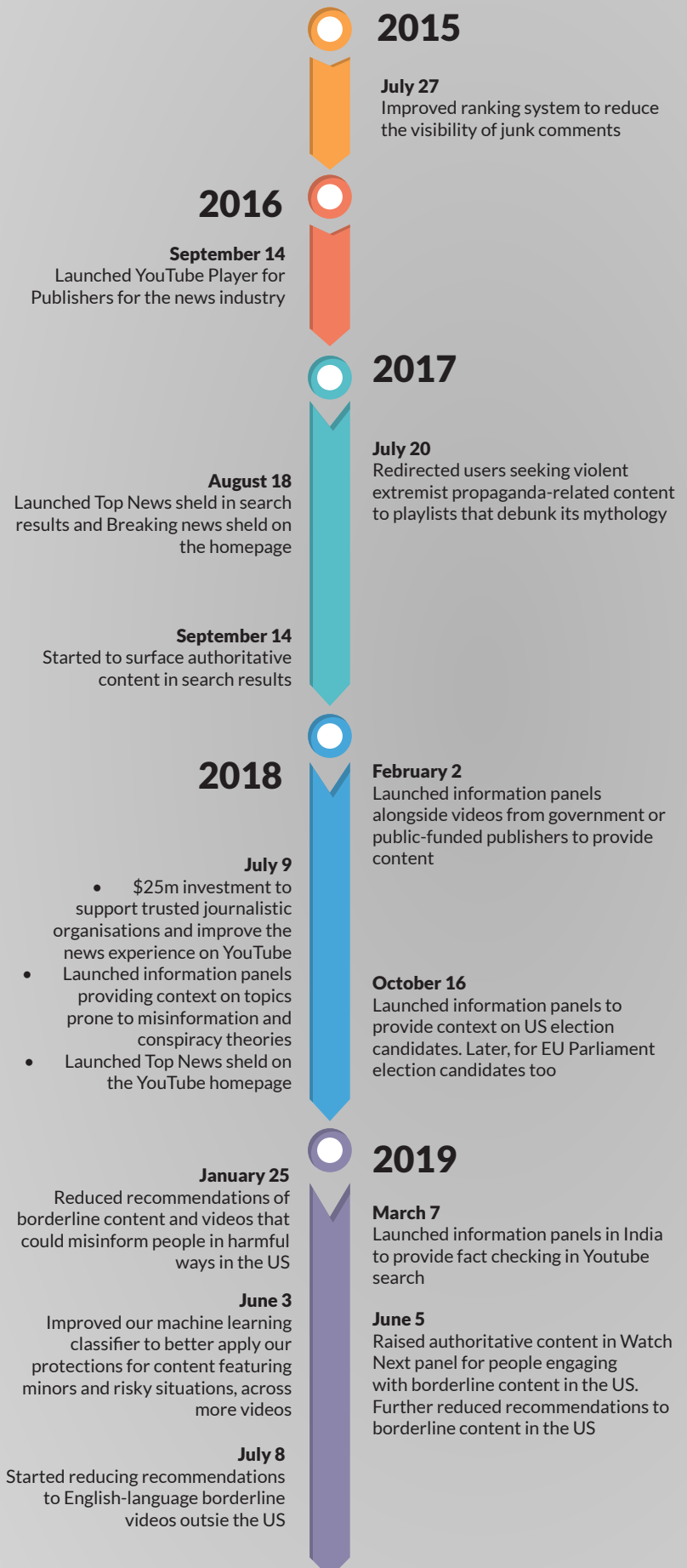
Ideally, Guillaume would like to build out a team and expand monitoring across a range of platforms, then be able to work with fact checkers, analysts, and journalists to make this data digestible for different stakeholders: individuals, researchers, and the employees at these tech companies themselves - many of whom don't want to look at their own problem. "Instead of small investigations, I want to look at the general issue. I want to make the data available and enable everyone to use it as they want."

