FROM UNSTRUCTURED DATA TO INTELLIGENCE

Uncovering the power of social media analytics

By Leendert de Voogd and Tara Beard-Knowland | November 2020

IPSOS VIEWS

GAME CHANGERS

Do you ever imagine what it must have been like to have been there at the beginning of market research? When researchers themselves were trying to figure out what was important and how to make the best sense of all the information? According to Google's Ngram Viewer[1], the concept of market research entered the broader lexicon about a hundred years ago but then built relatively quickly (see Figure 1). We are early in that journey with social intelligence and analytics (originally described as 'social listening' as it started with the monitoring of social feeds) but the fundamentals that make it good research are now emerging.

Figure 1 Evolution of the use of certain phrases in books

Source: Google Ngram Viewer

Social intelligence is the ability to collect, monitor and analyse available social media data feeds (including social media networks, blogs, forums, comments, etc.) to understand what is being said about a topic, brand, organisation or other entity.

The concept of social intelligence came early in the development of social media. However, it has roots in the technology sector, rather than in research. Although many research agencies (including Ipsos) were dabbling in it even a decade ago, it has only recently been taken seriously as a proper research discipline, rather than just a way to scrape together as much detail as possible and watch what people are saying about brands. Now, we know how social media and other unstructured datasets (vast and 'messy' open-ended data that cannot be easily classified into neat buckets) can tell us something useful in a consistent, replicable way without using technology just for technology's sake.

FROM TECHNOLOGICAL POSSIBILITY TO A RESEARCH-LED APPROACH

Over the last 15-20 years, significantly improved computing power and advances in algorithms and artifical intelligence (AI) have enabled us to process more data than was previously digestible. Whether you are talking about social media monitoring platforms or the many different AI-based analytic approaches, these technology and data science advancements are impressive. They help us to make sense of huge amounts of data quickly and efficiently. But they need more to constitute a useful and real research approach.

Across the industry, organisations from research agencies to technology providers to one-man-bands have tinkered with the possibilities of social media data. Service industry organisations harnessed some of the power of unstructured data early on, but this was only ever one dimension. The challenge is that, like quantitative and qualitative research, there are many, many different use cases.

We believe the distinction between what is research and what is not research is important. Sometimes, unstructured data can be used for non-research purposes, social media data in particular. However, there are many insight-led use cases. It is therefore critical to bring research discipline and rigour to it to ensure that it is done well, considering both the client's needs and the needs of the research participants. It is only by applying these principles that it will be considered as a legitimate research discipline, bringing clear value to clients.

Based on our experience over the years, we have observed three broad methodological building blocks of any meaningful social media intelligence programme: – the social media intelligence platforms, AI-led advanced analytics, and human-driven insight discovery. On their own, each building block brings something useful to answering critical client questions. Combined, they bring powerful insight. Like any research methodology, the capability to conduct the actual research is not enough, to do it well you must also have the requisite skills.

Significantly improved computing power and advances in algorithms and AI have enabled us to process more data than was previously digestible.

THE ROLE OF EACH BUILDING BLOCK

As the earliest type of social media data analysis, Software as a Service (SaaS) platforms have had more time to diverge and converge in terms of approaches. In research terms, SaaS platforms do more than just act as a way to gather as much data as possible, they should be gathering high quality data and allowing you to sift through it effectively and efficiently.

The wide range of data available made analytics a big growth area in the tech sector. There is a lot of data, and there is no way human beings can look at all of it or even most of it. Enter AI-enabled analytics. These are truly AI because they involve artificial intelligence in the form of NLP/NLU (natural language processing/understanding), machine learning and other approaches to find the patterns and decoding language (or pictures/videos). Applying algorithms and AI to our datasets helps us to see patterns. This is not just using AI for its own sake but ensuring it is targeted to the right questions and employing the best capabilities that will link to real client challenges. The role of Data Scientists cannot be

underestimated here. They are the ones who are selecting and developing meaningful algorithms to perform specific tasks. In our field, AI is too often advertised as a black box magical tool embedded in a platform, when in fact you need to combine different techniques to do good research. For example, simple vector models are great for topic modelling but are often inefficient for category specific sentiment models.

The technology will do exactly what you tell it to do, no more, no less. So, while SaaS platforms and AI-enabled analytics allow us to surface interesting facts, data and patterns, human-driven insight discovery takes these pieces and fits them together to unearth meaning. By applying research thinking to what we do and leveraging proven frameworks, there is a critical research role for human beings in insight discovery.

THE THREE KEY BUILDING BLOCKS OF SOCIAL INTELLIGENCE

Social media intelligence platforms: Software platforms designed to enable social listening from different sources of data, providing a real-time access to various metrics through the means of interactive dashboards.

AI-led analytics: Text, picture and video analytics designed to make sense of unstructured data using natural language processing (NLP), machine learning, data mining, statistical analysis, etc.

