WINTER 2022 / ISSUE 13

CREST SECURITY REVIEW

ISSN 2398-0540 (PRINT)
ISSN 2398-0559 (ONLINE)

Technology

CONTENTS

Highlights

CONVERGING SECURITY

Without evidence and guidance, organisations seeking to adopt convergence may be setting themselves up for failure – p18

CHINA'S DIGITAL DIPLOMACY

Social media platforms are places where China, amongst many other states, are seeking to increase reach and influence watching publics around the world – p20

TECHNOLOGY

CREST SECURITY REVIEW
Editor – Rebecca Stevens
Guest Editors – Prof Stacey Conchie and Dr Matthew Francis
Illustrator & designer – Rebecca Stevens
To contact *CREST Security Review* email
csr@crestresearch.ac.uk

PAST ISSUES

To download (or read online) this issue, as well as past issues of *CREST Security Review*, scan the QR code or visit our website:
crestresearch.ac.uk/magazine

FROM THE EDITOR

Emerging technology can give us a unique opportunity to anticipate, plan, counter, and respond to security threats as we move forward in an ever-changing world. But that same advancing technology can also be weaponised against us.

This issue of *CREST Security Review* highlights the new opportunities and problems that advances in technology bring through a behavioural and social science lens.

Looking at the issues surrounding data, David Buil-Gil (page 4) and his team explain how measurement error in crime data affects crime prevention. Emma Boakes (page 14) explains why cyber and physical security should collaborate to prevent new vulnerabilities in technologies that can be exploited through a cyber-attack.

We rely on biometric information, such as face, fingerprint, and voice, to provide a strong and permanent link between an individual and their identity. Yet this information can become compromised as Sophie Nightingale discusses face-morphing in the task of detecting identity fraud (page 6). Discussing the ethics and issues that surround biometrics, Ian D presses on the importance of law enforcement and researchers collaborating to develop an understanding of new biometrics (page 20).

On pages 16-19 we have a special new feature; a lightning piece consisting of four short articles focusing on specific studies in technology. 'Who are you? Your tech can tell you', as Oli Buckley analyses the way you type and Heather Shaw talks about the digital footprint in everyone's pocket (pages 16-17). On page 18, Leon Reicherts considers how chatbots could be designed to prompt us to stop and think, while Richard Philpot and Mark Levine analyse CCTV footage to better understand how people behave during dangerous emergencies (page 19).

How human analysts use and interact with AI systems is a recurring theme of technology discussions. Marion Oswald (page 8) explains why the success of machine learning depends on empowered people while Christopher Baber (page 26) explores the question of explanation in human interaction with AI systems.

We get the message: Carl Miller reports on what was found when bespoke algorithms analysed over 100,000 messages posted by Chinese diplomatic social media accounts (page 22), and as Isabelle Van Der Vegt, Bennett Kleinberg, and Paul Gill demonstrate on page 12, large-scale linguistic analysis may help security practitioners in making sense of violent and extreme communications.

As always, this issue includes some articles outside of the special topic: Erin Grace and Gina Ligon give us the low-down on NCITE; America's latest centre to lead on research tackling extremism and terrorism (page 28), Paul Gill and Zoe Marchment outline the results of a process evaluation of the Vulnerability Assessment Framework (VAF) on page 30, and Shanon Shah studies 'dark fandoms' to give lessons on how *not* to make a violent copycat (page 32).

You can find all the research that underpins these articles and some further reading in the 'Read More' section on page 34. Please let me know what you liked (or didn't) about this issue and what you would like to see featured in future issues. Write to me at *b.stevens@lancaster.ac.uk*

Rebecca Stevens
Editor, *CSR*

DAVID BUIL-GIL, JOSE PINA-SÁNCHEZ, IAN BRUNTON-SMITH & ALEXANDRU CERNAT

BAD DATA, WORSE PREDICTIONS

How measurement error in crime data affects crime prevention.

Increasingly sophisticated methods are applied to predict crime patterns from police records with little regard given to the quality of the data. We explore how this may affect crime prevention.

MEASUREMENT ERROR IN CRIME DATA

Police-recorded crime statistics are widely used for different purposes. Police forces use crime data to analyse geographic variations in crime and assist operational decisions. Policy makers draw on police records to justify policy changes. Neighbourhood watch groups use crime data to lobby for security. Crime researchers examine recorded crime trends to build theories of crime. But police-recorded crime data are severely affected by measurement error.

Crime records are incomplete. Not all victims report crimes to the police. And the police do not record all crimes reported. Many incidents, such as drug offences, do not even have direct victims who can inform the police. All this results in what is known as the 'dark figure of crime' - i.e., crimes not recorded in statistics. The percentage of crimes unknown to the police can be as large as 67% for damage, 63% for personal property offences and 61% for threats. Crime reporting also varies across population groups and recording practices differ between police forces, with a 2014 inspection concluding that between 63 and 71% of violent incidents and between 71 and 77% of sexual crimes reported to the police were not correctly recorded in crime registers.

Due to the ever-growing evidence that crime records are inaccurate, such records had the official designation of 'UK National Statistics' removed in 2014. Yet, crime aggregates are still used for crime prevention. Here we explore how crime prevention may be flawed by an uncritical acceptance of police-recorded data.

IMPACTS ON CRIME PREVENTION

We describe a set of ways in which police records are used in crime prevention efforts, and explore how these may be affected by underlying measurement error present in crime data.

Geographic crime analysis

Crime statistics are aggregated in geographic areas and visualised in maps to study the spatial concentration of crime and identify areas where crime is most prevalent. Geographic crime analysis assists the design of targeted crime control initiatives. For geographic crime analysis to accurately highlight high-crime areas, the proportion of crimes unknown to police forces should be spatially uniform. This is not the case. The 'dark figure of crime' varies across cities and neighbourhoods. Consequently, the police may underestimate crime in places with low reporting rates and overestimate it in places with higher reporting rates.

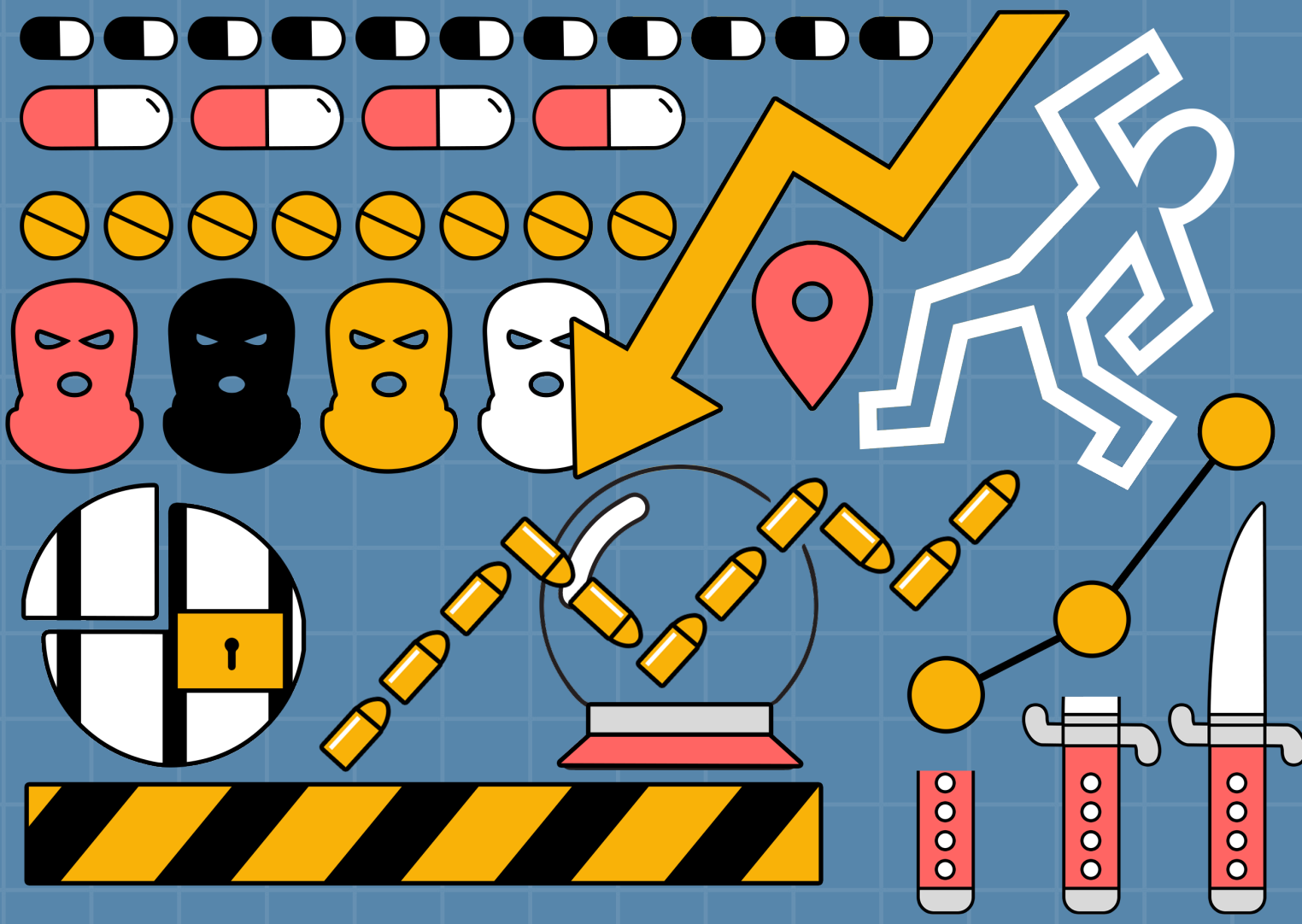
> **...crime forecasts suffer from high risk of inaccuracy and may result in disproportionate police control on historically over-policed communities.**

Crime trends analysis

Crime data are used to analyse changes in crime over time to identify whether crime is increasing and why. For crime trend analysis to accurately reflect changes in crime, the proportion of crimes missing from police records should remain stable across time. This is not always the case. Crime records are affected by changes in the way data are recorded, and crime reporting rates vary across years. In turn, trends recorded in police records and crime surveys vary significantly.

Predictive policing

Law enforcement makes use of predictive analytics to identify potential criminal activity before it takes place, to determine targets of police interventions. Historical crime data is used to train machine learning algorithms to forecast incidents. For predictive policing to accurately forecast crime, data used to train algorithms should not be affected by measurement error (or algorithms should account for it). This is rarely the case. Crime forecasts suffer from a high risk of inaccuracy and may result in disproportionate police control on historically over-policed communities.



Exploring the causes of crime

Crime records are used by researchers to explore the causes of crime. Statistical modelling is used to estimate the effect of a range of social, legal and environmental constructs on crime to assess if the presence of certain social conditions, policies or urban features is causing crime to increase. For statistical models to accurately estimate the effect of a variable on crime, it is necessary that crime data are not affected by measurement error. Crime researchers rarely account for this. Error induced by underreporting, underrecording and random errors (i.e., the 'dark figure of crime') may bias model estimates of the impact of security measures, economic conditions, disorder and other variables on crime.

A WAY FORWARD

The biasing effect of measurement error on crime research and crime prevention is widely recognised - but there are ways forward. Victim surveys record periodical data from randomly selected samples of respondents and provide relevant information about crimes known and unknown to police. Matching survey data with police records allows us to identify the prevalence of measurement error in crime records, and researchers are using this information to identify its potential effects on statistical outputs and to generate adjusted crime estimates. For instance, it enables identification of the geographic and temporal variation of the 'dark figure of crime', and accounts for the proportion of crimes missing from police records in statistical analyses, crime forecasts and crime prevention. We are also using it to test potential methodological solutions to mitigate the biasing effect of measurement error on model results. Simulation studies show that, in many cases, this problem can be minimised, or altogether eliminated by log-transforming crime rates. Of course, victim surveys are not error-free. This is why we compare and combine multiple crime data sources to enable crime estimates of improved precision for crime prevention.

Dr David Buil-Gil is a lecturer in Quantitative Criminology at the Department of Criminology of the University of Manchester, and a member of the Manchester Centre for Digital Trust and Society.

Dr Jose Pina-Sánchez is an associate professor in Quantitative Criminology at the School of Law of the University of Leeds.

Prof Ian Brunton-Smith is a professor of Criminology and Research Methods at the Department of Sociology of the University of Surrey.

Dr Alexandru Cernat is an associate professor in Social Statistics at the School of Social Sciences of the University of Manchester.

SOPHIE NIGHTINGALE

IDENTITY FRAUD IN THE DIGITAL AGE

Advances in technology are bringing new problems to the task of detecting identity fraud, including the relatively new phenomenon of face-morphing and the synthesis of facial images.

We rely on biometric information, such as face, fingerprint, and voice, to provide a strong and permanent link between an individual and their identity. Yet this information can become compromised — sometimes unintentionally, but other times as part of an attempt to steal another person's identity. Identity fraud is a societal problem that presents a significant threat to national security. Although not a new problem, technological advances allow for increasingly sophisticated means to commit identity fraud.

Consider how a known criminal on a government watch list might attempt to travel into a different country undetected. In the past they might have created a fake passport or had a similar-looking accomplice (who is legally able to travel) submit a renewal application using their photo. Now, the fraudster can rely on the relatively new phenomenon of face-morphing. Morphing enables a fraudster to digitally combine their face with that of their accomplice in a single image. The morphed image is submitted with the accomplice's passport application. If successful, the fraudster is issued a fraudulently obtained but genuine (FOG) passport; a real document that will bypass any counterfeit detection measures in place. It is now the task of the border control officials or automatic face recognition systems to detect the manipulated photo.

In applied settings, such as border security, it is important to accurately match faces of individuals unfamiliar to us with their photo-ID, yet people show surprisingly poor unfamiliar face-matching performance (Bruce et al., 1999). The morphing technique further complicates matters because the morphed image contains some of the fraudster's facial features thus making the ID-checkers' task even more difficult. Worryingly, there is growing evidence suggesting that accurate detection of these so-called 'morphing attacks' is limited, especially for high-quality morphs, and that training attempts have little effect on accuracy (Kramer et al., 2019; Nightingale et al., 2021; Robertson et al., 2017, 2018). Face recognition systems have also been shown to be vulnerable to morphing attacks (Nightingale et al., 2021; Scherhag et al., 2019).

> **...the morphed image contains some of the fraudster's facial features making the ID-checkers' task even more difficult.**

HOW CAN WE IMPROVE PEOPLE'S ABILITY TO DETECT FACE-MORPHING?

Borrowing from facial recognition studies, it has been shown that expert forensic facial examiners and untrained super-recognisers achieve higher accuracy on challenging face identification tasks than members of the general public (Phillips et al., 2018). Research has also shown that adopting a feature-by-feature comparison strategy can improve unfamiliar face-matching (Towler et al. 2017). These lines of research suggest that human face 'specialists' might show greater accuracy in morph detection tasks, and a featural rather than holistic approach might translate to improved accuracy in the task of face morph detection. These two possibilities remain to be tested.

Another possible solution is to modify the passport-issuance process. Researchers have suggested that the best solution to the face morphing problem is to have government officials acquire photos at the place of issuance (Ferrara et al., 2014). This live enrolment approach is already used in some countries and has recently been implemented in others in response to the threat

Adapted from @Art Huntington / Unsplash.com

of face morphing. Although this approach would solve the problem of digital face morphing, it still does not deal with the issue of physical identity fraud techniques, such as the use of hyperrealistic silicon masks (Robertson et al., 2020).

CAN ARTIFICIAL INTELLIGENCE HELP?

