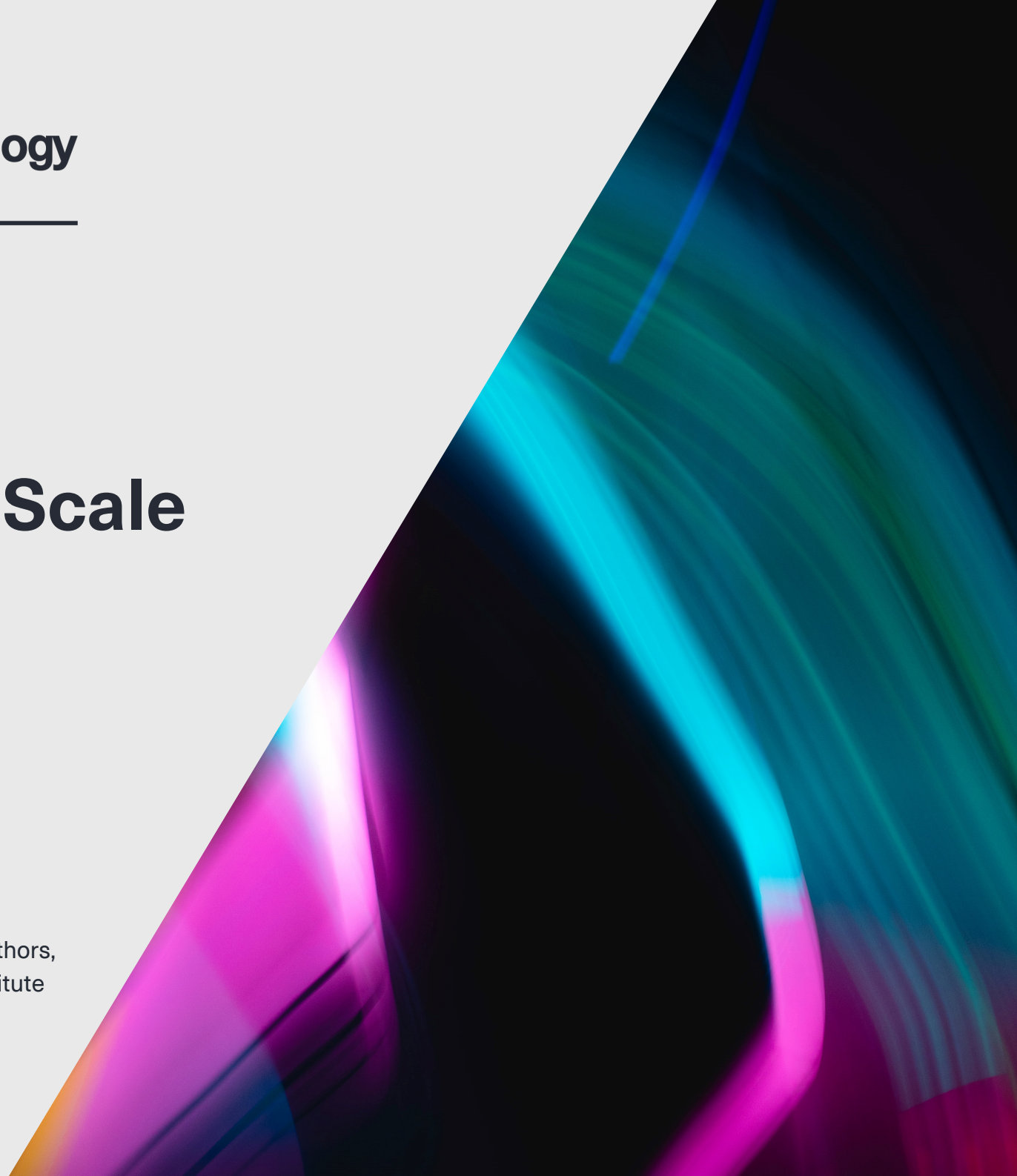
Centre for Emerging Technology and Security
EXPERT ANALYSIS

Voice Cloning At Scale

Ant Burke



The views expressed in this article are those of the authors, and do not represent the views of the Alan Turing Institute or any other organisation.