Human-driven insight discovery: Individual researcher contribution to finding insights from social media data using analytical frameworks.
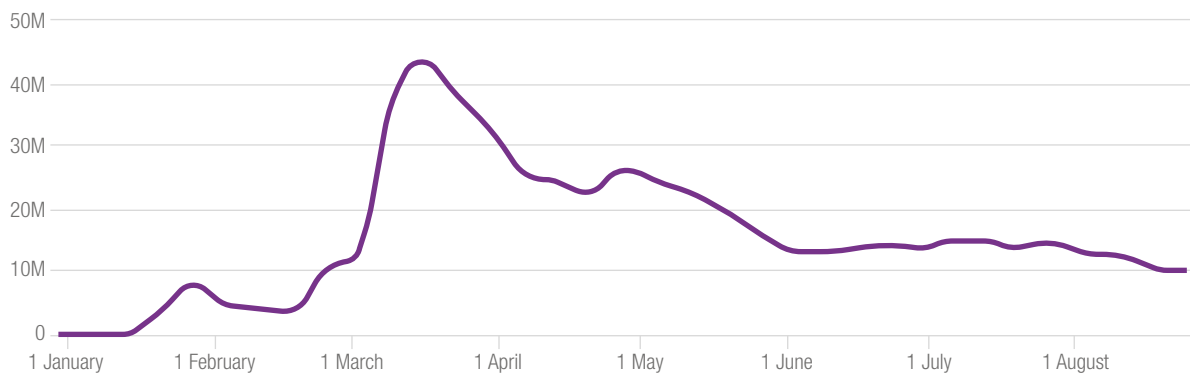
DEMONSTRATING THE BUILDING BLOCKS IN ACTION

In order to demonstrate what each of these methodological building blocks brings to a particular question, we will take the COVID-19 pandemic, and key lessons for brands, as a topic area. We have done (and continue to do) extensive work, across a wide range of clients and question types, to better understand the status and implications of the pandemic.

THE SOCIAL MEDIA INTELLIGENCE PLATFORM

Using a social media intelligence platform like Synthesio[2], we can uncover some interesting nuggets to understand what is happening. In the first instance it is just useful to see how the conversation is evolving compared to historical data that we can also collect post-event. For example, by the middle of May we had more than one billion tweets alone about COVID-19. Three months later, the figure had doubled (see Figure 2).
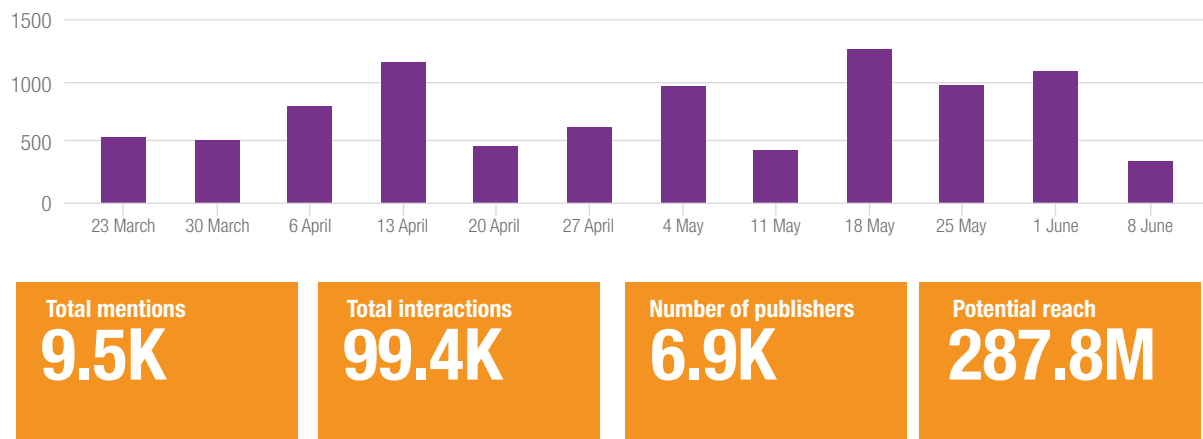
Using the platform, we can also explore more around particular topics to see at a topline level what is standing out before we dig into the data more deeply. This type of built-in AI that highlights abnormal trends can save a great deal of researcher time in chasing 'red herrings'. For example, thanks to the Signals feature of Synthesio, we can see that alcohol was a topic much mentioned with COVID-19, particularly from April to June (see Figure 3).

Figure 2 Original posts about COVID-19 on Twitter in 2020, excluding retweets (global base)

Source: Synthesio, an Ipsos company

Figure 3 Mentions of alcohol on social media

Total mentions	**Total interactions**	**Number of publishers**	**Potential reach**
9.5K	**99.4K**	**6.9K**	**287.8M**