The successful application of machine learning to develop an algorithm that can accurately discriminate morphed faces from real faces remains somewhat limited, in part because of the manual effort required to generate a high-quality landmark-based morph. Therefore, training sets typically consist of relatively small numbers of morphs.

One potential way to improve the capability of machine learning for detecting morphs is to draw on a popular artificial intelligence (AI) mechanism for synthesising content: generative adversarial networks (GANs) (e.g., MorGAN; Damer et al., 2018). A GAN consists of two neural networks — a generator and a discriminator — that are pitted against one another in a game-like scenario. The generator's task is to synthesise a facial image that the discriminator accepts as 'real'. The discriminator's task is to distinguish between real faces and those synthesised by the generator. Initially, the generator produces a random array of pixels and passes this to the discriminator. If the image is distinguishable from a real face, then the generator is penalised and over many iterations, it learns to synthesise increasingly realistic faces until the discriminator is no longer able to distinguish the synthetic from the real faces. Of course, this ability to synthesise content (so-called deep fakes) brings a unique set of threats to society (Nightingale & Farid, 2022); however, it also provides the infrastructure to produce high-quality face morphs at scale and, in turn, generate a substantial training set that could be used to train human facial examiners and to improve the accuracy of computational classification of morphed and non-morphed faces.

THE ARMS RACE CONTINUES

Rapid advances in technology continue to make it easier than ever to create sophisticated and compelling fakes. Although in a technological sense, these advances are exciting and an incredible achievement, inevitably there will be individuals who use these developments for harm. Therefore, we must work to keep these threats at bay.

Dr Sophie Nightingale is a lecturer in Psychology at Lancaster University. Her research examines the challenges and opportunities posed by digital technology, especially in relation to security and forensic identification processes.

MARION OSWALD

'GIVE ME A PING, VASILI. ONE PING ONLY'

WHY THE SUCCESS OF MACHINE LEARNING DEPENDS ON EMPOWERED PEOPLE

While we're seeing some promising developments in the introduction of machine learning and data science methods to support law enforcement risk assessments, we shouldn't assume our technology is answering the question we need to answer.

The quote in the title is from a 1990 Cold War film 'The Hunt for Red October.' Sean Connery plays Captain Marko Ramius, commander of the Soviet Union's newest submarine, which is fitted with an innovative propulsion system undetectable to passive sonar. As Captain Ramius and his officers want to defect to the United States, the story features a race between American and Soviet submarines to detect the Red October. The Americans need to make contact with it before the Russians find and sink it. Captain Ramius's famous '*Give me a ping, Vasili*' comes as the talented sonar officer Jonesy attempts to track the Red October using his underwater acoustics software.

Jonesy does not take the output of the software at face value. He knew they did not originally build it for tracking nuclear submarines but for detecting seismic anomalies. He used this knowledge to interpret the result in the complex situation and was supported by his commander because he was able to explain and justify his findings.

One moral of this story, which applies to today's preoccupation with data analytics, machine learning, and AI, is: **Don't assume your technology is answering the question you need to answer.**

To uphold this moral, we need to understand what AI tools are doing and the immediate and longer-term consequences of using them within our decision-making processes. Epstein argues that:

> 'In a truly open-world problem devoid of rigid rules and reams of perfect historical data, AI has been disastrous… When narrow specialization is combined with an unkind domain, the human tendency to rely on experience of familiar patterns can backfire horribly.'
>
> (Epstein, 2019)

Perhaps this might seem a touch harsh when set against claims made for some predictive techniques and diagnostic tools that appear almost daily. Here, I unpack what such predictive use cases are really doing.

WHAT ARE 'PREDICTIVE' USE CASES REALLY DOING?

Much is written about predictive risk assessments using machine learning methods, often based around random forest decision trees. But are they really predicting or risk assessing anything? They use group data from the past to make a prediction about an individual's future. It's more accurate to say they are categorising or comparing by comparison with certain characteristics of a specified group in the past and these characteristics will only be those that can be translated into a datapoint or a numeric scale. The question these tools are really answering is how do the characteristics of an individual (which can be translated into a datapoint or a numeric scale) compare to the characteristics (as translated into datapoints) of a specified group of people in the past.

If we want to understand and evaluate a tool, we need to know details like: what input data is being used and how has it been translated into datapoints? Are these data relevant to the question I need to answer? What is the analysis doing with these datapoints? And what uncertainties and provisos are attached to the analysis?

We know that in many public sector contexts, recorded data can be partial, entered in different formats, out of date or missing. For example, a BBC report on Greater Manchester Police's Integrated Operational Policing System quoted one

This article is a version of a keynote address given to the RUSI-Trustworthy Autonomous Systems conference 'Trusting Machines? Cross-sector lessons from healthcare and security' on 1 July 2021.

serving officer's concerns that 'there's a black hole where the recent intelligence should be.' If machine learning methods are implemented without a deep understanding of the underlying data, the impact of errors and missing information could be both amplified and hidden from the user.

PREDICTIVE USE CASES — DO THEY 'WORK'?

Law enforcement is increasingly expected to adopt a preventative, rather than reactive posture, with greater emphasis on anticipating potential harm before it occurs, identifying vulnerable individuals in need of safeguarding, and targeting interventions towards the highest-risk persistent and prolific offenders. Actuarial risk assessments have been used for many years to support such a preventative approach; what's new is the introduction of machine learning and data science methods to produce the algorithm and ever-increasing types and volumes of datasets.

High accuracy rates at the group level can often conceal very low accuracy rates for specific individuals or groups of individuals within that larger group.

High accuracy rates at the group level can often conceal very low accuracy rates for specific individuals or groups of individuals within that larger group. All individual predictions are associated with a confidence interval (a margin of error), which is often not taken into account when reporting the overall predictive accuracy of the tool (Babuta and Oswald, 2019).

To quote one of my favourite fictional detectives:

> 'While the individual man is an insoluble puzzle, in the aggregate he becomes a mathematical certainty. You can, for example, never foretell what any one man will do, but you can say with precision what an average number will be up to. Individuals vary, but percentages remain constant.'
>
> (Arthur Conan Doyle, 1890)

> **What question can the tool contribute to answering and is this the question we need to answer?**

Examples in health also illustrate the importance of validation and contextualisation of the AI output by an expert human. While AI-supported breast cancer risk prediction has produced promising results, researchers have highlighted the need for improvements based on additional clinical risk factors, closer consideration of strategic screening aims (early detection, reduction of false positives and so on) and validation on diverse patient populations and clinical environments. What question can the tool contribute to answering, and is this the question we need to answer?

An evaluation of a sepsis detection algorithm by academics at the University of Michigan claims that the particular tool has poor predictive value despite its widespread adoption in clinical settings. The research suggests that the tool does not catch patients at an earlier stage of sepsis (which is when you would want to catch them from a clinical point of view) and therefore does not do what its manufacturers state that it does (Wong et al., 2021).

DECISION-MAKING AUTHORITY

There's another reason why we should ask **whether the technology is answering the question we need to answer.**

Decisions in national security, policing and health are subject to important legal tests, including those set out in the human rights framework and in specific laws governing coercive or intrusive powers. There is a risk of relinquishing decision-making authority if we conflate algorithmic outputs with the answer to a legal test (Oswald, 2018).

Let's take the requirement for 'reasonable grounds for suspicion' to justify the exercise of police powers. According to Code A pursuant to the Police and Criminal Evidence Act 1984, 'generalisations or stereotypical images that certain groups or categories of people are more likely to be involved in criminal activity' cannot support reasonable suspicion. Probabilistic outputs based on reference class may not satisfy the requirement for reasonable grounds, as they fall within the exclusions of generalisations, category-based suspicion, and suspicion based on general association.

We've seen that algorithmic predictions effectively compare an individual against datapoints from a group in the past, and so are likely to be seen as equivalent to suspicion based on general association, as set out in code A of PACE.

All this is not to say that data analytics have no place in national security, policing, and healthcare — far from it (Oswald, 2020). We're seeing some very promising methods being developed to join the dots between different pieces of information to suggest connections between those involved in organised crime or previously unidentified crimes of modern slavery. In national security and policing terms, such analysis is a form of intelligence and therefore should be assessed and handled as such, with its potential uncertainties appreciated.

THE NAMING OF ALGORITHMS

As noted above, an algorithm might predict an average behaviour, but for an individual (especially when the algorithm's output could be used to 'do something' in the real world that might affect that individual's rights), badging something as predictive is potentially misleading and risks creating over-reliance. We should name these algorithms in a way that accurately describes what they do in a more specific and circumspect way, e.g., as an 'Organised Crime Group Association Suggester' or 'Public Order Deployment Suggester'.

RECOMMENDATIONS

I conclude by returning to Jonesy and Commander Mancuso and expanding on the recommendations that flow from their stories. We should:

- Ask what the tool was built to do.
- Ask what the tool is really telling us — question the headline.
- Ask what the tool is NOT telling us and what is missing or uncertain.
- Ask whether the output of the tool is relevant to the decision that needs to be taken.

Mancuso and his reaction to Jonesy is equally important in this story as it tells us the following about AI and empowered people:

- Operators and managers need appropriate training, knowledge and skills to understand AI tools.
- Skilled operators need discretion to decide how, if at all, to use the output, provided that they can justify their decision, and management should be supportive of the exercise of skilled discretion.
- Management should take a critical approach both to how AI works and the purposes for which it is proposed to be used.

Dr Marion Oswald works at Northumbria University and the Alan Turing Institute. Her research focuses on law, ethics, technology, policing and national security. She sits on the Advisory Board of the Centre for Data Ethics and Innovation.

ISABELLE VAN DER VEGT, BENNETT KLEINBERG & PAUL GILL

LINGUISTIC THREAT ASSESSMENT: CHALLENGES AND OPPORTUNITIES

Large-scale linguistic analysis may help security practitioners in making sense of violent and extreme communications.

Threat assessment generally involves the process of gathering information after a threat has been made to understand the risk of violence posed by a person. Usually, a threat will have been uttered in the form of verbal or written language. Nowadays, security professionals are confronted with assessing violent and extreme language on a large scale online. In light of these developments, we have been examining the application of computational linguistics to the study of grievance-fuelled targeted violence, including terrorism and mass murder. We call this approach 'linguistic threat assessment', in which our focus lies upon its computational implementation. This article highlights our main findings and the challenges and opportunities of this approach.

LINGUISTIC AREAS OF INTEREST

In our application of computational linguistics methods to the understanding of grievance-fuelled communications, we interviewed thirteen threat assessment professionals (with an average 18 years of experience) about their approach to anonymous threatening communications. Participants all read the same anonymous threat letter and subsequently discussed how they would assess the case. Although practice differed greatly between professionals — such as the cues paid attention to and the conclusions drawn from it — the responses in which linguistic information was used for assessment could be summarised as belonging to one of three areas of language, namely:

1. **Linguistic content:** *what* are people writing, i.e., in terms of word frequencies.
2. **Linguistic style:** *how* are people writing, i.e., in terms of grammar.
3. **Linguistic trajectories:** how does content and style develop over time.

Consequently, we leveraged these different areas of language for the study of grievance-fuelled communications. For example, we examined linguistic style in a study on abuse directed at politicians to infer gender gender, age, and personality traits based on language use in written abuse. Although we discovered some interesting gender and personality differences in the way participants wrote, the error margins for determining these traits based on language use alone were large, which means actionable predictions are difficult.

COMPUTATIONAL LINGUISTICS

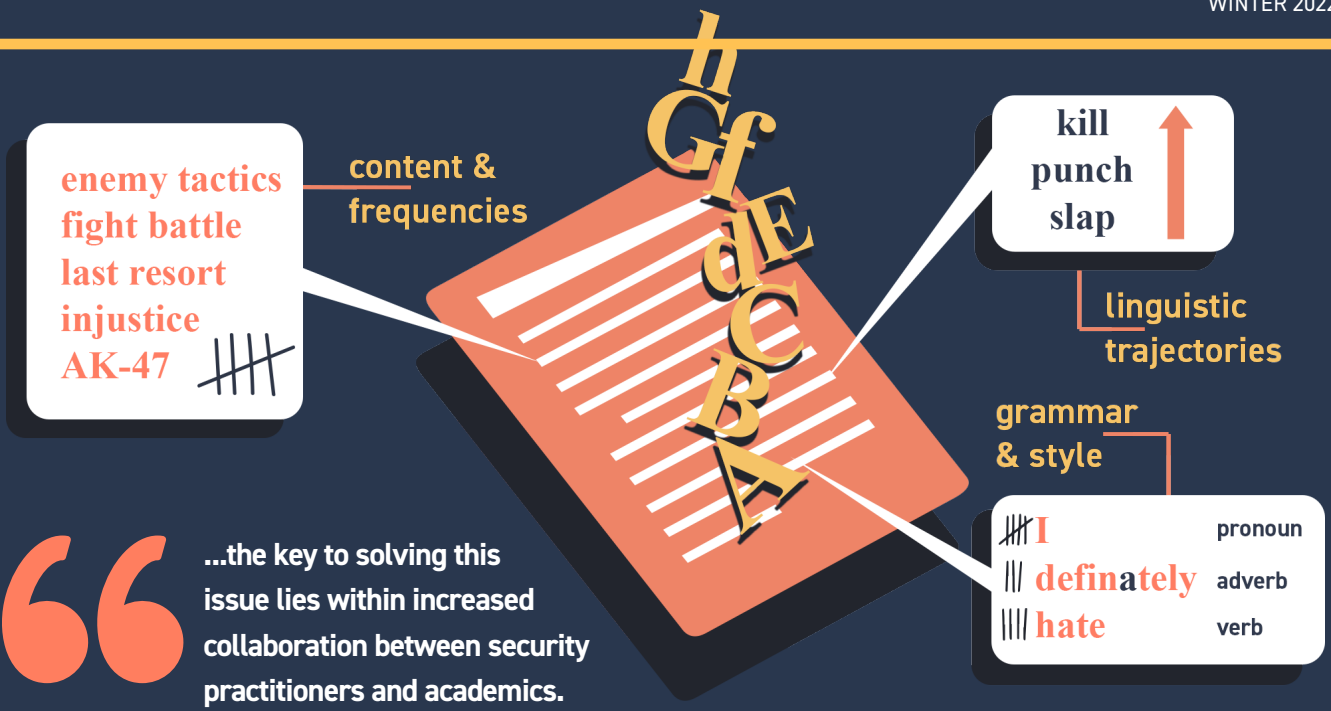
The branch of linguistics in which the techniques of computer science are applied to the analysis and synthesis of language and speech. *Oxford English Dictionary.*

We have also demonstrated the utility of measuring language over time (i.e., linguistic trajectories) to assess the effects of external events on an extremist group and the evolution of language on a far-right forum. Of particular relevance to security professionals is perhaps our study on the development of the 'Grievance Dictionary', which puts emphasis on the linguistic content (i.e., *what* someone conveyed).

THE GRIEVANCE DICTIONARY

The Grievance Dictionary is a tool specifically developed to analyse grievance-fuelled and/or threatening language at scale. It makes use of word frequencies to measure different (psychological) concepts in text. It is similar to the LIWC dictionary, which can measure a wide variety of psychological (e.g., friendship, sadness) and linguistic concepts (e.g., pronouns, swear words), but is specifically focussed on grievance-fuelled communications. Again, we started with consulting expert threat assessors (similar sample as stated above) and asked what they look for in a text when they assess a potential threat of violence.

From that expert exercise, we established 22 categories that make up the Grievance Dictionary, which includes categories such as weapons, murder, desperation, and planning. Next, we generated wordlists representative for each category and tested their validity using an online rating task, in which 2,318 participants on crowdsourcing platform Prolific assessed the 'goodness of fit' of 20,502 words for these categories. In applying the dictionary



> **...the key to solving this issue lies within increased collaboration between security practitioners and academics.**

to measure the aforementioned 22 concepts, we saw marked differences between different text samples. For instance, we saw that lone-actor terrorist texts scored higher on all but one measure (especially murder, soldier, and weaponry) when compared to right-wing extremist forum posts. The only category on which these samples did not differ, was our measure of loneliness.