Introduction

Artificial Intelligence (AI) voice cloning is an emerging form of 'deep fake' that can create new audio content that mimics a person's voice, based on a short recording. Improvements in voice cloning technology will create new opportunities for criminals and other malicious actors, and the UK security community will need to develop new capabilities to keep pace with this evolving threat.

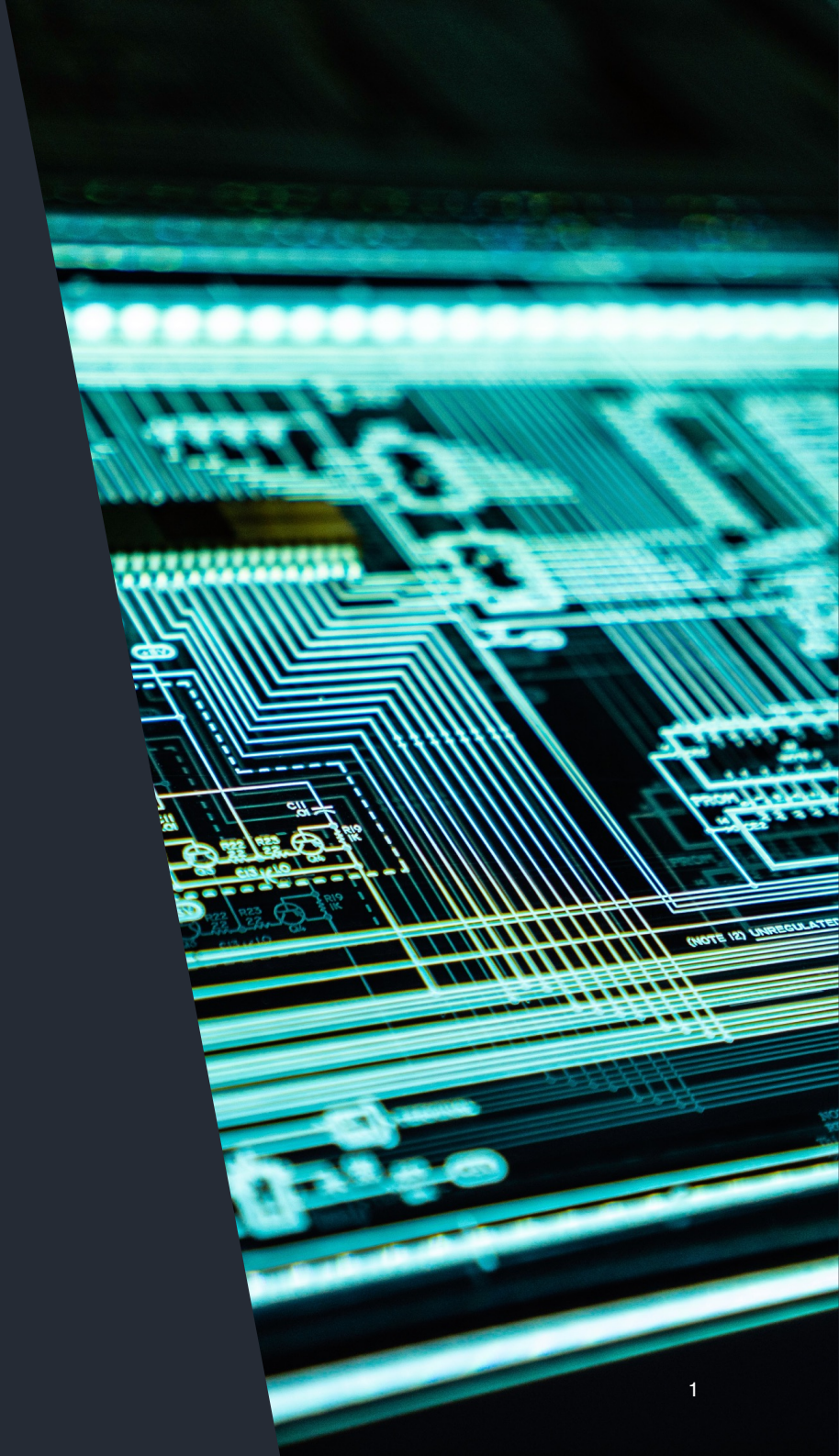
'Deep fake' is now a term within everyday discourse, with widespread coverage in many global mainstream media outlets. Numerous headline-grabbing demonstrations of convincing – yet entirely fake – videos have fuelled interest and created suspicion in public perceptions of digital content.[i] The creative industries are driving progress and there are many legitimate commercial uses of deep fake technology, as a natural extension of the rich domain of special effects.[ii]