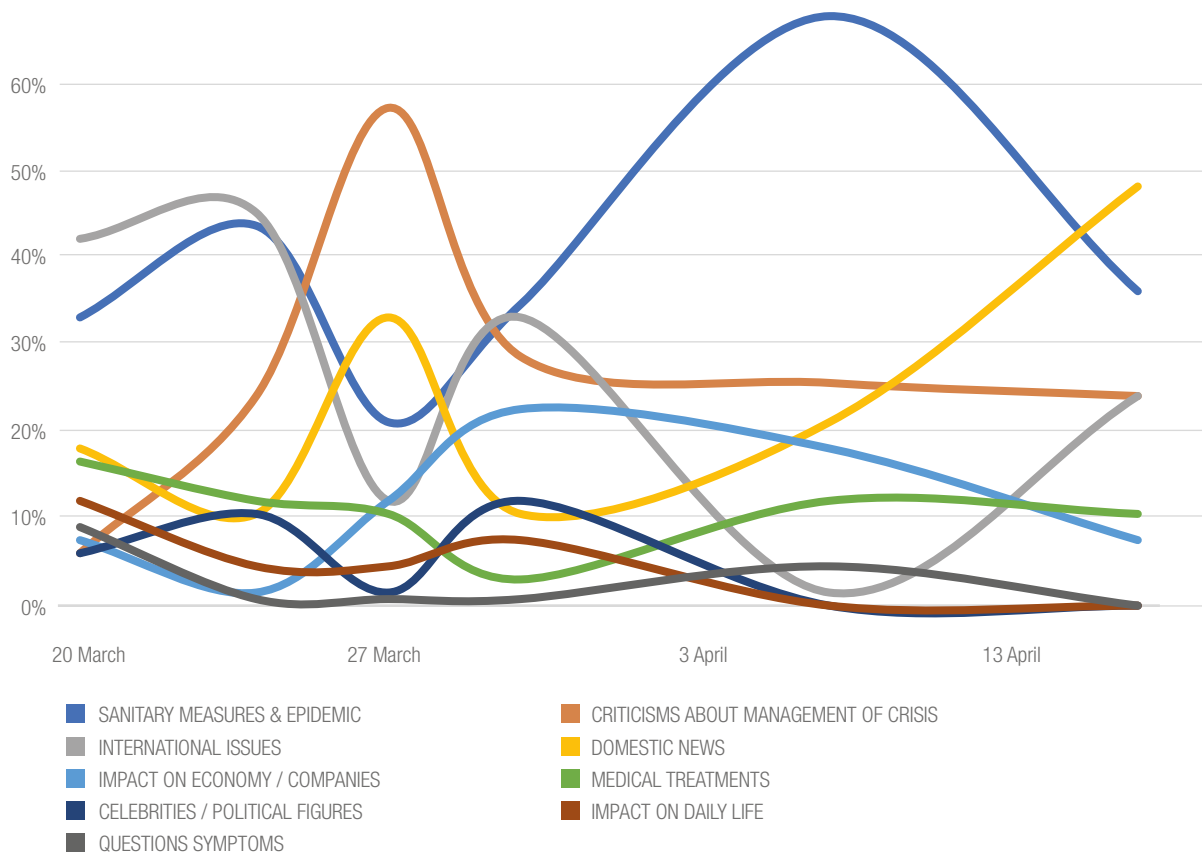*Source: Synthesio, social media posts written in English, March-June 2020*
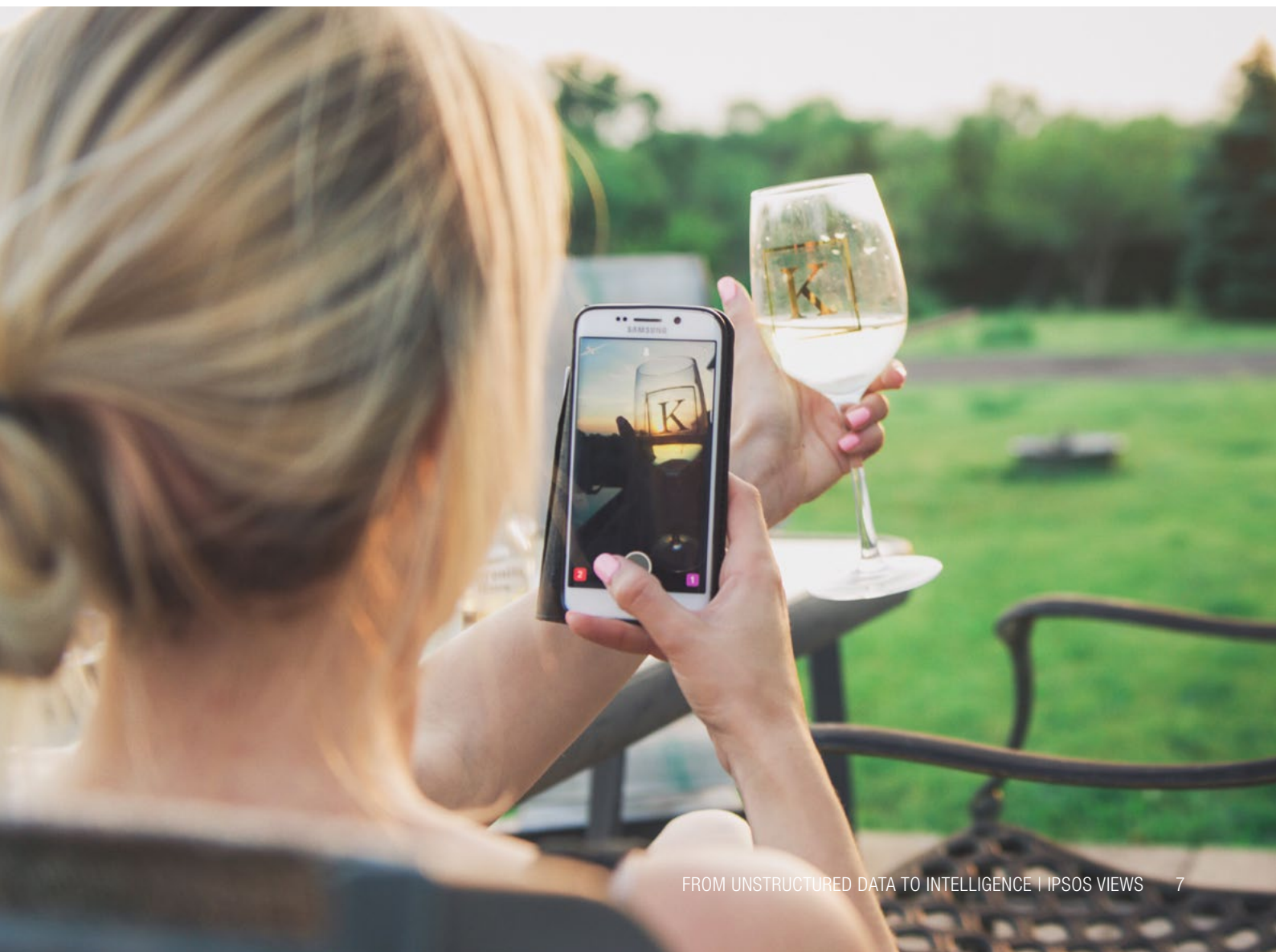
AI-LED ADVANCED ANALYTICS