These first analyses using the Grievance Dictionary demonstrate how it can be used to analyse large volumes of text, for instance in the case of a lengthy manifesto or an entire forum. In essence, these large volumes of text are condensed down into 22 comprehensible measures that are relevant to security professionals or researchers dealing with grievance-fueled violence. These measures can subsequently be integrated into a broader assessment of an individual or group of individuals, or can be used for research purposes in which different types of authors (e.g., different ideologies, violent vs. non-violent) are compared on Grievance Dictionary measures.

CHALLENGES AND OPPORTUNITIES

One challenging issue within the field of linguistic threat assessment is access to data. Targeted violence is a low base rate phenomenon, and the number of cases where the perpetrator produced linguistic material related to an incident will be even smaller. It is common procedure within this field to make use of lone-actor terrorist manifestos to better understand violent language use, as it is known these authors committed an act of violence. However, the sample size of lone-actor terrorist manifestos is small (our database counts approximately 25).

These manifestos are often compared to a larger sample of neutral, non-violent texts to assess linguistic differences. In doing so one of the main questions within this field remains unanswered, (which is what we are perhaps most interested in discovering), namely, which linguistic markers set apart a violent text written by an individual with violent intent, from an individual without such intent. That is, we want to know what — linguistically — sets apart the actualisers from the non-actualisers. Are there specific Grievance Dictionary categories that significantly differ between these groups? At present, we do not know because we do not have the data to study these questions.

When using extremist forum data, we simply do not know whether the individuals behind a post were in fact violence actualisers or not. In other words, the ground truth behind the data is not available to us. One notable recent initiative includes the use of a former extremist in order to identify the violent from the non-violent extremists on a forum. However, apart from this one paper, we believe the key to solving this issue lies within increased collaboration between security practitioners and academics . We expect that police or security practitioner databases contain a multitude of communications, which were initially seen as violent or extreme, and subsequently did or did not lead to violence.

Linguistic analysis of such data will be incredibly valuable for our understanding of (possible) links between violent language and behaviour. By sharing data, we can continue to increase our understanding of violent language and thereby further the field of linguistic threat assessment.

Dr Isabelle van der Vegt is an honorary research associate at the Department of Security and Crime Science at University College London and a scientific project manager at the Research and Documentation Centre for the Dutch Ministry of Justice and Security.

Bennett Kleinberg is an assistant professor at the Department of Methodology and Statistics at Tilburg University and an honorary associate professor at the Department of Security and Crime Science at University College London.

Paul Gill is Professor of Security & Crime Science at University College London.

EMMA BOAKES

CONVERGING SECURITY

Why cyber and physical security should collaborate, and what it takes to achieve this.

Organisations are increasingly reliant on internet-based technologies for physical assets such as building management systems, internet of things (IoT) devices and operational technology. Such technologies create new vulnerabilities that can be exploited through a cyber-attack. Indeed, the number of attacks where a vulnerability in cyber security has been used to target physical systems or vice versa, have been increasing (Symantec, 2019). Looking specifically at IoT devices, a 600% increase in attacks was reported in 2017 (Symantec, 2018). Gartner's *Predicts 2020* report highlighted "...incidents in the digital world have an effect in the physical world, as risks, threat and vulnerabilities now exist in a bidirectional cyber-physical spectrum".

To understand and mitigate threats that cross the boundary between what is cyber and what is physical, some organisations have integrated their security resources to encourage them to work more closely together. While intuitively it makes sense for security functions to converge, to date there has been little evidence to support this. Indeed, there remains a lack of guidance on how to effectively implement converged security. Without evidence and guidance, organisations seeking to adopt convergence may be setting themselves up for failure and even be implementing new structures and processes that will allow new vulnerabilities to emerge. Research is needed to build an evidence-base that will help organisations make informed decisions when deciding how to implement convergence.

> **Without evidence and guidance, organisations seeking to adopt convergence may be setting themselves up for failure...**

My research aims to provide such an evidence base. Structuring my research around evidence-based practice (Briner, 2019), I carried out three qualitative studies with security staff from a range of organisations and industries that operate converged security from around the world:

1. I conducted interviews with five senior security experts who have experience implementing convergence to start to identify a web of interconnected factors that support the implementation and operation of convergence.
2. I carried out a three-round Delphi study with a panel of 23 security professionals to validate the factors identified in the first study and to rate them on their importance for effective convergence.
3. Finally, 15 senior staff involved in the decision to converge in their respective organisations were interviewed using an epistolary interview technique (i.e., interviews using a series of written communications), carried out over email. These interviews identified how organisations decided to adopt converged security and the process and activities they used to design its implementation.

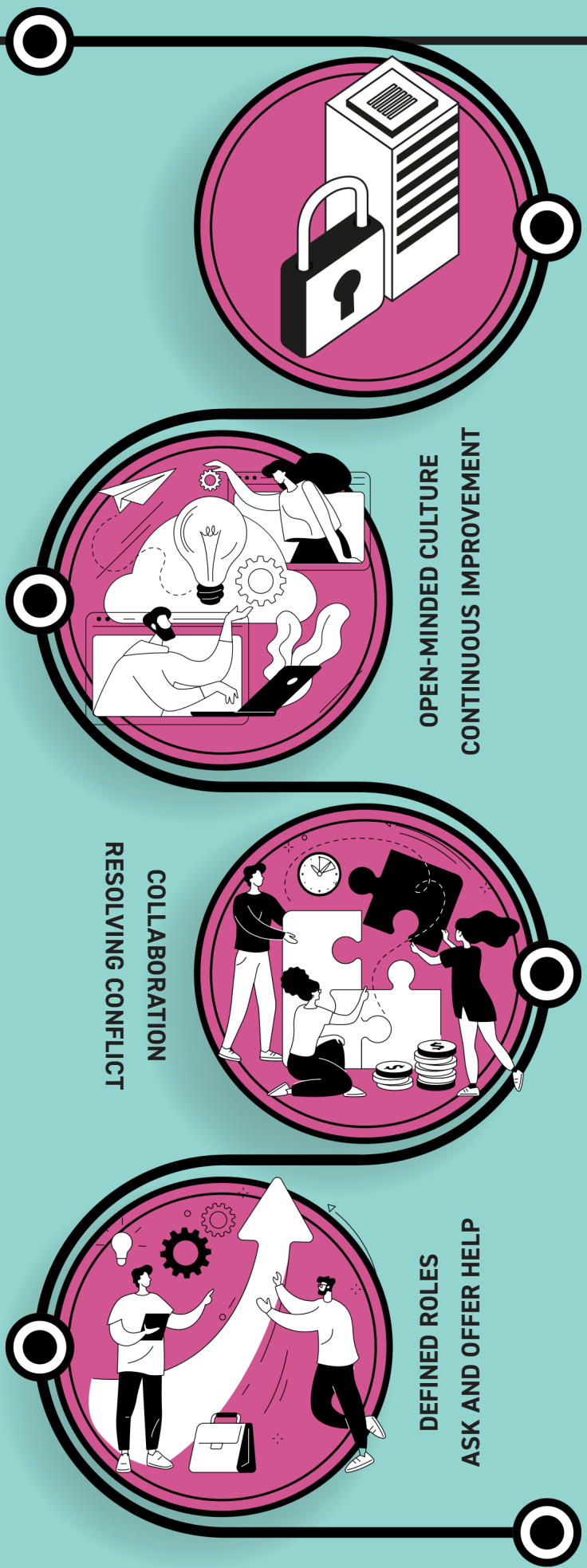
ESTABLISHING CONVERGED SECURITY

My research established that organisations adopt convergence in an effort to:

- Manage risk in the changing threat environment.
- Reduce complexity across the security function.
- Improve efficiency and make cost savings.

Convergence is often instigated by the insights of key security personnel but is also influenced by other organisations, government and industry associations.

The decision to adopt convergence is only one element of the decision-making process. My results showed that organisations have different ways of implementing convergence as it is dependent on organisational context. To achieve an appropriate and workable implementation of convergence, organisations



> **Convergence requires facilitation and active management to engage staff in the appropriate collaboration.**

need to draw on insights from within their security functions and consult with staff to capitalise on their first-hand experience of security in context.

ACHIEVING COLLABORATION

The establishment of organisational structures that bring security resources together under a common management and with a common goal are not enough to ensure convergence. Convergence requires facilitation and active management to engage staff in the appropriate collaboration. My research found that convergence relies on a web of interconnected factors, and to achieve collaboration, organisations need to cultivate each of these building blocks:

- Organisations need to foster a culture within security that encourages staff to be open-minded, promoting continuous improvement, setting the precedent that security will develop over time.
- Staff need to buy-in to the idea of collaboration, and management can play an active role in enabling this, from reviewing progress to resolving conflict.
- Collaboration is facilitated by staff having clearly defined roles and responsibilities. They need to be provided with opportunities to engage with each other formally and informally to help build working relationships and to enable them to ask for and offer help from each other.

WHAT DOES THIS RESEARCH MEAN?

The final stage of the research will be to use these findings to generate an evidence-based roadmap. The roadmap will specify the design decision organisations need to make when adopting convergence, and will help them identify the different sources of information they can use to inform those decisions.

The roadmap will also indicate the range of factors that organisations will need to consider to support effective convergence. The roadmap will, therefore, provide organisations with an evidence-based guide that helps them navigate the adoption and implementation of convergence in their context.

Emma Boakes is a final year PhD student at the University of Portsmouth. Her research explores security convergence.

Image credit: Adapted from *Gstudio*, *alestraza* and *Visual Generatio* | stock.adobe.com

OLI BUCKLEY

IT'S NOT WHAT YOU TYPED, IT'S THE WAY YOU TYPED IT...

Typing patterns can predict a user's name and native language.

The way we type says a lot about who we are, with the rhythm and cadence of our keystrokes as identifiable as our handwriting or signature. However, it doesn't stop there. Keystroke Dynamics — the study of typing patterns, enables researchers to identify characteristics about the person at the keyboard. This includes things such as handedness, hand size, mood or typing style.

Our work, *Collecting and Leveraging Identity Cues using Keystroke Analysis* (CLICKA), evolved this idea of user identification to derive personal characteristics unique to the individual. This work focused on determining the name and native language of an anonymous user, based solely on *how* rather than *what* they typed.

The first experiment centred on determining the name of an anonymous user by collecting typing samples from 84 users. Participants completed several typing exercises, where the timing of each keypress and release was recorded. The research hypothesised that a user would type a familiar combination of keys more quickly. As such, the data were subdivided into short phrases containing two characters (bigrams) and ranked according to their typing speed.

> **The name prediction achieved a balanced accuracy of 70% of the bigrams in a user's name.**

The second experiment used a similar approach to determine the native language of an individual. Here, 492 participants were recruited from five native languages (English, French, German, Italian and Spanish), with an event split across each group.

The research used machine learning classifiers to develop models capable of predicting both a user's name and native language. The name prediction achieved a balanced accuracy of 70% of the bigrams in a user's name. Native language prediction achieved a balanced accuracy of 71% when comparing English against everything else. When predicting based on all five language categories, the accuracy dropped to 45% — still considerably better than a random prediction.

The key takeaway of this project is that it is possible to predict identifying characteristics about a user based on their typing patterns. This often requires a small sample of data, with participants only typing 200 to 300 words.

Dr Oli Buckley is an associate professor in Cyber Security, School of Computing Sciences at the University of East Anglia.

HEATHER SHAW

THE IDENTITY IN EVERYONE'S POCKET

When people interact with their smartphones, the digital traces left behind can be used to infer their identity.

Around a quarter of an adult's daily behaviour is spent on their smartphone (Ellis et al., 2019; Shaw et al., 2020). As such, smartphone usage data can reveal important insights into a person's daily habits and can be used to infer their identity.

> **It was possible to find within a top-10 list, the person whom the application usage data belonged to 75% of the time.**

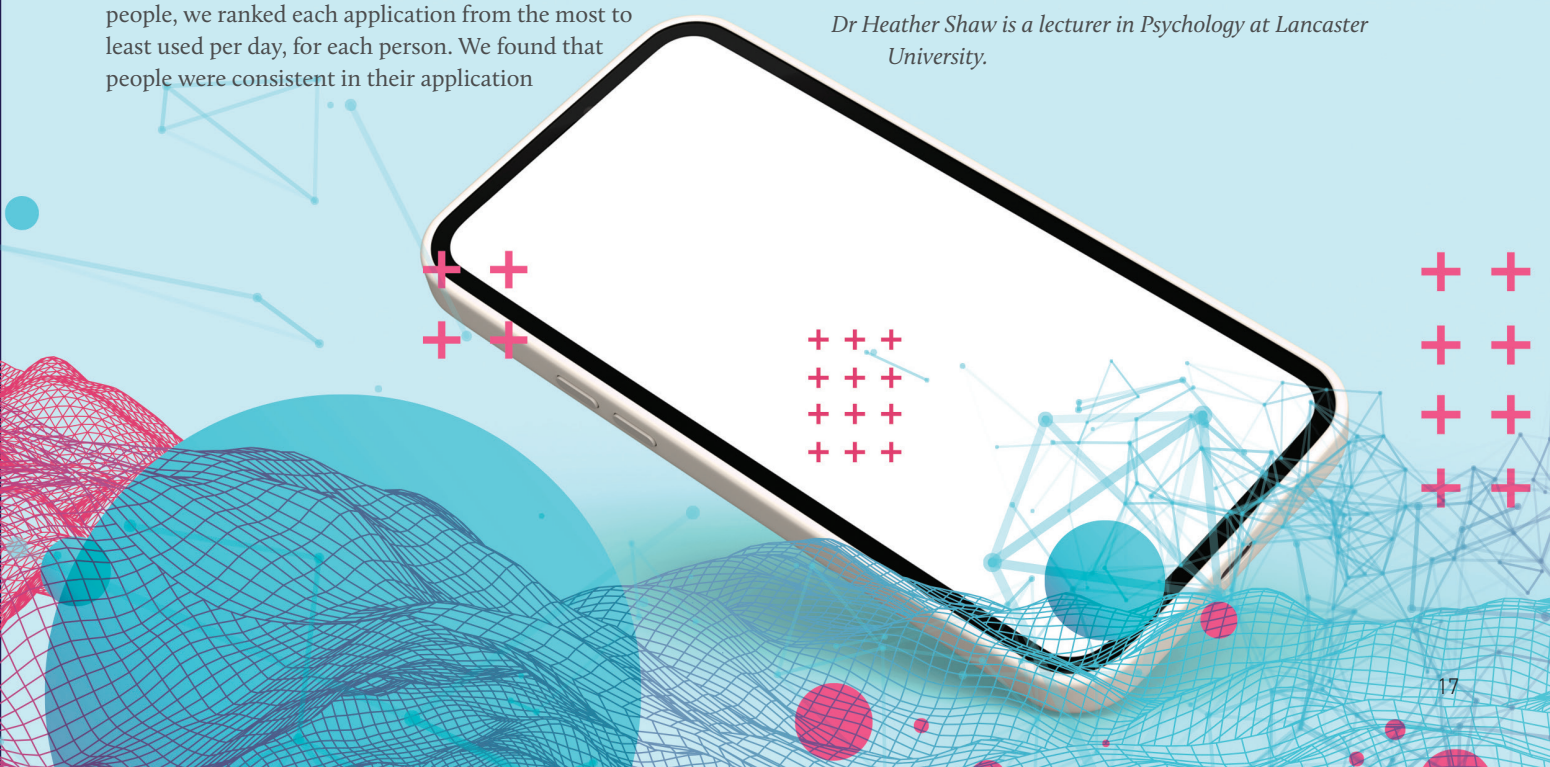
In our study of 28,692 days of smartphone data usage from 780 people, we ranked each application from the most to least used per day, for each person. We found that people were consistent in their application usage patterns on a day-to-day basis (e.g., consistently used Facebook the most and the calculator application the least). When we examined two randomly selected days from the same person, we found greater similarity in application use patterns than when we randomly selected two days that belonged to two different people.

