THE GOOD, THE BAD AND THE UGLY (NEWS)

What if wars could be won by words? Without a missile being fired, propaganda alone could swing the mood of a nation and undermine the actions of its leaders. Fake news? It's only just getting started.
By **Paul Dempsey**

"IT WOULD NOT BE impossible to prove with sufficient repetition and a psychological understanding of the people concerned that a square is in fact a circle."

'Repetition' captures the amplification of a message possible through social media. 'Psychological understanding' is a big goal in data mining. That could be said by any of today's propagandists.

These words, though, are actually those of Joseph Goebbels, Adolf Hitler's chief of propaganda. They remind us that few ideas behind fake news are new.

The current controversy is partly because such manipulation gains renewed momentum atop each wave of innovation in mass communications. The phenomenon stretches back to at least the invention of the printing press, yet seems to catch us off guard every time.

During each wave of fakery, we eventually wake up to the need for mitigation – usually both technological and social. The process under way now is no different.

Psychologists are seeking a finer understanding of how black propaganda works so they can inoculate society. You might think after so long that the topic is well understood, but some recent findings are surprising.

Meanwhile, experts in digital media forensics are refining tools that separate fair reporting and comment from the actively deceitful. As elsewhere in cyber security, they battle the rate at which digital fakery is evolving.

We must look at research on both sides to understand how fake news can be combated. Psychology is the best starting point.

Fake news as social disease

Researchers contend that fake news inserts itself into the civil discourse by exploiting basic human traits. The more important of these include:

- **'Confirmation bias' and 'motivated reasoning':** the former refers to our openness to information that supports our world views and the latter to our susceptibility to make decisions emotionally or dogmatically rather than rationally.
- **'Thought bubbles':** these represent the way in which we organise ourselves on social media around people, institutions and others with whom we agree, excluding dissenting voices (consciously or unconsciously).

Several psychologists argue that the critical first step in overcoming these weaknesses and thus reducing fakery's

CAPITAL PICTURES, ALAMY, GETTY IMAGES, REX FEATURES

Bamse bear was a famous Swedish cartoon character that was used to educate people about fake news

impact is to get effective warnings out to the public.

Researchers Dorje Brody from the University of Surrey and David Meier from Brunel University are working to model fake news using communications theory. They capture that received wisdom: "The mere knowledge that pieces of fake news may be in circulation goes a long way towards mitigating the impact of fake news."

However, warnings may not be enough. Other research suggests that we then need people to think more critically.

Professors Gordon Pennycook of Regina University and David Rand of Yale University asked a sample to assess a number of real and fake news articles and also take a cognitive reading test (CRT).

They found the sample's ability to detect fake news was not primarily undermined by motivated reasoning. Many subjects were good at detecting fakes that reflected their own political beliefs. This was, Pennycook has argued, perhaps because their views gave them better evaluation tools.

Instead, correlation with the CRT suggested that fake-news believers might be better defined by an unwillingness and/or inability to think analytically. "Our findings... suggest that susceptibility to fake news is driven," Pennycook and Rand write, "more by lazy thinking than it is by partisan bias *per se*."

Further work by this duo – with Tyrone Cannon – has then suggested that it is not only our biases that define our susceptibility.

"Using actual fake news headlines presented as they were seen on Facebook, we show that even a single exposure increases subsequent perceptions of accuracy, both within the same session and after a week," their research says.

"Moreover, this 'illusory truth effect' for fake news headlines occurs despite a low level of overall believability, and even when the stories are labelled as contested by fact checkers or are inconsistent with the reader's political ideology."

Our biases undoubtedly make some contribution. But we are kidding ourselves if we think that the credulous alone are vulnerable. Everything is rather a matter of degree. Everyone is at risk.

The 'whole of society' defence

This theory that fake news can be mitigated at a high social level is supported by experiences from two recent elections: 2018 in Sweden and 2017 in France.

The French Presidential election was explicitly attacked. The aggressors, most likely Russian, used much the same weapons as in the US in 2016. There was a large email leak. The 'anti-Russian' candidate, Emmanuel Macron, was smeared. Fake news was amplified by online trolls.

It didn't work. Macron got 66 per cent of the vote.

In an analysis for the Center for Strategic and International Studies, Jean-Baptiste Jeangène Vlimer says that, as well as multiple warnings from the Macron campaign, there was a concerted attempt to keep the threat of fake news in the public eye.