While platforms themselves can offer a first understanding of the main topics being discussed, an advanced AI-enabled analytic lens is often needed to get a more granular understanding of the true and spontaneous nature of people's concerns, needs and expectations. New deep-learning algorithms help us to better understand what key themes are coming through and tell us if they differ between different markets or if they are changing within the same market. For example, in Figure 4 we see that, in France, the volume of posts about key topics were changing week-by-week (and in some cases day-by-day, as the news emerged).
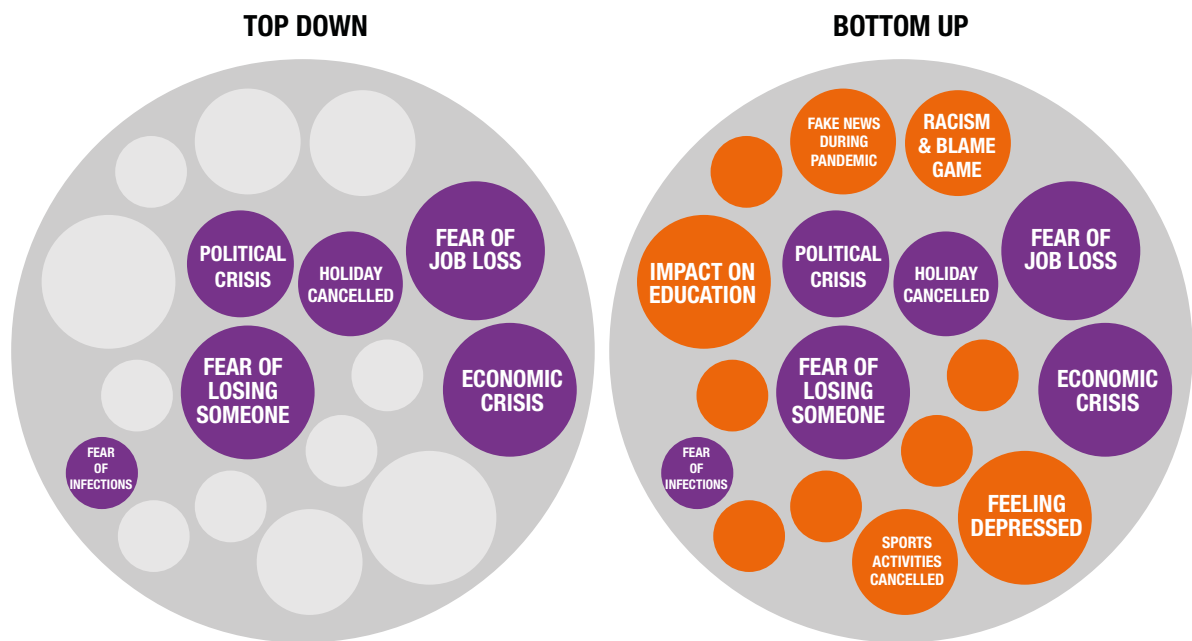
We can also use AI to understand emotion. Here we have focused on people's fears (see Figure 5) generated by the COVID-19 crisis. A typical top-down topic modeling approach would consist of quantifying the themes we expect to see in the dataset. One could have assumed that people were indeed concerned about their ability to keep their job or the fear of losing someone. Combining this top-down approach with a more consumer-centric one that leverages the power of deep-learning algorithms (we call this a bottom-up approach), we are able to surface additional new fears that we did not expect to see – such as the negative impact of the lockdown period on the education of kids and the emergence of racism and blame game during the crisis.

.