To explore if application use could identify a single person, we fed 4,680 days of application usage data (equating to 6 days per person) into machine learning models. The models learned people's usage habits from the 6 days of application data and then tried to predict a person's identity when presented with an anonymous seventh day of data. The model was able to identify the correct person one-third of the time. Daily smartphone use can therefore act as a digital fingerprint.

The results further showed that it was possible to find within a top-10 list, the person to whom the application usage data belonged 75% of the time. In practical terms, this means that an investigation seeking to find a criminal's new phone from knowledge of their historic phone usage could reduce a pool of ~1,000 people's phones to 10 phones, with a 25% risk of missing them.

Our results suggest that access to smartphone application use data allows for a reasonable prediction about a person's identity even when they are logged-out of their account. This identification is possible with no monitoring of the conversations or behaviours within the applications themselves. Therefore, it is important to acknowledge that application usage data alone could risk our privacy if it is misused. It also questions whether usage data should be protected in the same manner as other personal identifiers.

Dr Heather Shaw is a lecturer in Psychology at Lancaster University.



LEON REICHERTS

"OK GOOGLE, SHOULD I CLICK ON THAT EMAIL?"

Designing conversational user interfaces to make us stop and think.

In recent years, data analytics tools have been given new features that enable users to query complex datasets using typed or spoken natural language. Instead of having to learn and use complex query syntax, analysts can now ask questions directly 'to' the data. Research has shown how these new ways to interact with data can improve both the user experience and task efficiency. However, central to data analysis is also knowing *what* to ask and coming up with *meaningful* questions. How can the next generation of analytics tools help users to generate more meaningful questions? This is where chatbots and voice assistants (sometimes referred to as 'conversational agents') can really come into their own, by being programmed to probe users to scaffold their questioning when using data analysis tools.

In our research group, we have begun researching how to augment human cognition by having an agent embedded in the software to proactively prompt users when looking at different data visualisations. We have found that agent prompts — even simple ones — can shift the users' attention to aspects of the data they would have missed or overlooked. It can also help them generate more exploratory questions.

> **One such interface protects against phishing attacks by helping users think more about suspicious emails.**

Our next steps are to find out whether this proactive agent approach supports more extensive data analysis and decision making in various contexts. We want to test whether such agents may, to some extent, mitigate challenges such as overconfidence or confirmation bias. We are also exploring how conversational agents can be designed to get people to 'slow down and think' when they are about to make risky decisions. Such an interface protects against phishing attacks by helping users think more about suspicious emails, enabling them to examine specific aspects of the email before deciding whether to click the potentially harmful URL.

This line of research suggests there are new opportunities for extending the reach of chatbots and conversational agents beyond their current home; instead of answering users' queries they can instead question them, encouraging people to think in new ways.

Leon Reicherts is a PhD student at the UCL Interaction Centre and part of the Ecological Brain DTP (ecologicalbrain.org). His research focuses on how to design conversational interfaces to support human cognition when performing complex data analysis and decision-making tasks.

RICHARD PHILPOT & MARK LEVINE

CCTV ANALYSIS OF VIOLENT EMERGENCIES

Systematic analysis of CCTV footage of violent and dangerous emergencies can help us understand how people behave during times of heightened security threats.

Whether it is incidents of street violence or marauding terrorist attacks, the fact that these events are invariably captured on public space CCTV means we can build a robust evidence base about behaviour in real-life emergencies.

However, CCTV data can be complex, incomplete and lacking both sound and wider contextual information. One way to get around this (and in doing so, extract the most reliable evidence from the CCTV data) is building an appropriate ethogram — a list of relevant behaviours in a particular context.

Using an ethogram approach we analysed CCTV footage of street violence in the UK, the Netherlands, and South Africa and were able to show that contrary to conventional wisdom, bystanders intervened in more than 90% of aggressive public incidents.

These tended to be coordinated interventions, with three to four bystanders working together to calm the violence.

> **...contrary to conventional wisdom, bystanders intervened in more than 90% of aggressive public incidents.**

Bystanders were also at low risk of victimisation when intervening to help (Liebst et al., 2021).

In another micro-behavioural analysis of CCTV footage of an explosion in a single railway carriage (Philpot & Levine, 2021), we also showed how emergency response behaviours can be shaped by the actions of immediate others — but that the behaviours themselves can be different in different places. Proximity to the explosion site is seemingly less important than the behaviour of the people around you.

The kinds of analysis that can be done is often shaped by data availability. It's not always possible to collect data systematically, and access to CCTV footage from some incidents might be limited by ethical, legal or security concerns.

The strength of analysing CCTV data is that it not only provides a richer understanding of behaviour in emergencies (compared to research which uses self report methods), it also allows us to test the assumptions of existing models that underpin emergency preparedness. As more footage becomes available, we will continue to develop important new insights that improve security and resilience planning.

Dr Richard Philpot is a lecturer of Psychology at Lancaster University. Applying digital data, his research examines how citizens and emergency services behave and interact during spontaneous public space emergencies.

Mark Levine is a professor of Social Psychology at Lancaster University. His research explores the role of identities and group processes in pro-social and anti-social behaviour.

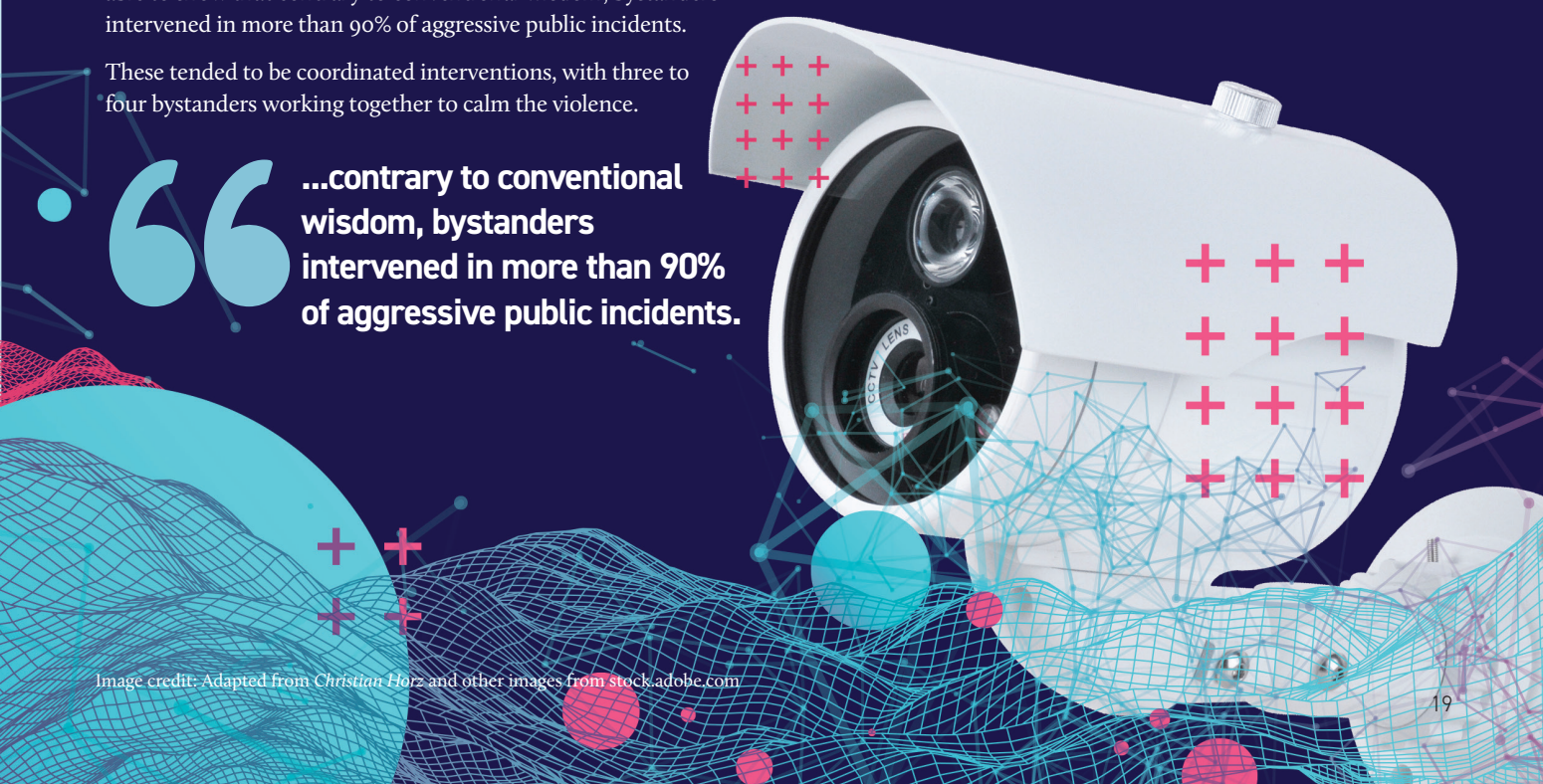


Image credit: Adapted from *Christian Horz* and other images from stock.adobe.com

IAN D

MAPPING A NEW BIOMETRICS LANDSCAPE

The development of new biometrics often stretches the ability of law enforcement organisations to train, test and apply these approaches to their work, let alone understand the ethical and scientific debates about their use and application.

Biometrics in the form of fingerprint and DNA are a mainstream element in law enforcement activity. As technology has moved forward, new biometrics have started to emerge in a wide range of areas such as face, voice, environmental geography, and other digital personal data. Invariably each involves capturing biologically derived data and the creation of a binary model against which further datasets can be compared to help indicate identity.

All have huge potential, if used appropriately, to safeguard the public. Many of these techniques are already in use in the private sector in areas such as customer authentication for access to banking records and facial comparison to unlock mobile phones. The challenge is understanding how law enforcement can employ these technologies to safeguard the public in a way that is socially acceptable, ethically aligned and legally compliant, while also better understanding the penetration of new biometrics, both now and in the future, within the private sector.

The multitude of modalities of new biometrics, ranging from gait analysis to voice analytics and image detection, is significant. A first step in understanding the landscape is to have a clear understanding of the relative strengths, weaknesses, opportunities, and threats posed by the different applications as a basis for future planning.

In addition, it is vital to understand the ethical framework in which new biometrics operate. Only by doing so can decision-makers ensure they are fully adhering to protecting the social contract with the public in a manner which is proportionate and necessary, within the legal framework. Part of this will come from gaining an in-depth understanding, through behavioural science, of the public perception of exploitation and acceptance of new biometrics such as voice analytics and facial recognition. In a world where technology has a global application, differing cultural views on new biometrics and the legislative frameworks surrounding them also need to be considered.

Feeding directly into the ethical debate are the issues of bias in systems and the malign use of new biometrics. As a community, we in law enforcement, need to understand the perception of bias and how it can be addressed in a manner that commands confidence across the full range of stakeholders. Equally, it is crucial to acknowledge, explore, and understand how new biometrics can be used for malign purposes both by states and through organised crime. Doing so is essential to understanding the wider narrative around new biometrics as well as the implications in other areas such as officer safety and the development of HUMINT relationships.

> "...it is vital to understand the ethical framework in which new biometrics operate. Only by doing so can decision-makers ensure they are fully adhering to protecting the social contract with the public"

Unlike traditional biometrics, new biometrics produce an identification based on the balance of probabilities. Combining modalities has the potential to increase the accuracy of identification. Academic research will play an important role in understanding how new biometric modalities relate to each other now and in the future.

If law enforcement is to mobilise the opportunities generated by new biometrics effectively, we need to have an understanding of the skills and training necessary to do so. Currently, such an understanding is fragmented, in some areas non-existent and in others being driven by commercial, rather than law enforcement

considerations. Academia has a vital part to play in supporting the development of these skill sets. Even at this early stage, it is clear that the complexity of these technologies will require a paradigm shift. Training will need to be carefully developed to ensure it is flexible, relevant, and cost-effective in a multi-disciplinary environment. With the potential for new biometrics to be used in both the overt and covert arenas, the design will need to address the needs of multiple stakeholders.

As a first step in this journey, the National Crime Agency (NCA) has partnered with CREST and staff at Lancaster University. This project will mine academic research in a number of key areas which relate to new and emerging biometrics. While focusing primarily on behavioural science, this will include a diverse range of disciplines such as data analysis, jurisprudence, and the full range of social sciences.

The information provided will then allow the NCA and other partners to better understand the biometrics landscape and how it is evolving and consider how academic involvement can shape the journey going forward.

New biometrics have the potential to be formative in shaping the nature of law enforcement in the 21st Century. The work being done between law enforcement and researchers will be a formative strand in developing our understanding in vital areas such as public perception, malign exploitation, and creating a skilled workforce.

Ian D is a senior manager at the National Crime Agency, responsible for the application of new and emerging digital biometrics.

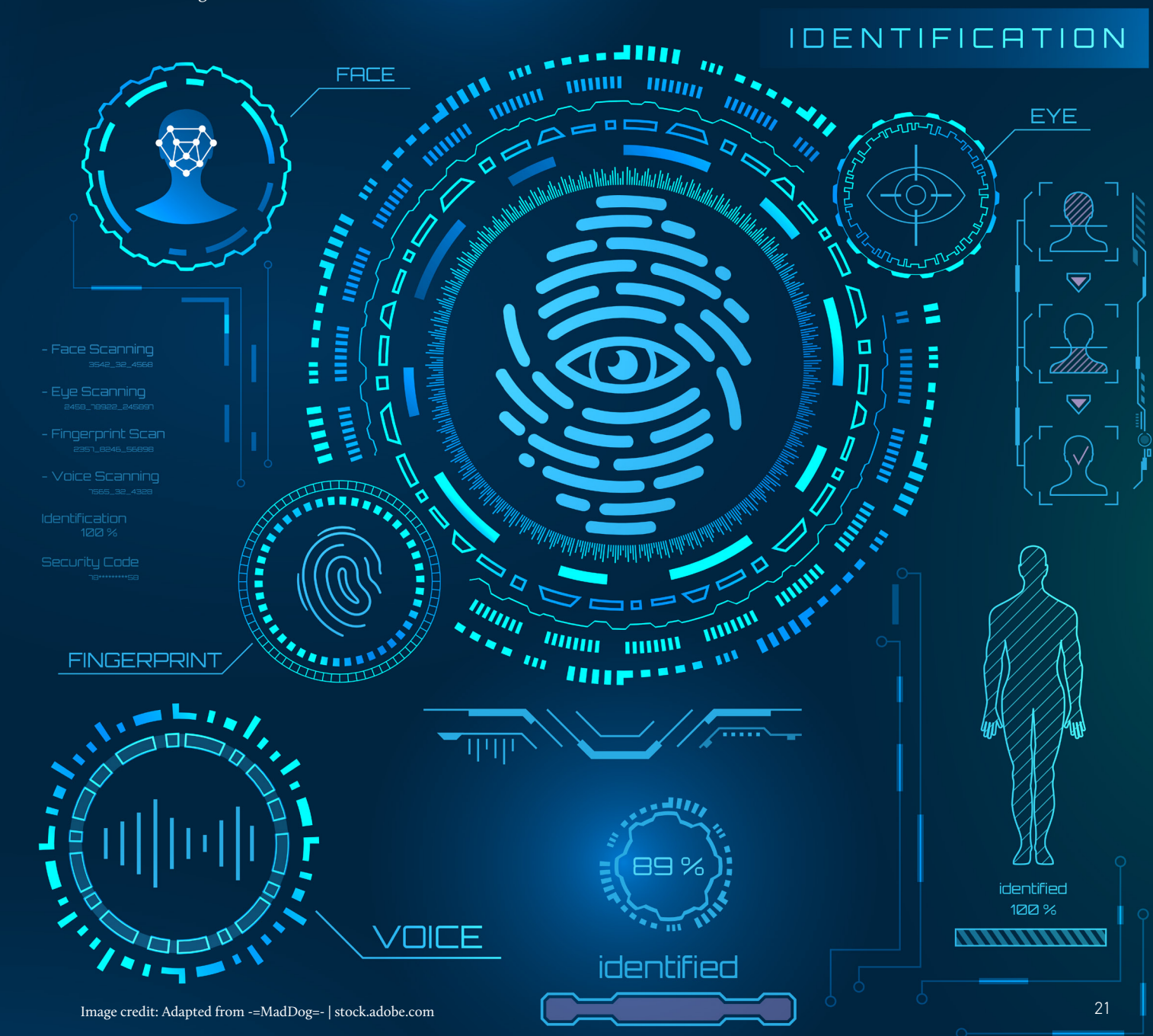


Image credit: Adapted from -=MadDog=- | stock.adobe.com

CARL MILLER

CHINA'S DIGITAL DIPLOMACY

Reading between the data lines...Carl Miller reports on what was found when bespoke algorithms analysed over 100,000 messages posted by Chinese diplomatic social media accounts.