Two independent state organisations – the Control of the Electoral Campaign for the Presidential Election (CNCCEP) and the National Cybersecurity Agency (ANSSI) – monitored the vote and issued frequent public warnings.

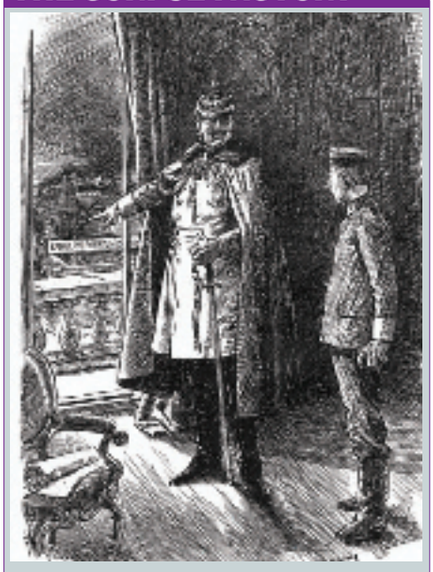
This contrasted with what had happened in the US, where President Obama's administration hesitated until late in the 2016 campaign. It was wary of being accused of minting its own fake news to favour Hillary Clinton.

There was a further contrast over the email leak. In the US, the media dissected emails stolen from John Podesta, Clinton's chief of staff, and drew stories from them for days. In France, the media focused on the timing behind the dump from inside Macron's En Marche organisation and the motivations of those behind it.

The Macron leak came just hours before the campaign went into a period of >

FAKE NEWS

THE CORPSE FACTORY



Fake news has a dishonourable history. One infamous example from the First World War was 'The German Corpse Factory'. Its implications stretched into the Second World War.

A rumour that Germany was rendering the bodies of its fallen to make fat and soap has been traced back to 1915, but began to spread more widely in 1917 through newspaper articles in China, Belgium, the Netherlands and the UK.

The Chinese version is thought to have been part of a British propaganda campaign aimed at bringing a new ally into the war. However, the other articles embellished the lie with reference to a specific factory.

They appear to have been spun around the mistranslation – very possibly deliberate – of one word from a brief item in a German newspaper about the pungent smell coming from a "carcass-utilisation" factory (*Kadaververwertungsanstalt*). Unlike its English homophone, '*Kadaver*' almost exclusively refers to the dead body of an animal. In German, a human corpse is a '*Leiche*' or '*Leichnam*'.

Some readers queried the translation, but the tale proved persistent. It spread widely enough to inspire a cartoon in the satirical magazine *Punch*: "And don't forget that your Kaiser will find a use for you – alive or dead."

The story was not officially debunked until 1925. Lingering anger allowed the fake to be exploited by Hitler. He cited it to prove that the British were liars who knew no bounds when accusing adversaries of war crimes.

It also long embarrassed senior officials in London. In 1942, reports reached them of a mass slaughter of Jews in the Warsaw ghetto. The chair of the British Joint Intelligence Committee doubted their credibility because of details that echoed "stories of employment of human corpses during the last war for the manufacture of fat which was a grotesque lie".

<pre-election purdah, when government activity is restricted. The perpetrators probably hoped this would make the contents harder to respond to. It served instead to make their actions look dubious.

French administrators acknowledge learning from the earlier experiences of others. This was even more the case for Sweden's elections in September 2018. That country had also suffered plenty of pre-digital campaign interference from Russia, and has long valued counter-measures. In a report for Harvard's Belfer Centre for Science and International Affairs, Gabriel Cederberg describes how Sweden adopted a 'whole-of-society' defence.

Coordinated communication by state security agencies was bolstered with fake news warnings from independent but respected figures such as King Carl XVI Gustaf and even a favourite cartoon character, Bamse.

Sweden's five largest media groups established a joint fact-checking unit. A telecoms company launched a board game called The Hunt for Truth. Looking to the future, Sweden even deployed counter-measures in schools.

"In March 2017, the government announced a nationwide curriculum reform to increase elementary and high school students' computer science skills and ability to recognise fake news," Cederberg writes. "The curriculum was officially launched in July 2018."

Those who sought to interfere failed in France. The consensus in Sweden is that they were barely seen.

Can we copy this model and breathe more easily? Have social engineers got their digital counterparts off the hook? Not really.

The technologies used in political interference so far are familiar and relatively unsophisticated. The next generation will likely be driven by AI, and is already raising fears about how an informed and wary voter can distinguish between reality and fakery.