Figure 4 Evolution of topics in France (AI-generated topic modeling)

Legend:
- SANITARY MEASURES & EPIDEMIC
- INTERNATIONAL ISSUES
- IMPACT ON ECONOMY / COMPANIES
- CELEBRITIES / POLITICAL FIGURES
- QUESTIONS SYMPTOMS
- CRITICISMS ABOUT MANAGEMENT OF CRISIS
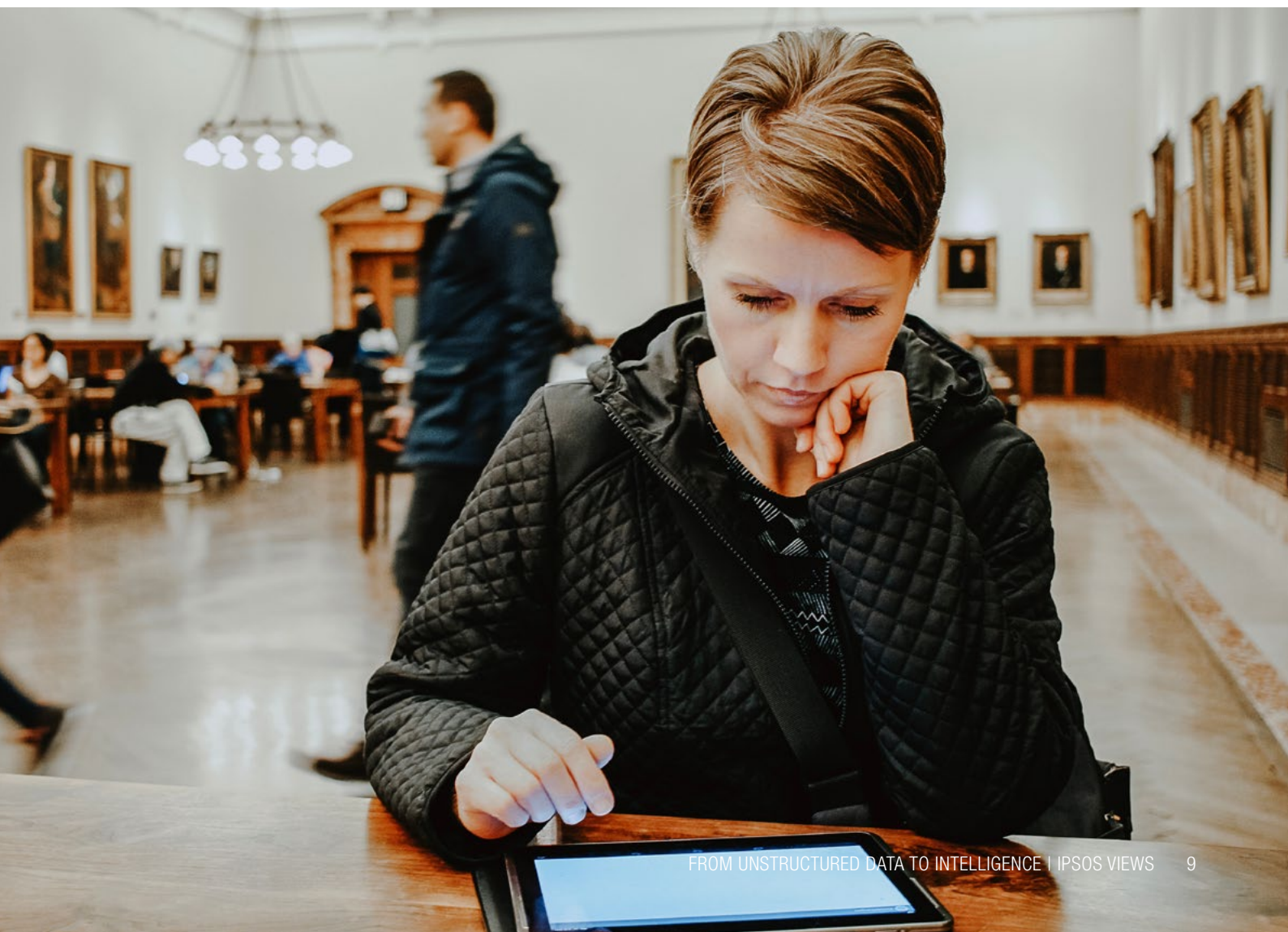- DOMESTIC NEWS
- MEDICAL TREATMENTS
- IMPACT ON DAILY LIFE

Source: Social Intelligence Analytics, Ipsos, March-April 2020

Figure 5 People's fears: surfacing the whole social insight spectrum

TOP DOWN

BOTTOM UP

POLITICAL CRISIS

HOLIDAY CANCELLED

FEAR OF JOB LOSS

FEAR OF LOSING SOMEONE

ECONOMIC CRISIS

FEAR OF INFECTIONS

FAKE NEWS DURING PANDEMIC

RACISM & BLAME GAME

IMPACT ON EDUCATION

POLITICAL CRISIS

HOLIDAY CANCELLED

FEAR OF JOB LOSS

FEAR OF LOSING SOMEONE

ECONOMIC CRISIS

FEAR OF INFECTIONS

FEELING DEPRESSED

SPORTS ACTIVITIES CANCELLED

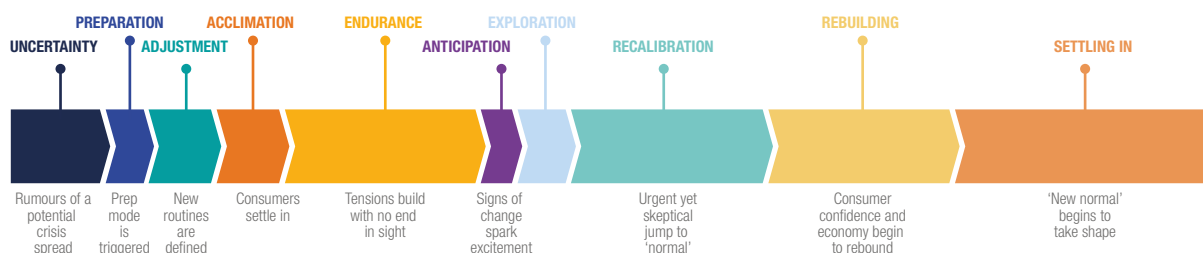*Source: Social Intelligence Analytics, Ipsos, social media posts written in English*