Across an average week in 2021, hundreds of Chinese diplomatic voices made themselves heard online, posting thousands of messages and provoking hundreds of thousands of reactions, challenges, questions, re-shares and responses. It was often consul generals rather than their more senior ambassadorial colleagues that led the conversation, a new generation of digitally savvy — so-called 'Wolf Warrior' — diplomats, more assertively pushing back against foreign criticism of China.

At the beginning of last year, BBC Monitoring (BBCM) and the CASM Technology set out to study this Chinese public diplomacy as it was happening across social media platforms. The point was to combine BBCM's deep linguistic expertise with CASM's social media research technology to build a research system that was both linguistically and politically sensitive, but also able to operate across the vast expanses of data that social media platforms routinely create.

MULTI-LINGUAL MACHINE LEARNING

For six months, BBCM language teams worked with CASM's technologists and their artificial intelligence research environment (Method52) to train a system of bespoke algorithms that could automatically analyse the messaging of China's diplomatic accounts across the four languages they most often used: French, Arabic, Spanish and English.

Creating a unified framework of themes across all four languages proved to be a significant definitional challenge, requiring a great deal of iterative engagement between each of the four language teams involved. To ensure the framework was reliably applied across all researchers, a small booklet was eventually produced detailing the criteria for inclusion in any theme.

In total, 34 algorithms were trained, most specific to the languages and themes being studied. Narrower themes of more specific language (e.g., COVID-19) tended to be more amenable to rapid training while broader, more linguistically diverse themes (e.g., 'China's culture and people') posed more formidable challenges to the machine learning. Eventually, these models performed with an accuracy of around 80% overall, calculated by comparing classifier outcomes with those of a human on roughly 2,500 randomly selected Tweets and Facebook posts.

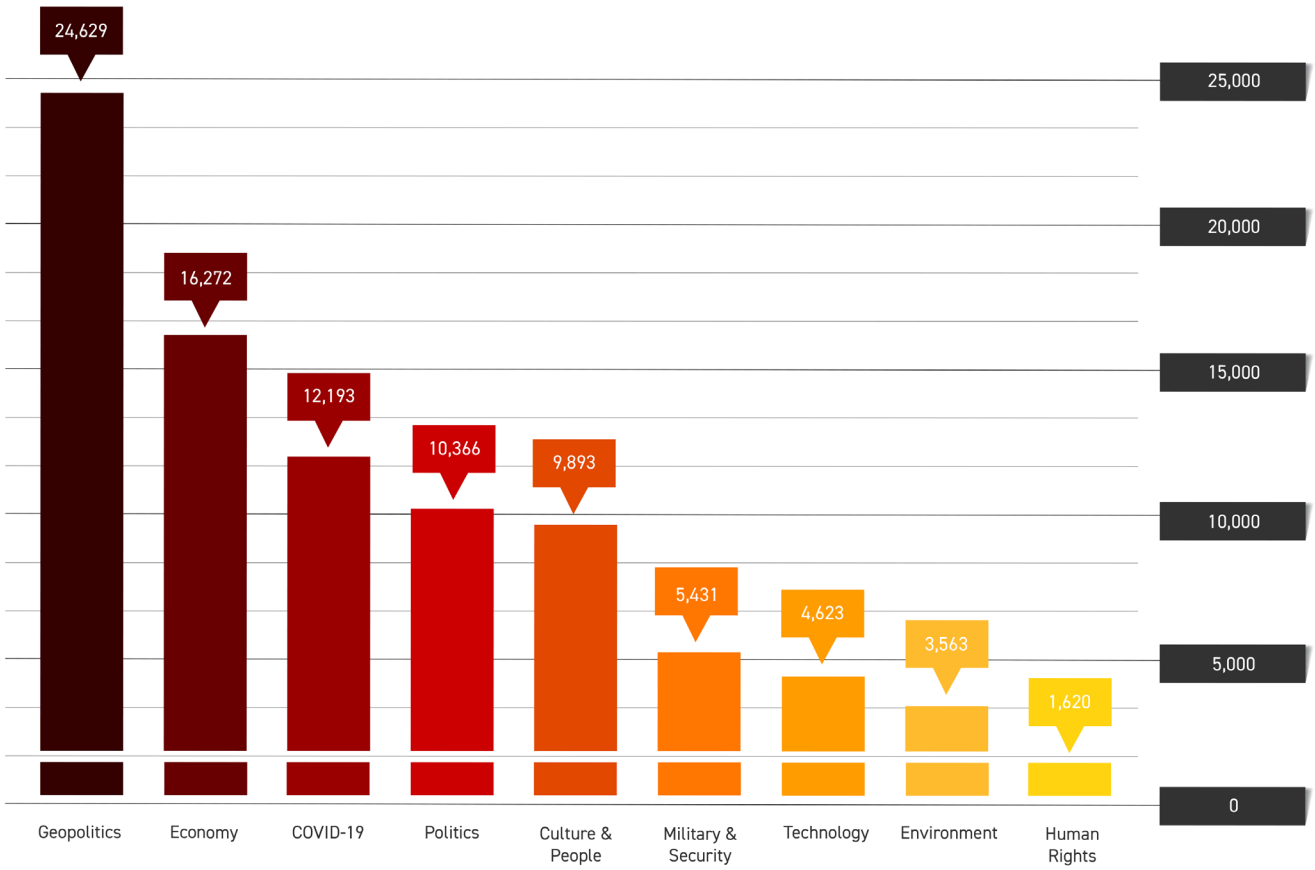
> **Social media platforms are places where China, amongst many other states, are seeking to increase reach and influence watching publics around the world.**

Here, we report on the output of this architecture: a window on over 100,000 separate messages sent by 393 Confucius Institutes, ambassadors, consular officials, and accounts from China's foreign ministry on Facebook and Twitter from the start of 2021 to the end of September that year.

The picture it paints is one of clear and reasonably stable global and strategic trends, but, as we'll see, also of important, sometimes dramatic, variation across time, theme, region and language.

GLOBAL PATTERNS

Beijing uses its network of diplomats around the world as the main way of getting its message out. Overall, they sent 102,883 messages which, generally, found an audience. In total, their messages (on Twitter) were retweeted 899,391 times across the period of study; an average of 12.4 times per Tweet. They also received a total of 5,883,361 'likes'; an average of 64.9 likes per message on Twitter and 38.7 on Facebook. Baselining this level of engagement is difficult because it is influenced by a number



Number of messages per theme

of factors; the followers of the messengers, the time when the messages are sent, the kind of messages that they are and the socio-culture mores of the audiences to the messages. However, it does represent a fairly significant audience in absolute terms.

THE THEME

Two thirds of China's messaging fell into one of the nine overall themes:

1. Geopolitics (26.3%)

More messaging was on Geopolitics than any other theme. This covered the factors, events and themes that governs and structures China's relationship with the world. This included any announcement, meeting or issue covering any of China's bilateral relationships. Especially key here was the China-USA relationship, Taiwan and Hong Kong.

2. The Economy (17.4%)

This second most popular theme included commentary regarding the production and consumption of goods and services and the supply of money as they relate to China. This covered Chinese economic development, reform, e-commerce, finance, taxation, marketing and advertising, transport infrastructure, trade, energy, mining, agriculture and industry. It also included specific economic programmes and projects, especially the Belt and Road Initiative.

3. COVID-19 (13%)

This included its impacts, countermeasures, vaccine development, controversy over the origin of its outbreak, 'COVID diplomacy' and its many social, political and economic implications.

4. Politics and Society (11.1%)

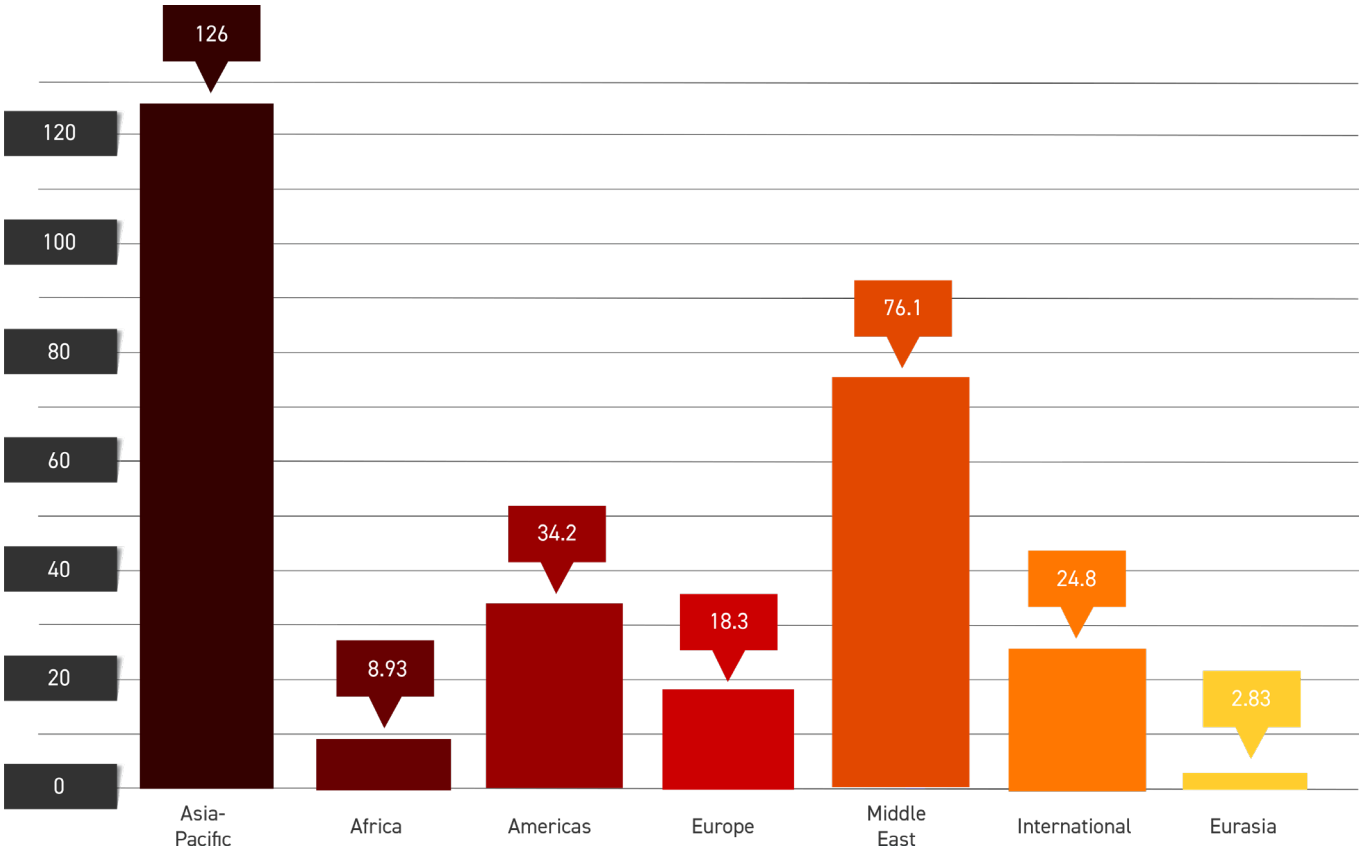
Messages in this theme were about how political power and influence are distributed and exercised to control, direct or influence events and the actions of people and officials in China. This included messaging related to the Chinese leadership within a domestic context, 'Xi Jinping Thought', corruption, crime, migration, welfare, protest and human rights. It specifically included treatment of ethnic and religious groups such as the Uyghurs of Xinjiang, and inhabitants of Tibet.

5. Chinese Culture and People (10.6%)

This was a broad theme that spanned China's culture(s), customs, its people(s), activities and events. This includes China's history, its 'food diplomacy', Chinese festivals, sports, the Olympics, Confucius Institutes, foreign students in China, its universities and educational exchanges, and outreach to the global Mandarin-speaking diaspora.

6. Military and Security (5.8%)

Messages related to the armed, intelligence or domestic security forces. This included armed forces modernisation, nuclear



Average likes received per message by region

weapons, robots, drones, cyber warfare, military exercises, defence diplomacy, policing and counter-terrorism operations.

7. Technology (4.9%)

This theme covered messaging specific technologies, especially the Internet and cyber-security, IP infringement, 5G, Huawei, space exploration, biotech and renewable energy. Also included in this theme were the discussions of technologies regarding security, national economic interest or space and military issues. Any technology related to COVID was excluded.

8. The Environment (3.8%)

The environment was discussed comparatively little, as only the eighth most popular theme. This covered anything relating to climate change, air pollution, environmental deterioration and responses to these challenges that could include policies, technological solutions, and changing attitudes.

9. Human Rights (1.73%)

The least common of the nine themes covered any messaging related to both domestic and international human rights. This included criticism of the West's human rights record, speech, religious and press freedoms, Xinjiang and Tibet, as well as State monitoring surveillance and the social credit system. The other messages tended to fall into a series of smaller and miscellaneous themes or were too event-specific to place into a broader category.

REGIONAL VARIATION

Within the global trends, there was a high degree of variation between China's diplomats based in different parts of the world and in different languages. We present these contrasts, below, as a series of synoptic regional profiles although the reader should note these are only based on our analysis of English, French, Arabic and Spanish and not any other language that accounts from each region might use.

1. Asia Pacific: Key region

The Asia Pacific was the key region where China's digital diplomats were based. It saw sharp increase in message volumes from mid-March onwards and then sustained higher message volumes for the rest of the period of study.

While Europe actually had more accounts than the Asia Pacific, messages from accounts in the Asia Pacific also saw on average around twice as many reposts and likes as messages from any other region, suggesting, perhaps, that China's most visible and influential online sources are disproportionately concentrated in this region.