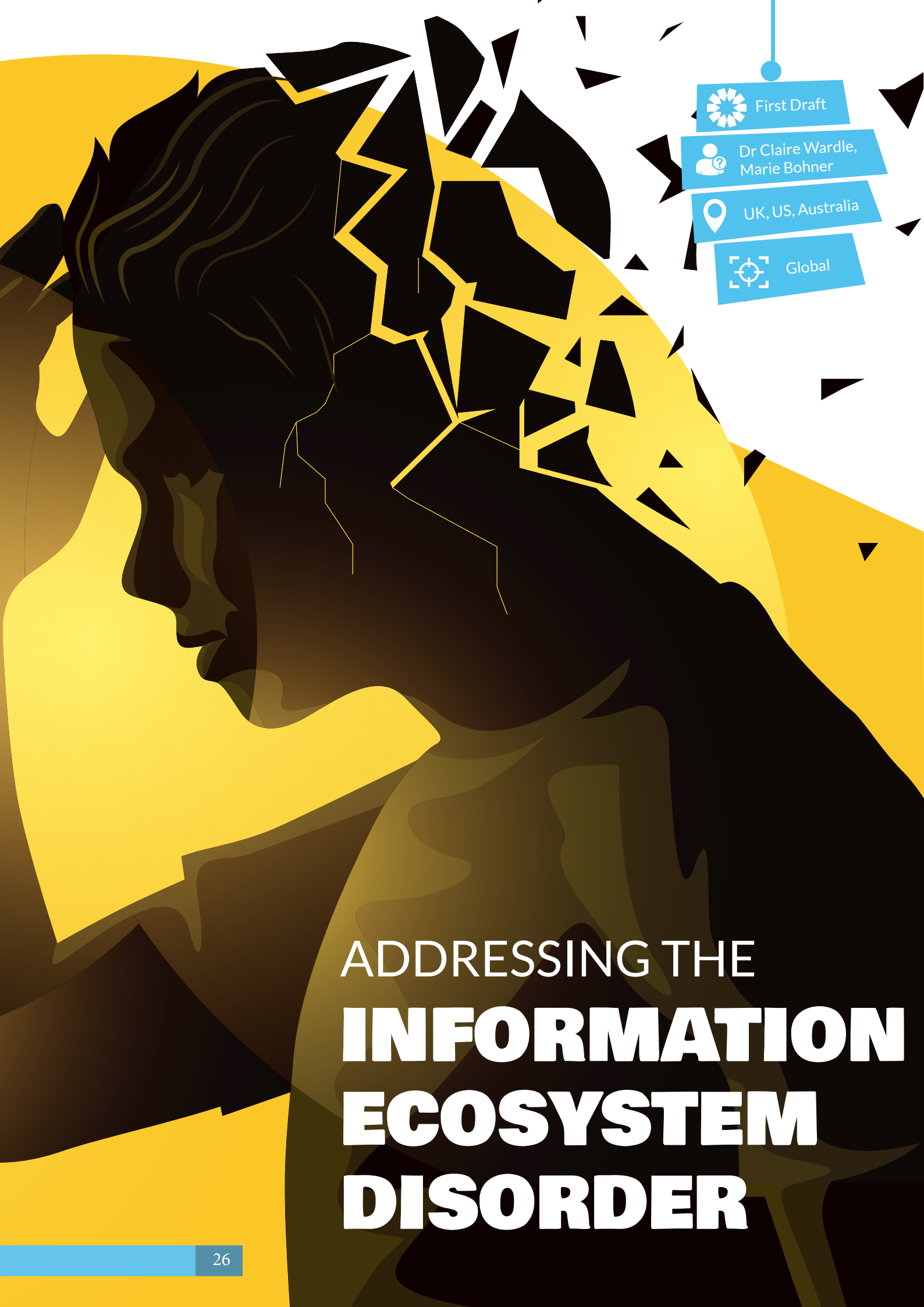
INSPIRING CHANGE

AlgoTransparency has had enormous impact. Guillaume himself has achieved media recognition as a whistle blower of sorts (he recently appeared in the film "The Social Dilemma"), as the policy discussion has increasingly turned towards the need for algorithmic transparency. The impact on platforms is more difficult to understand. "YouTube changed their algorithm 30 times to address the exact issues I was talking about", Guillaume notes, referring to their actions in 2017 on the amplification of terrorist content. But they did not take action regarding disinformation like the flat earth content that he had brought to their attention.

YOUTUBE KEY PRODUCT AND POLICY LAUNCHES TO RAISE AUTHORITATIVE VOICES TO REDUCE THE SPREAD OF BORDERLINE CONTENT SINCE 2015

- Raise authoritative content
- Reduce spread of borderline content





-  First Draft
-  Dr Claire Wardle,
Marie Bohner
-  UK, US, Australia
-  Global

ADDRESSING THE
**INFORMATION
ECOSYSTEM
DISORDER**

First Draft was founded in 2015 as a nonprofit coalition of nine founding partners with the mission of protecting communities from harmful misinformation. That original coalition has expanded greatly since, and now includes international partnerships with newsrooms, universities, companies (recently they have begun work with Spotify), institutions such as the WHO or UNICEF and other NGOs.