HUMAN-DRIVEN INSIGHTS DISCOVERY

From the human-led insights perspective, we can begin to drill down more into the whys and wherefores of what is happening and the implications in terms of changing consumer behaviour and activity. This helps us to understand what it might mean for brands or organisations. This is especially useful when you can tie it back to other frameworks. For example, the Ipsos Pandemic Adaptability Continuum[3] (see Figure 6).

Figure 6 The Ipsos Pandemic Adaptability Continuum

UNCERTAINTY | PREPARATION | ADJUSTMENT | ACCLIMATION | ENDURANCE | ANTICIPATION | EXPLORATION | RECALIBRATION | REBUILDING | SETTLING IN

Rumours of a potential crisis spread | Prep mode is triggered | New routines are defined | Consumers settle in | Tensions build with no end in sight | Signs of change spark excitement | Urgent yet skeptical jump to 'normal' | Consumer confidence and economy begin to rebound | 'New normal' begins to take shape
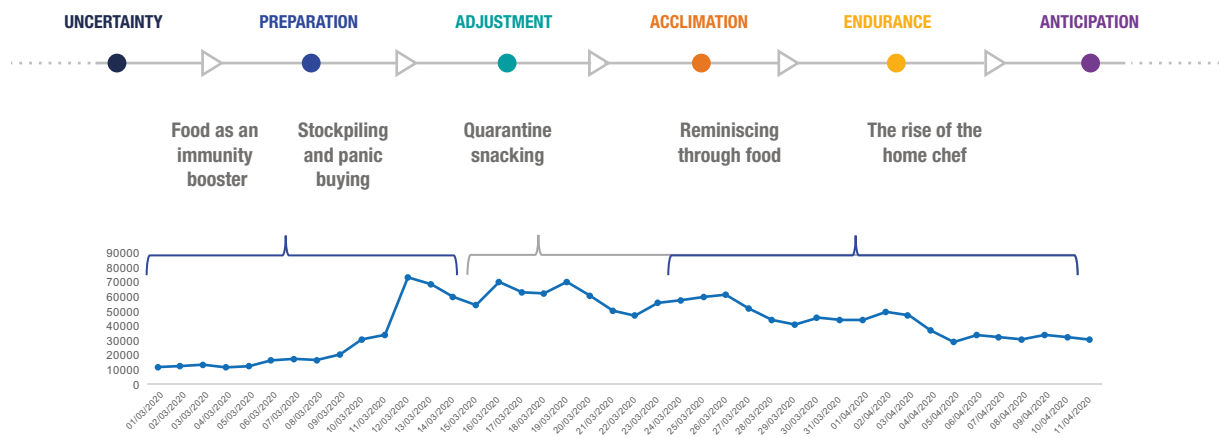
Source: Ipsos Global Science Organisation