2. Africa: Multi-lingual messaging about COVID-19

Africa was the only region to see significant volumes of messaging in all four languages and a quarter of all messages (over 23,000) were sent from accounts based there. The messaging tended towards COVID-19 as a theme, but attracted

extremely low levels of engagement with an average of just two reposts per message. Differing levels of engagement are explainable through a number of factors, including the specific visibilities of China's diplomats in the region, the prevailing socio-technological norms of the populations living there and any regional trends around technology use that act as a backdrop to all of the behaviours in this report.

3. The Americas: English and Spanish language soft power

In the Americas, China's diplomatic accounts tended to emphasise the soft power topics of the economy and China's people and its society, with the least concentration on geopolitics. More Spanish messages were sent from accounts in this region than anywhere else. In total, the region saw the third-highest number of messages and also the third-most level of engagement with those messages.

4. Europe: Multi-lingual soft power to a less engaged audience

The activity of China's accounts based in Europe were similar to the Americas. In Europe, too, there was a relative preference for soft power topics across English, French and Spanish. Perhaps the greatest distinction was found not in the messaging but the behaviour of the audience, with China's accounts based in Europe seeing roughly half the levels of engagement and amplification as those based in the Americas, an average of 3.46 reposts and 18.3 likes per message.

5. Middle East: Low number of messages provoking a larger response

The Middle East saw a relatively small number of messages that were highly engaged with, second only to the Asia-Pacific region on average. Naturally enough, this region saw more messaging in Arabic than any other, and also a pronounced emphasis on political themes, including sharp spikes of activity not seen in any other region, in March, April and July (detailed more fully in the report). This region posted the highest proportion of 'Human Rights', messaging, albeit still very low in absolute terms; 4.1% of the region's output, compared to an average across regions of 1.77%.

6. Xinjiang: Most commonly mentioned entity

Much like the themes themselves, the entities being mentioned by China's diplomats changed significantly over time.
The project used multi-lingual automated Named Entity Recognition technology to identify these entities — peoples, places or organisations — within the messages collected. Many were related to specific events; mentions of 'Xi Jinping', for instance, increased on at least three occasions in February, late April and July, with the latter occasion also met with an uncharacteristically high number of messages mentioning 'Beijing' and the 'Communist Party of China'. This coincided with the 100th anniversary of the founding of the Communist Party of China, an event we observe in greater detail below.

Strikingly, 'Xinjiang' — the Uyghur Autonomous Region — was the most commonly mentioned entity across five of our nine themes ('culture and people', 'the economy', 'geopolitics', 'human rights' and 'military and security'). Volumes of messages mentioning 'Xinjiang' saw a number of sharp increases throughout the report period, most prominently during the first half of 2021. These 'Xinjiang spikes' (as we call them) occurred across February/March, again in April and a third in late May. Each tended to represent vocal opposition from diplomats and embassies to the UK, US, Canada and EU's coordinated sanctions and blacklisting of several officials over alleged human rights abuses in Xinjiang.

CONCLUSION

Social media platforms are places where China, amongst many other states, are seeking to increase reach and influence watching publics around the world. They know that opinions and attitudes can be formed there, and that they are an opportunity to make their case, raise the issues considered to be priorities, and respond to the criticism and messaging of other states.

The consequences of this are, of course, possibly very wide-ranging. China's messaging matters for activists, journalists, and really any of the planners and the strategists who work on the vast variety of different issues and areas around the world that it touches. For the UK, analysis of this messaging can provide insight into the thinking and priorities of China, as well as the prospect for strategic and tactical counter-communications of their own.

Researching geopolitics must suit the modes that the phenomenon itself now takes and this collaboration was as much interested in the method and technology used by the research as the topic itself. It was an attempt to blend together powerful machine learning with human linguistic and subject matter expertise to create an approach that was both sensitive to language and context, but also capable of handling data scales far beyond those of a manual analyst. In doing so, the contribution we hope to make is of an empirical, data-driven system that can provide a window into the way in which governments and others are using social media platforms to project certain narratives and messages around the world.

Geopolitics and influence, perhaps even statecraft itself, is changing. And as it does so, the ways we understand, track, measure and evaluate these phenomena must be just as data-rich as the environments where they now so routinely play out. The full report will be published on BBC Monitoring's website.

Carl Miller is co-founder of CASM Technology, a team of technologists working to develop social media research methods. He is also the research director of the Centre for the Analysis of Social Media at Demos. Find him on Twitter: @carljackmiller

CHRIS BABER

WHY AI SYSTEMS NEED TO EXPLAIN THEMSELVES

Chris Baber and his team's work for CREST explores the question of 'explanation' in human interaction with Artificial Intelligence (AI) systems.

WHY ARE YOU TELLING ME THAT?

AI systems provide information based on complex algorithms and often massive collections of data. While explanations to help guide understanding of AI systems and the decisions they reach are necessary, explanation should not be solely about the algorithms and data that AI systems use. The point of explanation is not only *how* the decision was reached, but *why* the decision was reached, and what impact these decisions have on our beliefs and actions. Explanation should account for the consequences of the decision. As we suggest below, explanation as it relates to the *why* and the *consequence*, is too complex to be left to the developers of AI systems and instead should be achieved through supporting conversation between users and the AI system to negotiate what would make a useful answer to the question 'why are you telling me that?'

EXPLAINABLE AI

Our concept of explanation combines three elements:

1. Perception of the situation
2. Background knowledge
3. Definition of relevance (of a decision to the situation).

The perception that people and AI systems have of their immediate situation should not only relate to the data that are available but also the environment in which the analysis occurs or activity that occurs within the environment. From this, one can see that a human analyst would most likely 'know' more than the AI system in terms of wider, less tangible perceptions, just as the AI system would clearly 'know' more than the human in terms of the wealth of data available to it. For example, in medical applications, AI systems will outperform humans in the ability to scan millions of cases and discern patterns and associations — far more than a human physician (even a specialist in a particular branch of medicine) is likely to see over the course of their career. This is because contemporary AI systems continue to prove remarkably robust at solving well-defined problems, often achieving levels of performance that spectacularly outperform human counterparts, particularly in areas like board games or image classification. The definition of 'performance' here favours the AI system. However, in the medical arena, outcome is arguably more important, and here, comparison of the accuracy of diagnosis tends to show the human experts perform as well as AI systems.

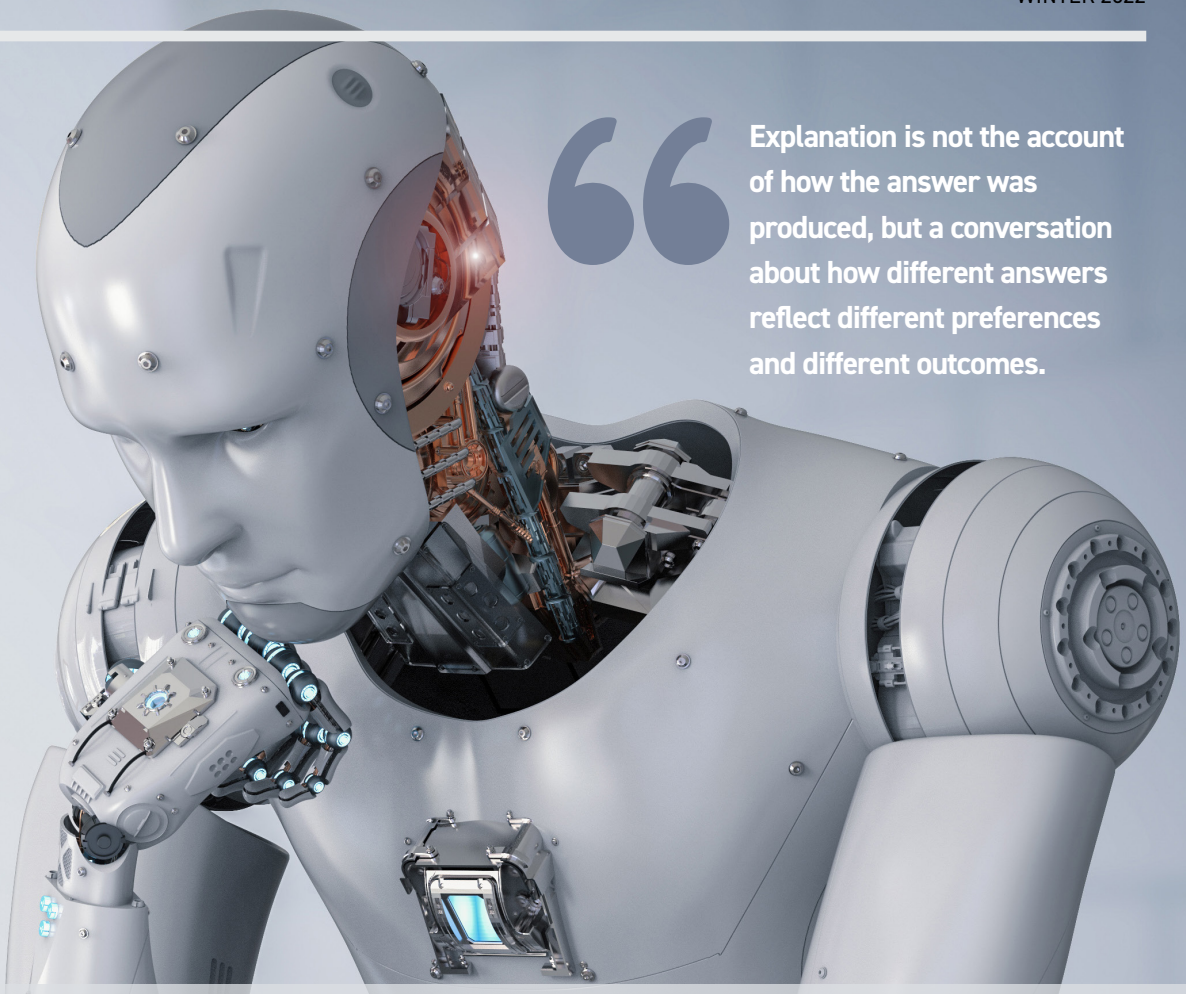
THE IMPORTANCE OF HYPOTHESIS TESTING

Where there are differences, these are not because the human is unable to produce a 'correct' (i.e., plausible for that situation) response, but because the AI system is often not able to juggle competing or ambiguous solutions. The experienced human physician can weigh up competing hypotheses, which lead to questions they ask the patient to seek other information. That is, the process of diagnosis involves the forming and testing of hypotheses through evidence collection informed by prior experience and expertise. We used AI tools (i.e., reinforcement learning) to model the use of information in human decision making and proposed that, in the absence of other sources, the optimal decision should accept the recommendation of an AI system only when its confidence exceeds 94%.

WHEN HUMANS INTERACT WITH AI

For human interaction with AI systems, differences in perception of the situation and background knowledge create different ways in which the conversation can be managed.

For example, recommender systems (which can suggest films, books, recipes, gifts, potential dates, etc.) assume that you and the AI system share the same interpretation of the situation (i.e., the criteria that define movies, such as genre) and the same definition of relevance (i.e., matching criteria to a list of recommendations, such as labelling the same movies as



> **Explanation is not the account of how the answer was produced, but a conversation about how different answers reflect different preferences and different outcomes.**

action-adventure). Any differences between what the AI system recommends can be easily handled by editing the criteria or rejecting the suggestions until you find one that you like. In this way, the conversation is not about agreeing with the answers but about agreeing on how best to define your taste.

If, for example, the AI system recommends you watch *Highlander 2*, then it (probably) has a definition of relevance that differs from yours. In this case, there are two broad options. The first is to adapt the AI system's definition of relevance to better match yours. However, the other is to 'nudge' you into adapting to the one that the AI system has decided is optimal. For the latter option, let's assume that the AI system is providing 'health' advice and decides that the choices you make (for food, alcohol, tobacco, or exercise) are not optimal. It might introduce goals, reminders, or instructions to encourage changes in behaviour.

For this to be successful, the AI system needs to have a correct model of an optimal outcome, and you, the user, need to accept that the solution is optimal. In all cases, the outcomes for you (i.e., an enjoyable film night or healthier lifestyle) are the more important explanation points, as opposed to the algorithms that got you there.

Stuart Russell, in his 2021 Reith Lecture on *Living with AI,* defined 'traditional AI' as seeking to optimise a decision in terms of given data and criteria, but posited that 'future AI' ought to be designed to appreciate that humans might not know the exact criteria for a 'correct' decision or their true objectives.

To shift from finding patterns in data to determining questions to ask, an AI system would need to change, so that the AI system is able to reason about its own reasoning and decision-making. Rather than blandly presenting an 'answer', AI systems ought to be able to discuss options available to their human users, with the AI system predicting the likely consequences of different options.

In this way, explanation is not the account of how the answer was produced, but a conversation about how different answers reflect different preferences and different outcomes. But the differences between how people and AI systems reach their decisions need not be as far removed as might be imagined.

Our work has shown that, for decisions which involve the selection and judgement of information, the strategy that a person uses can be modelled using AI algorithms and this suggests that it might be possible to find a common language through which AI systems and people are able to review and negotiate their decisions.

You can read more about this project at: crestresearch.ac.uk/projects/human-engagement-through-ai

Professor Chris Baber is Chair of Pervasive and Ubiquitous Computing at the University of Birmingham.

Image credit: *phonlamaiphoto* | stock.adobe.com

ERIN GRACE & GINA LIGON

NCITE

THE DESIGNATED COUNTER TERRORISM AND TARGETED VIOLENCE RESEARCH CENTRE FOR THE US DEPARTMENT OF HOMELAND SECURITY

NCITE conducts and shares research on the who, how, where, when, and why of terrorism and targeted violence that occurs inside the United States.

NCITE, or the National Counterterrorism Innovation, Technology, and Education Center has been the US Department of Homeland Security's (DHS) chosen academic partner for counter terrorism and targeted violence studies since 2020. NCITE conducts research and workforce development projects by leveraging interdisciplinary expertise across the social and technical sciences. NCITE has over 50 psychologists, sociologists, criminologists, political scientists, business and strategy professors, computer engineers and IT innovators focusing on the pressing case of violent extremism. These experts are drawn from 19 academic institutions in the US and UK, with a large number at the University of Nebraska Omaha (UNO). At NCITE HQ in UNO's Rod Rhoden Innovation Center, we have the largest number of dedicated PhD level extremist violence scholars of any academic terrorism centre in the US.

WHY DOES NCITE MATTER NOW?

Countering terrorism and targeted violence is always important, but especially so now. In its most recent Strategic Intelligence Assessment (May 2021), the US Federal Bureau of Investigation and DHS jointly said the greatest terrorism threat to the US is "posed by lone offenders, often radicalized online, who look to attack soft targets with easily accessible weapons".

> **"...the greatest terrorism threat to the US is "posed by lone offenders, often radicalized online, who look to attack soft targets with easily accessible weapons".**

Extremist violence is an especially relevant threat area as we see an increase in ideological-based violence, reflective of the rise in anti-government and anti-authority beliefs, racially and ethnically motivated attacks, and general civic destabilisation. The latter has been exacerbated by the ongoing coronavirus pandemic, deepening partisan divide, and an omnipresent online culture that elevates, accelerates, and mainstreams once-sidelined conspiracy theories and extreme beliefs. NCITE researchers are seeing a copycat effect in the way adherents of violent ideologies across the spectrum see, borrow, and use terrorism tactics and techniques from one another.

WHAT ARE THE MAJOR FOCI OF NCITE?

NCITE is focused on four thematic areas, on which scientific advancements are generated through annual grants provided by the DHS Science and Technology Office via our centre:

1. The Nature of Counter Terrorism and Targeted Violence Operations

Here we explore the nature of counter terrorism from two perspectives: 1) understanding tactics, ideologies, and connections of terrorists, and 2) equipping DHS's counter terrorism professionals with the knowledge and tools they need to anticipate emerging, novel threats.