With offices in three time zones (London, New York, and Sydney), and a team of 52 employees, First Draft is monitoring around the world and around the clock. First Draft is addressing an “information ecosystem disorder”, says Marie. This work is guided by their Information Disorder Framework which Claire Wardle, co-founder and US director, first laid down in a 2017 report for the Council of Europe with co-author Hossein Derakhshan. This framework helped put First Draft on the map, but it is one of many approaches the organisation has pioneered since its founding. CrossCheck, a collaborative approach to reporting around elections, has been used in the US, UK, France, Germany, Brazil, Nigeria and Spain, and inspired similar initiatives in many other countries. “Elections are our DNA”, Marie explains. First Draft has developed a strong editorial department, which produces daily and weekly briefings from their global monitoring and they contribute regularly to the research discussion. Through their global Partner Network of newsrooms, fact checkers, human rights organisations and technology platforms, they have been working to share knowledge and set standards in the areas of collaboration, training and research.

Rooted in strong local partnerships and collaborations, First Draft takes a global approach in order to gain a cross-border view of information



THE DECEPTIVE SEVEN: COMMON TYPES OF MISINFORMATION & DISINFORMATION



Satire or parody
No intention to cause harm but has potential to fool



Misleading content
Misleading use of information to frame an issue or individual



Imposter content
When genuine sources are impersonated



Fabricated content
New content that 100% false made to deceive and do harm



False connection
When headlines, visuals or captions don't support the content



False context
When genuine is shared with false contextual information



Manipulated content
When genuine information or imagery is manipulated to deceive

disorder. *“This approach aims to overcome individual biases, to compare stories, and to share similar problems,”* Marie explains.



**WHILE THE MEDIA TEND TO
FOCUS ALL ON THE SAME THING,
WE TRY TO CONCENTRATE ON
WHAT PEOPLE ARE LOOKING FOR
ANSWERS TO**

Recently, First Draft has been exploring what they identify as data deficits, areas where a lack of relevant and readily accessible information may lead to misinformation. They are trying to identify people’s questions, for example, using Google Trends, or else through direct conversation with their audiences. To better understand misinformation and data deficits in closed spaces, they apply crowd-sourcing strategies, such as tip lines, where members of the public flag content for examination.

The organisation is growing quickly, but rapid expansion during the pandemic has been a challenge. While they have had plenty of work to do around elections - particularly the US election - they still lack long term, flexible finance that guarantees an organisation sustainability and peace of mind. *“It’s uncomfortable being 100% dependent on funding, so we are developing diverse revenue streams to ensure we can continue supporting those relied upon for accurate information in the years to come”.* Meanwhile, there is no lack of work to be done. While continuing their focus around democracy and health, particularly vaccines, they are already looking ahead to other global topics that are vulnerable to disinformation, such as the looming economic downturn and the climate crisis. There is more awareness and acknowledgement of the threats and risks of online disinformation, an urgent need to empower societies with knowledge and skills, and for collaborative efforts to increase. After 5 years building experience and expertise, First Draft is needed now more than ever before.

CAPACITY BUILDING

First Draft is building capacity across all imaginable stakeholders. In addition to their series of trainings for journalists (from personal and cyber security to the ethics of investigations, etc.), their Local News Fellowship Program around the US 2020 election equipped local and regional journalists in social media monitoring. They pioneered information crisis simulations to help build resilience, and have a growing number of resources targeted at more general audiences as well.



Dr Claire Wardle



Marie Bohner

“MACHINE LEARNING OFTEN GETS IT WRONG,

I WANTED TO GO A STEP BACKWARDS”

Account Analysis



Luca Hammer



Germany



Global

Luca began the project behind Account Analysis one weekend in 2017, in response to a growing debate about bots.

RESEARCH

DATA ANALYSIS

OSINT

TECHNOLOGY DEVELOPMENT

Journalists were talking about social bots who influence elections and debate online.” While he knew a handful of existing tools to automatically identify bots on Twitter, usually by giving accounts a score, it was difficult to understand how those scores were produced. The research community tends to agree that most of the tools used to identify bots are flawed. *“It’s nearly impossible to reliably identify bots from the outside; platforms can do this better because they have access to more data points.”*

His approach has been to build a tool that wouldn’t deliver a score, but rather allow users to evaluate for themselves. Account Analysis lets users study Twitter accounts by looking at criteria like how many retweets they post, which hashtags they use, which websites they link to most often.