This helps us to orient the findings and anchor them into something bigger. For example, when we look within food behaviour as the COVID-19 crisis emerged and then took hold, we saw the way that consumers were moving into different phases (see Figure 7). This means that new opportunities were arising for companies.
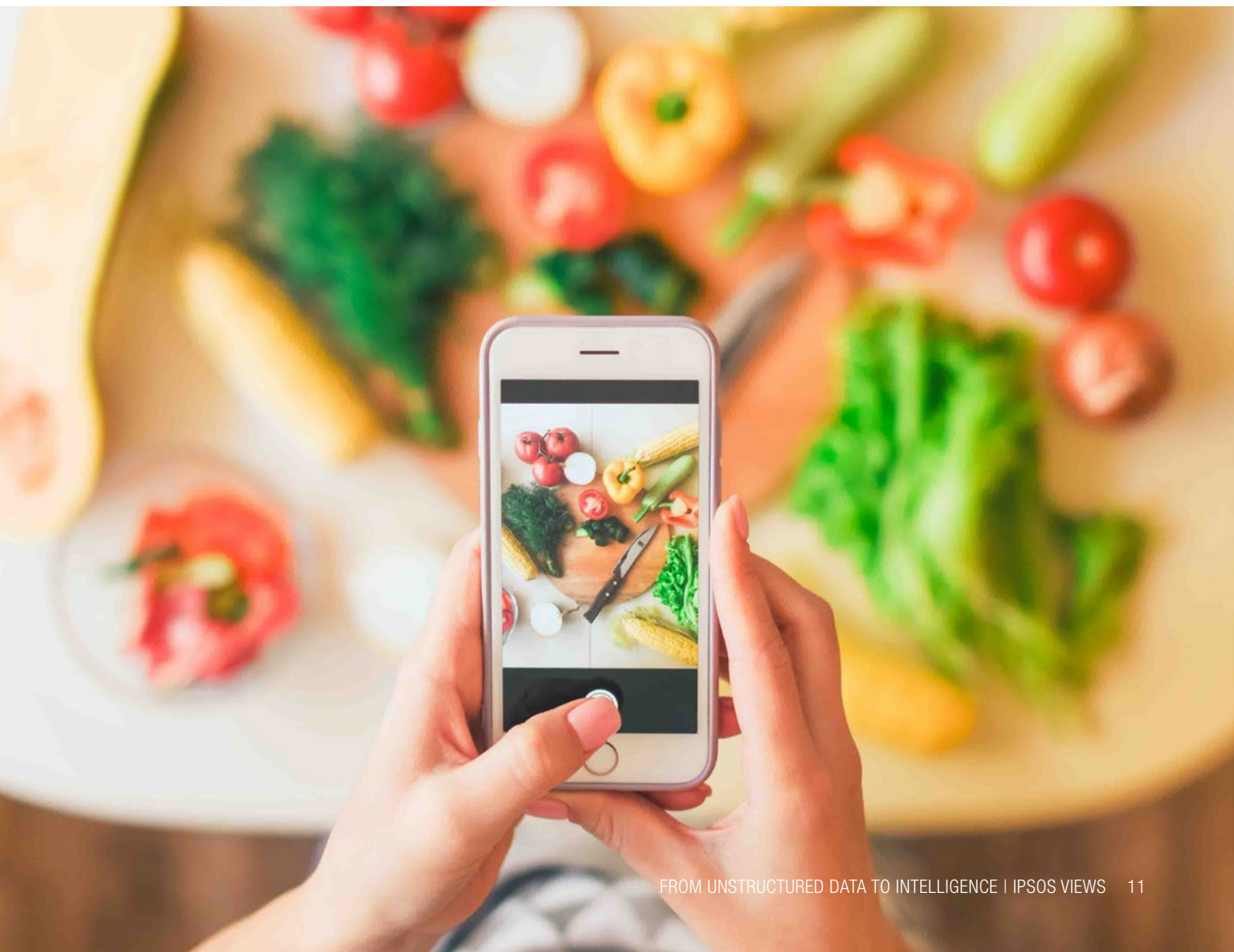
Having a human look at the data can also prevent drawing wrong conclusions. At the start of the pandemic, we began tracking online conversations about COVID, but as our analysts reviewed the evolution of the data, we saw the decline in mentions.

As a result, we might have thought that the topic was not so important anymore. On the contrary, COVID was part of our daily lives… People did not have to mention it anymore in their social posts. Using the strength of our analysts, we quickly developed a robust lifestyle-focused view to understand the many layers to our pandemic experience – life in the home, entertainment, health and wellness, etc. More often than not, relying simply on a trendline does not tell you the whole story. This illustrates how each layer can add new and important information to help us better understand what is happening.

Figure 7 Social conversation reveals how this emotional journey has unfolded with food

UNCERTAINTY   PREPARATION   ADJUSTMENT   ACCLIMATION   ENDURANCE   ANTICIPATION

Food as an immunity booster   Stockpiling and panic buying   Quarantine snacking   Reminiscing through food   The rise of the home chef

Source: Synthesio, social media posts written in English, March-April 2020

THE MYTHS AND REALITIES OF SOCIAL INTELLIGENCE AND ANALYTICS

As with any emerging practice, it is easy to misunderstand what is possible and what is being done. This makes it easier to get taken in by the promise of a diamond of an approach, which turns out only to be particularly brilliant glass. We'd like to dispel some of the most commonly occurring myths.

Myth 1: Social media data is worthless – it is all Twitter data and it doesn't tell me anything

Reality: Twitter data can sometimes dominate results, if you allow it to. Sometimes that is all to the good. Even the micro-blogging, short and sweet text has value in it, especially in its volume. But Twitter data isn't right for every single question or approach. Just as you would consider how to build a good sample in a qualitative or quantitative research project, it is important to consider your universe in gathering social media data. You must make sure you can access a wide variety of sources in your chosen platform before launching a social media research programme.[4] On several occasions, specialised blogs and forums are more insightful than Twitter. It is also critical to treat the data coming from various sources as an heterogenous mass. Indeed, a tweet about "Tesla" for instance will carry a very different meaning layer compared to a "Tesla" usage experience review or a "Tesla" post in a forum on the future of mobility. Experienced analysts are needed to bring this to life. Pretty much in the same way as we split and report our survey data comparing target groups, socioeconomic or attitudinal criteria.

Myth 2: Social intelligence is only for understanding our PR and social media marketing efforts and to manage crises

Reality: Social media data and intelligence can be useful in those cases, but there are so many more questions that it can help with. We have used it reliably in many cases, including to guide and supplement segmentation and audience understanding; to understand a market landscape and broader context; to understand signals of macro trends and deep dive into micro- and nano-trends; to surface insight about why people engage (or not) with particular products and services; to understand specific brand moments that comprise a brand experience; and to get a clear picture of what makes people tick and what ticks them off on a wide range of topics. This is by no means an exhaustive list, but these are approaches we've taken time and again to create actionable insights.

Myth 3: We can just look for what's interesting out of all of the social media data

Reality: Given enough time and money, you might eventually get there (while also burning an incredible amount of electricity), but resources are rarely unlimited. As the saying goes: "If one does not know to which port one is sailing, no wind is favourable." Therefore, it is critical to create some boundaries around your research, so that you know you've found something interesting. Your human expertise and industry knowledge will help you to develop some relevant boundaries.