2. Nationwide Suspicious Activity Reporting Initiative

Our focus is on strengthening the Nationwide Suspicious Activity Reporting Initiative: the formal tips reporting mechanism run by the federal government. Research looks both at the social barriers that prevent effective reporting and the technical innovations that make sorting complex information more efficient and precise.

3. Terrorism and Targeted Violence Prevention Program Evaluation

Work is focused on generating scientific evidence about the efficacy — and areas for improvement — for targeted violence and terrorism prevention programmes. Creative ways to evaluate the varied approaches to violence prevention is the main goal for this important and understudied area of research in terrorism studies.

4. Counter-Terrorism Workforce Development

We seek innovative workforce development research for the counter-terrorism community. The goal of this theme is to strengthen and professionalise the hardworking analysts, policymakers, and other members of the counter-terrorism workforce to ensure they are equipped with the latest training and technology to do their jobs and keep our communities safe.

HOW CAN YOU GET INVOLVED?

The best way to get involved is by signing up to our mailing list and attending our virtual and in-person events. We have several upcoming employment, scholarship, and fellowship opportunities to join us in Omaha! NCITE also run an annual call for funded projects and we welcome engagement with international academic and practitioner communities — check the website *www.unomaha.edu/ncite* for details. We look forward to welcoming these project teams to NCITE. Information on our previous funded teams can be found in our NCITE Year One Annual Report (see the Read More section).

WHAT IS NEXT FOR NCITE?

Our vision is to become the US's premier academic consortium for counter terrorism and targeted violence studies, now and beyond the 10-year duration of our DHS Center of Excellence designation. Picture NCITE as a place where an array of law enforcement, government agencies, non-profits, and corporate partners send their workers for professional development and where students across academic disciplines eagerly come for a unique opportunity to become part of the antidote to extremist violence. Picture Nebraska as a place leading the US from its centre, helping pull people in from the extremes to reduce violence, build resilience, and create a more stable future.

Professor Gina Ligon is the Director of the National Counterterrorism Innovation, Technology, and Education (NCITE) Center. She is also the Jack and Stephanie Koraleski Chair for Collaboration Science.

Erin Grace is the Strategic Communications Manager at the NCITE Center, and a former career journalist who believes in the power of storytelling.

PAUL GILL & ZOE MARCHMENT

EVALUATING THE CHANNEL PROGRAMME'S VULNERABILITY ASSESSMENT FRAMEWORK

Paul Gill and Zoe Marchment outline the results of a process evaluation of the Vulnerability Assessment Framework (VAF).

Since 2012, the UK government has used the Channel process to bring multiple agencies together to help prevent vulnerable people from being drawn into violent extremism. The VAF is an assessment guide used as part of the Channel process to identify an individual's vulnerability to becoming involved in (violent) extremism. Channel seeks to identify those at risk, assess the nature and extent of that risk, and develop suitable support plans to mitigate the risk. VAF assessments are required to inform decisions regarding whether and how to intervene with such individuals to prevent them from becoming radicalised and progressing further towards harmful behaviour.

Through a practitioner survey (n =181) and semi-structured interviews (n =13) we looked at the real-world use of VAF in existing risk assessment and management practice within Channel and developed a picture of:

1. Practitioner backgrounds.
2. Experiences of using, writing, and gaining information for the VAF.
3. The availability, utility and forms of training and guidance.
4. Potential improvements to be made.
5. Barriers to the risk assessment and management process.

THE RESULTS

Most survey participants agreed or strongly agreed that each of the VAF's 22 factors were useful for understanding the overall risk in most cases. However, respondents commonly expressed that the VAF needs to be more user friendly and could be condensed through reviewing and re-sorting risk factors. Respondents requested the inclusion of a summary conclusion section, a management plan section and a section dedicated to noting significant changes between VAF assessments.

Interviewees clearly and consistently expressed a need for an instrument to assist in decision-making. Various benefits include ensuring the assessor:

1. Does not miss crucial details.
2. Thinks of issues that did not immediately spring to mind.
3. Makes the results easier to digest by focusing the mind on three core areas.
4. Specifies why factors are irrelevant, thereby helping the bigger risk assessment.
5. Provides record-keeping and justification of actions conducted.

Respondents requested the inclusion of a summary conclusion section, a management plan section and a section dedicated to noting significant changes between VAF assessments.

The importance of training was evident. Some interviewees mentioned elements of the VAF are less applicable to those individuals with a mixed or unclear ideology, those who are non-aligned with a specific group (e.g., potential lone actors), and those interested in school shootings. Suggestions for standardised training included:

1. Practising filling out a real case and submitting the workings for feedback.
2. The practice of formulation and other fundamentals of risk assessment and management.
3. Greater focus on the factors and how to interpret them in different ideological contexts.
4. Demonstrations of good and poorly completed VAF assessments.
5. Refresher training.
6. Technical guidance on operating the relevant computer systems the VAF sits on and interacts with.
7. Overviews of available interventions to choose from.

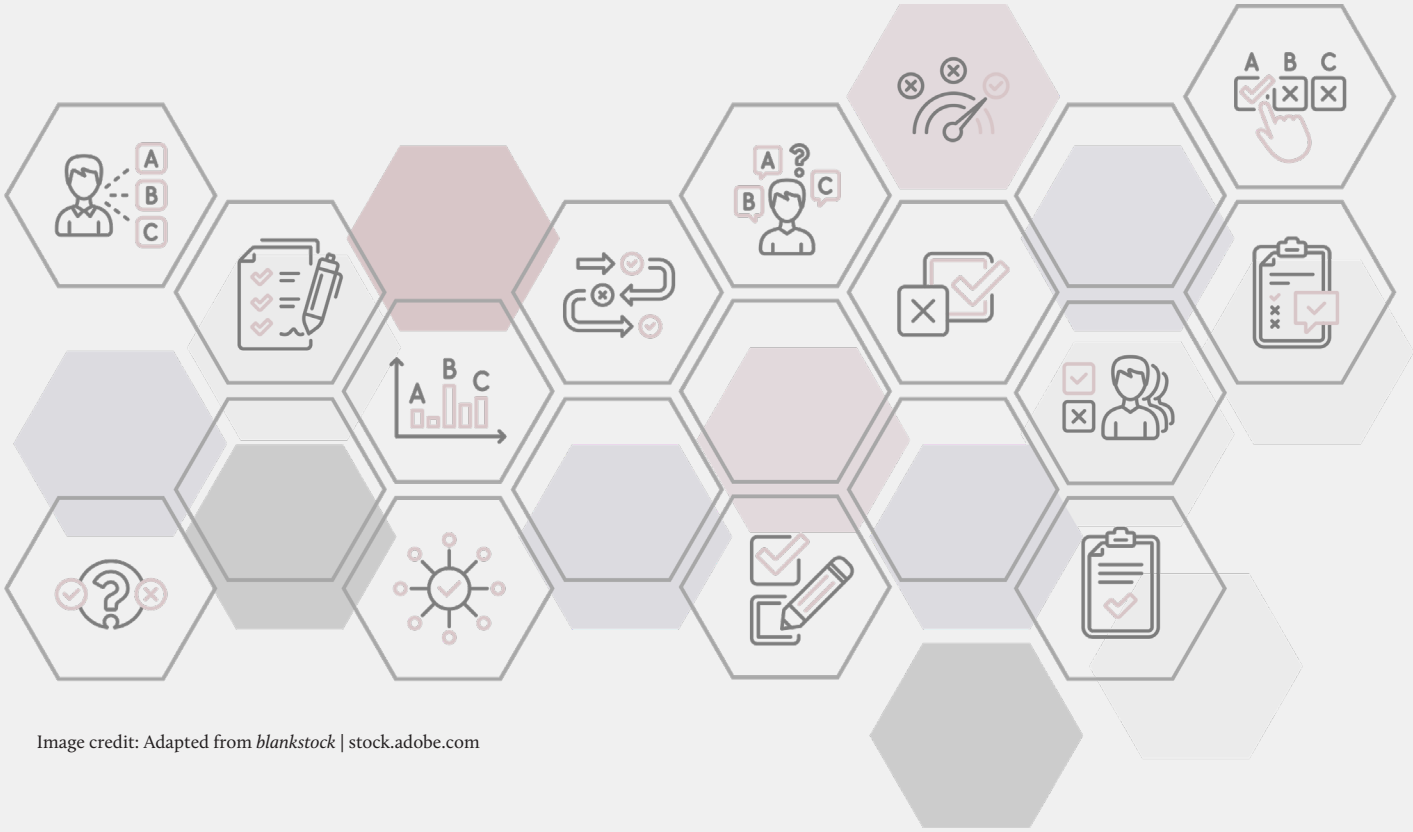


Image credit: Adapted from *blankstock* | stock.adobe.com

Of course, all of this has immediate resource implications. While official training is available, some local areas conduct mock VAF exercises on previously concluded cases that they feel are beneficial because of its safe environment. Other local initiatives additionally provide their own training to Channel panel chairs and partners and share VAF best practice at regional meetings.

Many participants noted how long and potentially unwieldy the VAF is and questioned whether there was a need for so many factors, especially when compared to other instruments used in other parts of policing. However, others felt that with sufficient time and investment, the VAF becomes more useful, overcoming initial feelings of being overwhelmed.

We asked participants about the ease or difficulty in determining the presence of risk factors. Largely, the consensus was that there is no one particular factor that is consistently harder to obtain information on than others. However, online information is difficult to obtain for many practical and technical reasons. The ease of gathering information is case-dependent and highly reliant on building good relationships with local partners.

Many factors become easier to glean information on once the individual is engaging first-hand with the process. For example, issues concerning grievance/injustice, access to networks, and potentially substance misuse are unlikely to be held by partner agencies. Although engagement with Channel is voluntary, individual engagement levels can vary and if low, determining the presence of these factors becomes very difficult. Other factors might be difficult to glean information on for adults (e.g., family attitudes) because families are typically not consulted in such cases.

CONCLUSIONS

Findings derived from the practitioner survey and interviews demonstrate the need to:

- Update guidance documents to demystify some parts of the process.
- Build greater clarity around risk factors.
- Make the practitioner using the VAF feel they are being action-oriented toward building a management plan, rather than simply being a filler of forms.

The VAF is now deployed in a different context than it was built for. The rise of new extremist entities, the morphing of old ones, the rise of the online space as a contributing factor, and the adoption of smaller low-tech and less-sophisticated terrorist plots, may mean some of the guidance requires a re-write. Such a re-write may include considerations of protective factors, comments regarding the relevance of risk factors in each case rather than their simple presence and mention a suite of other issues brought up in the surveys and interviews above.

Evaluations such as ours, and excellent ongoing research by many colleagues on protective factors and new threats can help ensure tools such as the VAF stay up to date and continue to provide valuable information to agencies involved in managing terrorist risks.

Paul Gill is Professor of Security and Crime Science at University College London.

Dr Zoe Marchment is a postdoctoral research associate at University College London.

SHANON SHAH

HOW (NOT) TO MAKE A VIOLENT COPYCAT: LESSONS FROM 'DARK FANDOMS'

Studies of fan cultures, or fandoms, contain insights about 'copycats' that can shed new light on the pathways that perpetrators of violent extremism might take.

Debates about violent extremism (especially jihadist and far-right varieties) often focus on the role of religious belief or political ideology, or both, in motivating the perpetrators. Yet acts of mass public violence are not solely carried out by explicitly religious or political actors. 'Dark fandoms' (groups that are fascinated by people and events central to an act of violence or atrocity) have also added to public concerns about copycat violence.

One paradigmatic example of a 'dark fandom' is the online communities dedicated to Eric Harris and Dylan Klebold, the perpetrators of the 1999 Columbine High School shooting. Earlier studies of 'Columbiners' tended to portray them as deviants, whose admiration for Harris and Klebold was equated with approval of their violent acts.

According to more recent research, however, many fans may have empathised with the bullying (real or purported) experienced by Harris and Klebold, but stopped short of condoning their actions. In other words, dark fandoms are not straightforward incubators of violence — they attract different types and degrees of interest.

Dark fandoms can help us understand how copycat violence can be influenced by three overlapping factors — identification with role models, the intersections of ideological content and practical tactics, and the dynamics of online and offline interactions.

ROLE MODELS

On the one hand, fascination with particular individuals is by itself not a sufficient indicator of the potential for copycat violence. This is true of the Columbiner fans who express empathy for Harris and Klebold whilst not condoning their actions. This ambivalence is also present amongst 'Aumers' — fans of Aum Shinrikyo, the Japanese new religious movement responsible for several violent crimes including the deadly 1995 sarin gas attack on the Tokyo subway.

On the other hand, adulation of Harris and Klebold was clearly present amongst Lindsay Souvannarath and James Gamble, the would-be perpetrators of the foiled 2015 Valentine's Day plot in Nova Scotia, Canada. But, crucially, Souvannarath and Gamble also condoned the actions of Harris and Klebold and wanted to emulate them.

IDEOLOGY AND TACTICS

To use a cookbook analogy: recipes can be followed closely, or they can be improvised and changed. 'True' copycats can thus be understood as perpetrators who seek to reproduce the original recipe as closely as possible.

> **"True" copycats can thus be understood as perpetrators who seek to reproduce the original recipe as closely as possible.**

In the case of the foiled Valentine's Day plot, two aspects of the original recipe help to identify Souvannarath and Gamble as 'true' copycats — their personal adulation for the shooters and the conscious desire to emulate their methods. Souvannarath and Gamble also shared the shooters' disdain for white middle-class, suburban lifestyles and values. Agreement with the perpetrator's beliefs, which can sometimes be expressed implicitly, is therefore another necessary ingredient that defines copycat violence.

Different combinations of these factors can produce different variants of copycat violence. Harris and Klebold, for example, were partly inspired by the Oklahoma City bombing of 1995. According to their diaries, they wanted to exceed its death toll.

Meanwhile, Timothy McVeigh (the Oklahoma City bomber) was partly motivated by vengeance for the Waco Siege of 1993, during which the actions of US enforcement agencies triggered a series of confrontations that culminated in the deaths of 76 members

of a millenarian movement. This explicit vengeance against the US state was absent for the Columbine shooters. McVeigh can therefore be seen as an avenger, a description that does not quite apply to Harris and Klebold. If anything, they drew inspiration from McVeigh's tactics rather than his ideology.

SETTING

The foiled Valentine's Day plot illustrates how online and offline settings influenced Souvannarath's trajectory. After graduating from college in 2014, she started forming friendships only online, where she met and forged a relationship with Gamble through Columbiner networks. The pair began plotting on social media and eventually met in person to carry out their attack before it was stopped by the police.

Souvannarath's journey raises questions about the extent to which online environments provide cognitive openings for radicalisation. Also, do they create a separate-but-parallel virtual reality or an extension of an individual's social reality? Do they encourage aggression by facilitating anonymity?

Meanwhile, it is unclear whether several of the copycats inspired by the Norwegian far-right terrorist Anders Breivik actually belonged to any online or offline communities. We must therefore question if online or offline forums are a necessary or sufficient factor in the making of copycat violence.

Effectively identifying different routes towards violent extremism, therefore, requires us to discern the nuances in *how* people relate to violent perpetrators, their actions, and their motivations in different settings. Online 'dark fandoms' are an important phenomenon to develop better understandings of the distinct and varied ways individuals and groups respond to violent atrocities.

Dr Shanon Shah is a tutor in Interfaith Relations at the University of London's Divinity programme and a researcher at Inform, based at King's College London. His research focuses on social justice trends in contemporary Islam and Christianity, and the intersections of esoteric religious movements and political action.