The lAIr of the liar

Walt is an online AI tool developed under a team led by Mike Tamir, head of data science in the Advanced Technology Group at Uber, lecturer at UC Berkeley and crusader against fake news.

Given any article's URL, Walt undertakes textual analysis based on natural language processing, deep learning, neural networks and a human feedback loop that leverages wisdom-of-crowds theory. It then places content – based on its headline and body copy – within six categories: journalism, wiki, opinion, satire, sensational, and agenda-driven.

Tamir likens Walt's assessments to "the nutrition label on a packet of food". Beyond the flashy packaging, how much good stuff is there inside? He claims the tool has achieved 90 per cent plus accuracy (you can try it for yourself at fakerfact.org).

Right now, it has two immediate drawbacks.

First, "90 per cent plus" is not yet high enough to use Walt as a filter to pre-empt fake news. That level of intervention will need close to six-sigma accuracy. Still, it shows how rapidly digital media forensics are evolving.

Second, you must therefore go to Walt as you might go to a fact-checking site like Snopes. The tool's effectiveness depends on the user's awareness of fake news and suspicions about an article. We still need that social prodding.

Walt then faces a third looming challenge. As much as it is born of AI, AI may soon be its enemy.

Almost all fake news today is written by humans. That which is not is easy to detect, and the manually crafted guff is not much better.

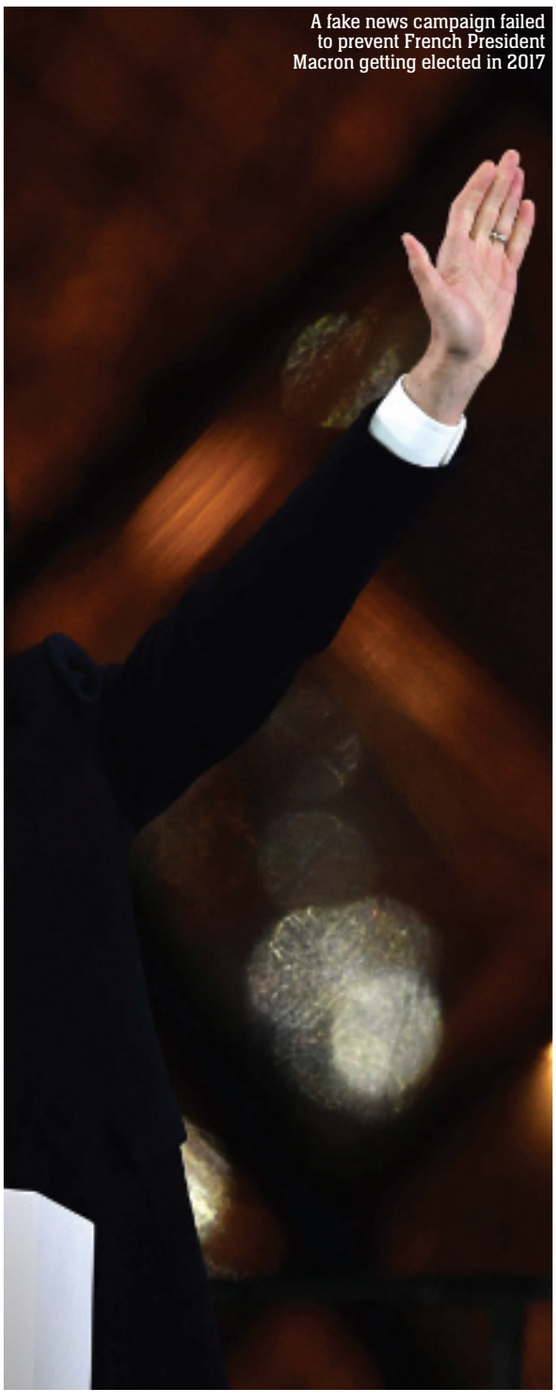
Journalism has rules about structure and content. There are style books governing the prose. You typically have to be trained. Another way of looking at this is to think of when you last saw a newspaper article or TV bulletin in a drama. Chances are, it felt phoney. Now consider that that was the result of one professional writer trying to mimic another.

But what if you could master those rules and perform mimicry on their terms?

Generative adversarial networks (GANs) are an emerging AI technique. They basically comprise two conferring AIs. One generates 'candidates' and the other evaluates them until they get an optimised strategy.