Alongside these legitimate commercial developments, the terms 'disinformation', 'fake news' and 'troll factories' have also become part of everyday discourse, as the public has become increasingly aware of malicious use of the internet and social media to influence, confuse, divide and manipulate audiences for political or strategic gains.

In 2018, the UK Ministry of Defence (MOD) published a joint concept note on Human Machine Teaming,[iii] warning of the future threat from manipulative AI technology:

> *When untrained amateurs, or automated social engineering web robots (bots) can produce fake videos at a higher quality than today's Hollywood computer-generated imagery, forgeries are likely to constitute a large proportion of online content. Such forgeries will challenge trust in, and between, institutions.*

The global information and political landscape has changed dramatically in this time. Disinformation is increasingly embedded in many of the digital platforms that citizens rely upon to make sense of their world. The adage 'Everyone is entitled to his own opinion, but not his own facts' now feels like a relic of a bygone era.[iv] It is within this context that we explore the maturity of AI voice cloning and the risks it may pose to UK security.



Where are we now?

While AI voice cloning is a here and now technology,[v] its malicious use by hostile actors has so far been limited to niche and small-scale fraud attempts. As the technology improves and becomes cheaper and easier to access and use, this situation is likely to change.[vi]

Voice cloning can be seen as an 'off-the-shelf' commercial product, with a plethora of start-up companies now offering AI cloned voice services. Resemble.ai[vii] is one such example, targeting creative and marketing industries with AI voice cloning and content generation – underpinned by cutting-edge machine learning research.[viii] Notably, the developers have demonstrated proactive consideration of the ethical risks associated with their product, publishing their ethical position and releasing tools to support countering disinformation.[ix]

But AI voice cloning is no longer the preserve of start-ups. Microsoft recently announced the availability of voice cloning technology within the Azure Cloud 'Cognitive Services' suite. Microsoft has implemented strict controls to reduce the risk of misuse, requiring potential customers to provide detailed business cases for approval.[x] Not all vendors will do this, nor would any malicious actor that seeks to develop their own capabilities. As noted by Resemble.ai, defending against malicious actors relies on the ability to distinguish real from fake; being able to quickly, accurately and reliably determine whether content is captured from the real world or a forgery. Significant research has already been conducted into deep fake detection, and such research must continue to develop new capabilities to appropriately identify and triage deep fake content.[xi xii xiii xiv]

AI progress has moved hand in hand with cloud computing progress. Previously the preserve of well-resourced institutions, cloud services have made advanced compute and telephony services available to anyone with a credit card. While developing complex vast AI technology still requires significant resource, powerful capabilities are now available to consumers for hundreds or thousands of dollars.[xv xvi]