“For researchers to fully understand what’s going on, I’m convinced that you have to see the data yourself.” Luca believes it isn’t enough to just “get the numbers” of how often contents are shared. In order to understand influence and impact, researchers need to be able to see the data directly. This requires more communication between platforms, the researcher and programmer communities. Twitter’s API is by far the most accessible for researchers: anything public on the platform is public on the API. Telegram, which Luca is beginning to focus on more closely, is similarly unrestricted. Facebook, who’s API is notoriously difficult to use, has a less clear cut user-facing definition of private and public space, which contributes to the challenge. Still, all platforms could do more to make their APIs accessible for research.

Luca is less interested in bots nowadays. “I’ve discovered bot-nets on Twitter and published about them, but I think their impact is low. The platforms have great tools and try to stop spamming.” He’s more focused on platform manipulation. “It doesn’t matter if it’s one person or several people, I’m interested in what they do and what they try to achieve.” Account Analysis helps him study this because it lets him understand how those accounts work. He’s also interested in false information sharing in big public channels on Telegram. In the future, he aims to make more tools on more platforms accessible to more people. He also has plans to add features, including tools to be able to analyse debates, not simply accounts.

KEEPING IT FREE

Account Analysis has both a free and paid version (for 15 euros a month, the tool essentially works more quickly). He developed this ‘freemium model’ after finding the business case for marketers, who use the tool to evaluate the accounts of influencers and politicians. This business case covers his server costs and allows him to continue to provide the tool for free. His primary motivation is to serve the OSINT and journalist community. He also offers workshops for journalists.



© Tony Gigov

2020, A TURNING POINT IN OUR RESPONSE TO DISINFORMATION?

This research was conducted at a moment when the disinformation challenge has never seemed higher. From the Covid-19 health crisis and parallel 'infodemic' to elections in the US and Belarus, 2020 has been a tumultuous year for our information ecosystems. It has also been a significant year for regulatory response to disinformation. The European Democracy Action Plan, a roadmap to enhance democratic integrity and resilience across the European Union, was released in December. It was followed by the Digital Services Act and the Digital Markets Act, regulatory packages to clarify the role and responsibilities of online platforms and increase accountability for online disinformation.

Regulation of online platforms will reduce opacity and better equip anti-disinformation actors like those interviewed here, but regulation is not a silver bullet. As a diffuse and rapidly evolving set of challenges, disinformation requires a broader response. Key to this response is a thriving, decentralized

civil society ecosystem. Disinformation is a horizontal challenge by which more and more actors find themselves confronted. A decentralized architecture can catalyze multiplier effects and build capacity and resilience among those currently at the periphery of the disinformation threat (climate activists, health professionals, etc.).

Our aim in this project has been to try to understand this growing civil society network - their struggles and their successes - in order to better support their activities and fortify this ecosystem as a whole. Far from a comprehensive survey, this research can only scratch the surface of this vast landscape of actors. And yet even with this partial view or snapshot, we have been able to highlight trends and shared challenges, to pinpoint areas where collaboration is possible and where more support is necessary. It is our hope that these findings will further empower these actors and those like them, and provide evidence for the competent authorities to act.



ACKNOWLEDGEMENTS

This report is made possible by the support of the Friedrich Naumann Foundation for Freedom. Thanks are also due of course to the actors interviewed and featured here. They were generous in taking the time to participate in this project, and earnest in their contributions. The report that follows is rich because of their openness. Particularly admirable was their ability to discuss their vulnerabilities and struggles. In the face of these challenges, their tireless optimism is striking and kinetic.

ABOUT EU DISINFOLAB

The EU DisinfoLab is an independent non-profit organisation focused on tackling sophisticated disinformation campaigns and documenting the disinformation phenomenon in Europe. As a small civil society organisation, EU DisinfoLab acts as a facilitator within the community of disinformation experts, putting emphasis on building partnerships and fostering collaboration between relevant initiatives. We aim to serve as a gathering place for experts and organisations to exchange best practices, cooperate, and develop new approaches to countering disinformation. We seek to amplify the voices of our community of partners and contribute collective expertise to policymakers, through research, advocacy, and policy recommendations. Our activities and partnerships are global, but we have a particular focus on EU legislation, EU institutions and EU Member States.

WANT TO KNOW MORE?

In addition to this magazine, you can visit the website dedicated to this project.

Visit
disinfo.eu/manyfaces

to find out more about the actors and initiatives featured here. You can also find other materials related to this project, including our recommendations for how different stakeholders can help foster a resilient, decentralized and harmonized civil society ecosystem.