Given Walt's existence, we can imagine a GAN exercise for written fake news. Rules, style books and the millions of genuine online articles could be fed in as training data. The 'candidates' AI would generate cyber hacks, while on the other side a Walt-like evaluator marked them.

It need not only be written fakes that might soon be optimised like this. The term 'deepfakes' refers to the use of deep learning to generate fake videos into which real



A fake news campaign failed to prevent French President Macron getting elected in 2017

FAKE NEWS

MACEDONIAN MONEY

In 2016, the Macedonian town of Veles was exposed as the home to more than 100 fake news providers supporting Donald Trump's presidential campaign. Unlike Russia's Internet Research Agency, the Veles cluster had no political interest. It backed Trump for the money.

By feeding the Twitter and Facebook armies amplifying material that damaged Hillary Clinton, the Macedonians found that they could each earn about €1,000 a month in click-thru revenue from Google AdSense and others. That was more than twice the national average income.

The scam's pioneers even set up how-to classes for local teenagers. This was easy money.

After the election, social media finally cracked down on the Veles cluster. However, its leaders pledged to be back for the US's 2018 and 2020 elections. In fact, they appear to have been back even sooner and even closer to home.

In September, Macedonia's government held a referendum seeking a mandate to change the country's name. The background was a diplomatic dispute with Greece. Resolving it would clear a path for Macedonia to join the EU and Nato.

Russia was opposed and is suspected to have launched another interference campaign. However, Team Veles now also had a real interest.

Veles was the hotbed for global fake news

Under current Macedonian law, its 2016 fakery was not illegal. Were the country to join the EU and Nato, that would change.

Veles' fakers opened the spigot – or at least that is what one of its members says via ProtonMail – producing anti-name change 'news' and looking to depress turnout.

When the results came in, the referendum failed to hit the 50 per cent turnout mark that, alongside a 'Yes', would have given the government its mandate. Nationally, it was 34.8 per cent. In Veles, it was 28.2 per cent.

people are inserted. Its evolution also draws on GANs.

Deepfake technology has been making headlines since the beginning of the year. The wide release of the first apps was exploited by internet pond life to create revenge porn. But in April, wider fears were sparked. Writer-director Jordan Peele released a video on YouTube in which he mimicked former President Obama, making him appear to call his successor "a total and complete dipsh*t".

Neither deepfake- nor GAN-based written technology is infallible yet. But it is getting better all the time.

Earlier this year, Professor Siwei Lyu and colleagues from SUNY-Albany described the use of eyeblink-based forensics to detect deepfakes. They spotted a flaw. AIs were receiving still images as training data. Stills seldom show a subject with eyes closed, but humans blink every 2-15 seconds. The AIs generated creepy-looking, unblinking subjects.

The work attracted plenty of attention, including within the deepfake community. At least one member emailed Lyu to let him know that the problem had been identified. He said that savvier users had already started using video in the training data.

This highlights another important factor in fake news's potential evolution. Many of its foundational technologies are cheap. Moore's Law economics drives down the cost of the CPUs and GPUs. The Google-developed TensorFlow framework for machine learning is open-source. The source code behind the most popular deepfake generator – FakeApp – has also been freely placed on GitHub.

Alina Polyakova, a fellow at the Brookings Institution who closely tracks Russia's cyber capacity, is not convinced that society is ready to deal with these emerging threats and their diversifying sources. It ain't just about Putin.

'Deepfake' technology used by director Jordan Peele

"Detecting automated accounts, often called 'bots', will become more difficult as these accounts appear increasingly human," she writes in Brookings' recent report on Order from Chaos. "They will be able to adapt to human reactions, tailor messaging and exploit human emotions."

Then, there is the question of time. Fake news's dissemination exploits a widespread and synchronised promotion of black propaganda. One signature strategy is for many bots and trolls to point at the same URL within a very short period of time. It follows a thought widely (but wrongly) attributed to Mark Twain: "A lie can travel halfway around the world while the truth is still putting on its boots."

This presents another conundrum. AIs may ultimately create nearly, perhaps even completely, undetectable fakes, but in the medium term they may only need to produce them to the 'good enough' standard. The propagandist's benchmark is not perfection, but enough damage before the truth arrives.

There are a lot of smart people trying to make sure that black propagandists fail. But their success is not assured. Spread the word, and if you have any well-founded ideas, you will likely find an audience.

Or let's borrow from Jordan Peele. He signed off his Obama deepfake warning with this: "Stay woke, bitches." •