READ MORE

Read more about some of the research that our contributors mention in their articles. We've flagged up those that are open access and given links to online versions where they are available. For full references and citations please visit the online version at *crestresearch.ac.uk/magazine/technology*

CHRIS BABER: WHY AI SYSTEMS NEED TO EXPLAIN THEMSELVES

Acharya, A., Howes, A., Baber, C., Marshall, T. (2018). Automation reliability and decision strategy: A sequential decision making model for automation interaction. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 62(1), 144-148. *https://bit.ly/350260L*

Baber, C., McCormick, E., Apperly, I. (2022). A Human-Centred Process Model for Explainable AI (No. 7332). *https://bit.ly/3MoglAP*

Chen, X., Starke, S.D., Baber, C., Howes, A. (2017). A cognitive model of how people make decisions through interaction with visual displays. *Proceedings of the 2017 CHI conference on human factors in computing systems*, 1205-1216. *https://bit.ly/3vwTB63*

EMMA BOAKES: CONVERGING SECURITY

Briner, R. (2019). The Basics of Evidence-Based Practice. *People* *https://bit.ly/3IjzsDO*

Symantec (2018). *Internet Security Threat Report* (Volume 23). *https://bit.ly/3IwZxiI*

Symantec (2019). *Internet Security Threat Report* (Volume 24). *https://bit.ly/3BSe7Pj*

Witty, R., Voster, W., Thielemann, J., Olyaei, S., Care., J. (2019). *Predicts 2020: Security and Risk Management Programs.* Gartner.

DAVID BUIL-GIL, JOSE PINA-SÁNCHEZ, IAN BRUNTON-SMITH & ALEXANDRU CERNAT: BAD DATA, WORSE PREDICTIONS

Akpinar, N., De-Arteaga, M., Chouldechova, A. (2021). The Effect of Differential Victim Crime Reporting on Predictive Policing Systems. In *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 838–849. *https://bit.ly/3JO6HiQ*

Biderman, A. D., Reiss, A. J. (1967). On Exploring the "Dark Figure" of Crime. *The ANNALS of the American Academy of Political and Social Science*, 374(1), 1-15. *https://bit.ly/3sjKeEy*

Brantingham, P. (2018). The Logic of Data Bias and Its Impact on Place-Based Predictive Policing. *Ohio State Journal of Criminal Law*, 15(2), 473-486. *https://bit.ly/3hdngJ1*

Buil-Gil, D., Moretti, A., Langton, S. H. (2021). The Accuracy of Crime Statistics: Assessing the Impact of Police Data Bias on Geographic Crime Analysis. *Journal of Experimental Criminology*, 1-27. *https://bit.ly/3sh2HS3*

Cernat, A., Buil-Gil, D., Brunton-Smith, I., Pina-Sánchez, J., Murrià-Sangenís, M. (2021). Estimating Crime in Place: Moving Beyond Residence Location. *Crime & Delinquency*. *https://bit.ly/3HjJseU*

Lum, K., Isaac, W. (2016). To Predict and Serve? *Significance,* 13(5), 14-19. *https://bit.ly/3BQwuUP*

Martin, R. A., Legault, R. L. (2005). Systematic Measurement Error with State-Level Crime Data: Evidence from the "More Guns, Less Crime" Debate. *Journal of Research in Crime and Delinquency,* 42(2), 187-210. *https://bit.ly/3IkVwhv*

Pina-Sánchez, J., Buil-Gil, D., Brunton-Smith, I., Cernat, A. (2021). The Impact of Measurement Error in Models Using Police Recorded Crime Rates. *https://bit.ly/3M28CCi*

UK Statistics Authority. (2014). Statistics on Crime in England and Wales (Produced by the Office for National Statistics). *Assessment Report 268*. London: UK Statistics Authority. *https://bit.ly/3tbuNO9*

LIGHTNING ARTICLES

- **OLI BUCKLEY: IT'S NOT WHAT YOU TYPED, IT'S THE WAY YOU TYPED IT...**

  More information on Oli Buckley's research can be found on his CREST project page: *https://crestresearch.ac.uk/projects/clicka/*

- **MARK LEVINE: CCTV ANALYSIS OF VIOLENT EMERGENCIES**

  Levine, M., Philpot, R., Kovalenko, A. G. (2020). Rethinking the by-stander effect in violence reduction training programs. Social Issues and Policy Review, 14(1), 273–296. *https://bit.ly/3hdL7bh*

  Levine, M., Taylor, P. J., Best, R. (2011). Third parties, violence, and conflict resolution: The role of group size and collective action in the microregulation of violence. *Psychological Science, 22*(3), 406-412. *https://bit.ly/3pfoqou*

  Liebst, L. S., et al., (2021). Cross-national CCTV footage shows low victimization risk for bystander interveners in public conflicts. *Psychology of Violence, 11*(1), 11–18. *https://bit.ly/3pfocxE*

  Philpot, R., et al., (2020). Would I be helped? Cross-national CCTV footage shows that intervention is the norm in public conflicts. American Psychologist, 75(1), 66–75. *https://bit.ly/3M2Si4z*

  Philpot, R., Levine, M. (2021). Evacuation behavior in a subway train emergency: A video-based analysis. *Environment and Behavior, 54*(2), 383-411. *https://bit.ly/3MouVbK*

- **HEATHER SHAW: THE IDENTITY IN EVERYONE'S POCKET**

  Ellis, D.A., Davidson, B.I., Shaw H.; Geyer, K. (2019). Do Smartphone Usage Scales Predict Behavior? *International Journal of Human-Computer Studies*, 130, 86-92. *https://bit.ly/3HdBXGB*

  Shaw, H., Ellis, D.A., Geyer, K. Davidson, B.I., Ziegler, F.V., Smith, A. (2020). Quantifying smartphone "Use": Choice of Measurement Impacts Relationships Between "Usage" and Health. *Technology, Mind and Behavior*, 1(2). *https://bit.ly/3IgaEwu*

  Shaw, H., Taylor, P.J., Ellis, D., Conchie, S. (2021). Behavioral consistency in the digital age. Psychological Science, (in press). *https://bit.ly/35mxSm6*

- **LEON REICHERTS: "OK GOOGLE, SHOULD I CLICK ON THAT EMAIL?"**

  Reicherts, L., Rogers, Y. (2020). Do Make me Think! How CUIs Can Support Cognitive Processes. *Proceedings of the 2nd Conference on Conversational User Interfaces*, 1–4. *https://bit.ly/3LYNTiU*

  Reicherts, L., Wood, E., Duong, T. D., Sebire, N. (2022). It's Good to Talk: A Comparison of Using Voice Versus Screen-Based Interactions for Agent-Assisted Tasks. *ACM Transactions on Computer-Human Interaction*, 29(3), 1-41 *https://bit.ly/3shTs44*

CARL MILLER: CHINA'S DIGITAL DIPLOMACY

Bartlett, J. (2014). Vox Digitas: Social Media is Transforming How to Study Society, *Demos*, *https://bit.ly/3JNZNKy*

Brandt, J., Schafer, B. (2020). How China's "wold warrior" diplomats use and abuse Twitter. *Tech Stream. https://brook.gs/3vmG4Ol*

Chaguan. (2020). China's "Wolf Warrior" Diplomacy Gamble, *The Economist*. *https://econ.st/3MopREo*

Mazumdar, B.T. (2021). Digital diplomacy: Internet-based public

diplomacy activities or novel forms of public engagement?. *Place Branding Public Diplomacy*. https://bit.ly/3BNUYoY

SOPHIE NIGHTINGALE: IDENTITY FRAUD IN THE DIGITAL AGE

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J., Burton, A. M., Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied*, 5(4), 339. https://bit.ly/3hkKcGo

Damer, N., Saladie, A. M., Braun, A., Kuijper, A. (2018). Morgan: Recognition vulnerability and attack detectability of face morphing attacks created by generative adversarial network. *In IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 1-10. https://bit.ly/3BNWkJ6

Ferrara, M., Franco, A., Maltoni, D. (2014). The magic passport. In *IEEE International Joint Conference on Biometrics*, 1–7. https://bit.ly/3JRSvW5

Kramer, R. S., et al., (2019). Face morphing attacks: Investigating detection with humans and computers. *Cognitive research: principles and implications*, 4, 1-15. https://bit.ly/35qFV14

Nightingale, S., Agarwal, S., Härkönen, E., Lehtinen, J., Farid, H. (2021). Synthetic faces: how perceptually convincing are they? *Journal of Vision*, 21(9), 2015. https://doi.org/10.1167/jov.21.9.2015

Nightingale, S., Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy, *Proceedings of the National Academy of Sciences*, 119 (8). https://bit.ly/353cZwC

Nightingale, S. J., Agarwal, S., Farid, H. (2021). Perceptual and computational detection of face morphing. *Journal of Vision*, 21, 1-18. https://doi.org/10.1167/jov.21.3.4

Phillips, P. J., Yates, A. N., Hu, Y., Hahn, C. A., Noyes, E., Jackson, K., ... O'Toole, A. J. (2018). Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *Proceedings of the National Academy of Sciences*, 115, 6171-6176.https://bit.ly/3HkMkbJ

Robertson, D. J., Kramer, R. S., Burton, A. M. (2017). Fraudulent ID using face morphs: Experiments on human and automatic recognition. *PLoS One*, 12, https://doi.org/10.1371/journal.pone.0173319

Robertson, D. J., Mungall, A., Watson, D. G., Wade, K. A., Nightingale, S. J., Butler, S. (2018). Detecting morphed passport photos: A training and individual differences approach. *Cognitive Research: Principles and Implications*, 3, 1-11. https://doi.org/10.1186/s41235-018-0113-8

Robertson, D. J., Towler, A., Sanders, J., Kramer, R. S. (2020). Hyper-realistic masks are extremely hard to spot – as our new research shows. *The Conversation*. https://bit.ly/33SwN5c

Scherhag, U., Rathgeb, C., Merkle, J., Breithaupt, R., Busch, C. (2019). Face recognition systems under morphing attacks: A survey. *IEEE Access*, 7, 23012-23026. https://doi.org/10.1109/ACCESS.2019.2899367

Towler, A., White, D., & Kemp, R. I. (2017). Evaluating the feature comparison strategy for forensic face identification. *Journal of Experimental Psychology: Applied*, 23, 47-58. https://bit.ly/3Inmghd

MARION OSWALD: 'GIVE ME A PING, VASILI. ONE PING ONLY'

Babuta, A., Oswald, M. (2020). 'Data analytics and algorithms in policing in England and Wales: Towards a new policy framework'. *RUSI Occasional Paper*. https://bit.ly/3HhO3OH

Babuta, A., Oswald, M., Janjeva, A. (2020). 'Artificial Intelligence and UK national security: policy considerations'. *RUSI Occasional Paper*. https://bit.ly/3BNDkL3

Oswald, M. (2018). 'Algorithmic-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power'. *Phil. Trans. R. Soc. A*, 376(2128). https://bit.ly/3t2k1cO

Oswald, M. (2020). 'AI and national security: learn from the machine, but don't let it take decisions'. https://bit.ly/3JWRxIu

Oswald, M. (in press). 'A Three-Pillar Approach to Achieving Trustworthy Use of AI and Emerging Technology in Policing in England and Wales: Lessons From the West Midlands Model'. *European Journal of Law and Technology*. https://bit.ly/3t2clau

SHANON SHAH: HOW (NOT) TO MAKE A VIOLENT COPYCAT: LESSONS FROM 'DARK FANDOMS'

Bangstad, S. (2021). What has Norway learned from the Utøya attack 10 years ago? Not what I hoped. *The Guardian*. https://bit.ly/3pfqtMo

Berger, J. M. (2019). The Dangerous Spread of Extremist Manifestos. *The Atlantic*. https://bit.ly/36zvxol

Gill, P. (2012). Tracing the Motivations and Antecedent Behaviors of Lone-Actor Terrorism: A Routine Activity Analysis of Five Lone-Actor Terrorist Events. *International Center for the Study of Terrorism*. https://bit.ly/3JQOeCs

Lamoureux, M. (2019). The Woman Who Plotted a Valentine's Mass Murder Shares How the Internet Radicalized Her. *Vice*. https://bit.ly/3peXnga

Monroe, R. (2019). The cult of Columbine: how an obsession with school shooters led to a murder plot. *The Guardian*. https://bit.ly/33QMwS9

Osaki, T. (2014). Aum cultists inspire a new generation of admirers. *The Japan Times*. https://bit.ly/3sifhR4

Vidal, G. (2008). The Meaning of Timothy McVeigh. *Vanity Fair*. https://bit.ly/3t4KpD2

Wright, B. L. (2019). Don't fear the nobodies: A critical youth study of the Columbiner Instagram community. *Mississippi State University*. https://bit.ly/3JYCmOV

ISABELLE VAN DER VEGT, BENNETT KLEINBERG, & PAUL GILL: LINGUISTIC THREAT ASSESSMENT: CHALLENGES AND OPPORTUNITIES

Kleinberg, B., van der Vegt, I., Gill, P. (2021). The temporal evolution of a far-right forum. *Journal of computational social science,* 4(1), 1-23. https://link.springer.com/article/10.1007/s42001-020-00064-x

Meloy, J.R., Hoffman, J. (Eds.). (2021). *International Handbook of Threat Assessment*. Oxford University Press.

Pennebaker, J. W., Francis, M. E., Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, 71, https://bit.ly/3McQXZ5

Scrivens, R. et al., (2021). Comparing the Online Posting Behaviors of Violent and Non-Violent Right-Wing Extremists. *Terrorism and Political Violence*, 1-18. https://doi.org/10.1080/09546553.2021.1891893

van der Vegt, et al., (2021). The Grievance Dictionary: understanding threatening language use. *Behavior research methods,* 1-15. https://bit.ly/3BQgsKM

van der Vegt, I., et al., (2020). Online influence, offline violence: language use on YouTube surrounding the 'Unite the Right' rally. *Journal of computational social science*, 333–354. https://bit.ly/36FZCmq

van der Vegt, I., Kleinberg, B., Gill, P. (in press). Predicting author profiles from online abuse directed at public figures. *Journal of Threat Assessment and Management.*

van der Vegt, I. et al., (in press). Assessment procedures in anonymously written threats of harm and violence. *Journal of Threat Assessment and Management.*

van de Vegt, I.; Kleinberg, B., Gill, P. (2020). Too good to be true? Predicting author profiles from abusive language. *arXiv:2009.01126v2*

CENTRE FOR RESEARCH AND EVIDENCE ON SECURITY THREATS

CREST Security Review provides a gateway to the very best knowledge and expertise. Its articles translate academic jargon to 'so what' answers and illustrate how behavioural and social science can be used effectively in everyday scenarios.

THE CENTRE FOR RESEARCH AND EVIDENCE ON SECURITY THREATS

CSR is produced by the Centre for Research and Evidence on Security Threats (CREST). CREST is funded by the UK's Home Office and security and intelligence agencies to identify and produce social science that enhances their understanding of security threats and capacity to counter them. CREST also receives funding from its core partners (the universities of Bath, Lancaster and Portsmouth). Its funding is administered by the Economic and Social Research Council (ESRC Award ES/V002775/1), one of seven UK Research Councils, which direct taxpayers' money towards academic research and training. The ESRC ensures the academic independence and rigour of CREST's work.

CREST has established a growing international network of over 140 researchers, commissioned research in priority areas, and begun to tackle some of the field's most pressing questions.

> '*CREST Security Review is a fantastic means by which we can keep practitioners, policy-makers and other stakeholders up-to-date on the impressive social and behavioural science occurring not only at CREST, but around the world.*'
>
> Professor Stacey Conchie, CREST Director

For more information on CREST and its work visit **www.crestresearch.ac.uk** or find us on Twitter, @crest_research