Myth 4: No one talks about this topic and, therefore, I can't do any research using social intelligence

Reality: Sometimes, we can be a little too specific in our requirements. This is not uncommon with brand research. It does not mean that there is no value that social media data can provide – in fact it can help steep us more in the consumer reality. Consumers might not talk about a specific brand's product – for example a specific brand of frozen peas or tinned tomatoes – but they will talk about very relevant topics, such as how they cook at home, favourite flavours and dishes, and what they aspire to bring to their mealtimes.

Myth 5: Social intelligence will be fast and cheap

Reality: It can be faster and cheaper than more traditional methodologies for sure. It can also give you deeper insight into what is moving people because it is unprompted. Because there are layers of refinement needed to prepare a strong and relevant dataset as well as human-led insight, good social intelligence and analytics requires a certain level of investment.

Myth 6: AI does it all

Reality: Analysis of any unstructured data would be incredibly restricted without AI. In the first place, it would be nearly impossible to find the data we wanted to review. The power we have as human beings is to be able to find meaning and importance in the factors that AI brings to our attention. AI is powerful and can show us patterns we might not have noticed ourselves. But it takes a human being to say *why* that is important, assuming it is a human being who understands the question and the broader category and topic.

Myth 7: All tools are created equal

Reality: Each tool has its pros and cons and there are always trade-offs depending on what you want to do. There are many different factors to consider, depending on what's most important to you and your organisation – data sourcing, level of cleaning, ease of use, built-in widgets, volume of data, level of in-built analytics and AI, language and geographic coverage, etc. What's more, these are changing all of the time. It can feel overwhelming to stay on top of it, but regular reviews of needs versus what you can get are important.

Myth 8: It is very easy to become an expert in these approaches. Only a select few can possibly do this

Reality: Like any research approach, a little knowledge can be dangerous. You cannot become an expert in all aspects of social intelligence and analytics. But it is possible for a keen and curious mind to learn how to make the most of unstructured data and how to execute worthwhile and reliable social media data research and insight.

The three building blocks cover many different skills. Within the human-led insights strand, those who have a mix of quantitative and qualitative skills and a good understanding of social media in general to respect the specificities of the different data sources generally are the ones more suited for this kind of project.

We also believe that the Insight Function within our client's organisation need to learn more about how to use social intelligence and unstructured data in a meaningful way. It is critical to get the basics right to prevent two risks for our clients: falling into the shiny object trap (and misuses) or simply missing the opportunity of leveraging this type of information. We are deploying efforts with our clients to empower them so they can more easily navigate the complexity of the social intelligence and analytics world.

KEY TAKEAWAYS

The area of social intelligence and analytics is exciting and still emerging. Like the emergence of market research as an overall discipline, we amass learning as we go about the ways in which we can create useful meaning and insight from the datasets.

When using unstructured data, especially social media data, to gather insights, there are some important factors to consider as you evaluate the pros and cons of different approaches:

- **What do we really want to know about?** To avoid boiling the ocean and wasting unnecessary time and therefore money, you need to be targeted in what you are asking.
- **Is this something people are likely to be talking about?** Feasibility is critical and any good research practitioner of social intelligence will do a feasibility check before running a project.
- **What sources will be most important to answering my questions?** There are many different sources. People post on forums for different reasons than they post on Twitter and again for different reasons than they share something on Instagram. Therefore, no two data sources are the same, and it is important to consider the universe we want to analyse data from.
- **What hypotheses and expectations can we already formulate?** This is good practice for any research project but especially for such open-ended

datasets. Sometimes, you will already know what you want, and it will be better to structure the data around the needs (e.g. you have certain attributes that you want to track). At other times, you will be in 'discovery' mode. If you identify this early on, the approach will be better tailored to get to your desired outcome.

- **How much depth do we want?** Like any research approach, you can dive in more deeply or keep it in the shallows. This applies to both AI-led and human-led elements of the project. Deeper will get more insight but it will also likely take more time and more money.
- **How will we use the data?** This is a more technical, but important, question which will tell you what is legal based on GDPR. It is important to consider this. Different companies take different approaches to interpreting what is publicly available data and what is not. Ensure the policies of any organisations you retain align with your own policies.

These are all factors we guide you on in order to ensure a successful research programme. Social intelligence and analytics have a great deal to offer as research tools, and they can also represent a very useful complement to other sources of insight, ensuring a 360° understanding of consumers and citizens' behaviours and expectations. In promoting pure bottom-up approaches and leveraging cutting edge AI, we believe that social intelligence and analytics are truly changing the nature of market research.

REFERENCES

1. https://books.google.com/ngrams/info
2. https://www.ipsos.com/en/social-listening-tools-synthesio
3. https://www.ipsos.com/en-us/knowledge/consumer-shopper/Ipsos-Pandemic-Adaptability-Continuum
4. https://www.ipsos.com/en/bigger-innovations-require-bigger-data

FROM UNSTRUCTURED DATA TO INTELLIGENCE

Leendert de Voogd Global Service Line Leader, Social Intelligence Analytics, Ipsos

Tara Beard-Knowland Country Service Line Leader, Social Intelligence Analytics, Ipsos MORI

GAME CHANGERS