Voice cloning in practice

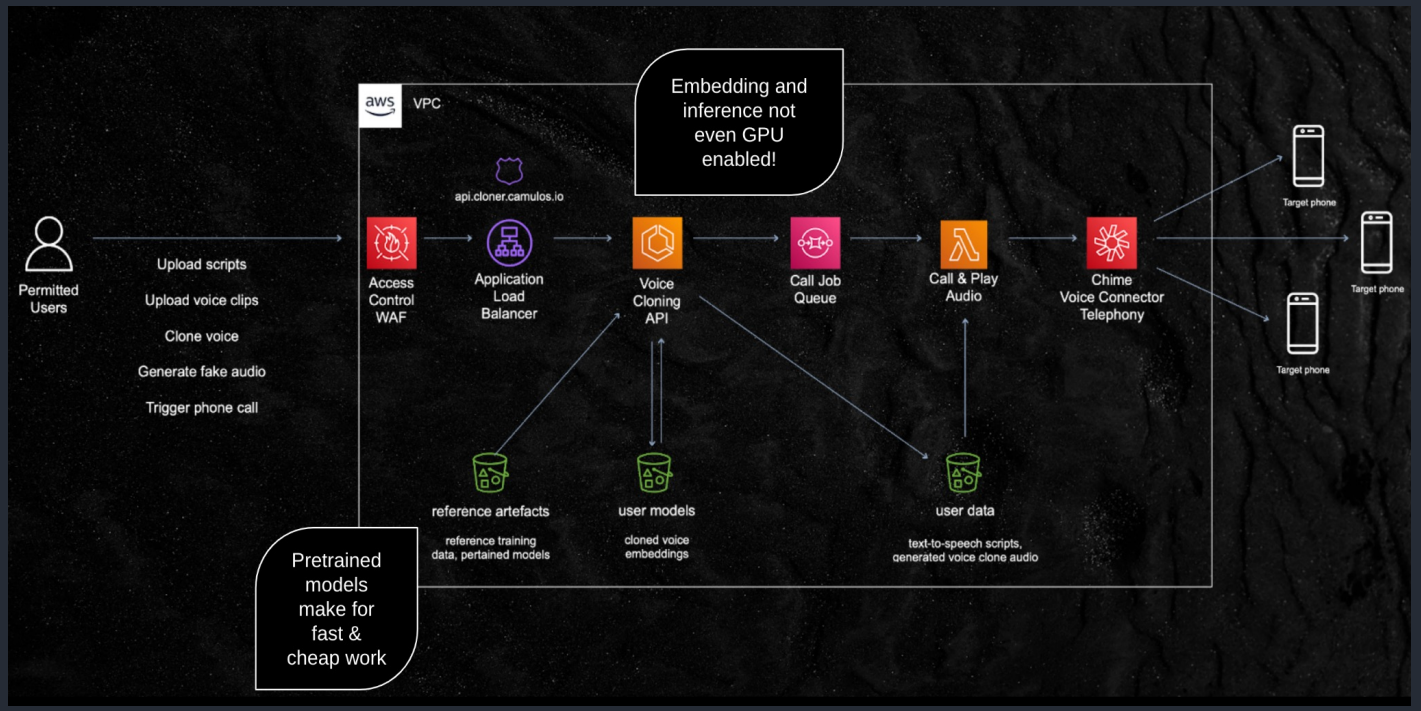
We conducted an AI voice cloning proof-of-concept project.

We set out to build a scalable voice cloning platform and integrate voice cloned audio into a telephony system, to mimic a system that allows users to create and use cloned audio to influence individuals directly.

By combining Amazon Web Services (AWS) managed telephony services, open-source AI (code and models), modern software development practices and automation, it was possible to build a rudimentary yet scalable voice cloning telephony platform in a matter of a few hours, spread over several days. The graphic below illustrates the architecture of the proof-of-concept demonstrator.

This shows the asymmetry of effort and outcomes in building deep fake capabilities - one person with an open-source project and an AWS subscription is now able to create a highly scalable voice cloning telephony service. A novice team could do a great deal more with ease.

A multidisciplinary team of AI researchers, machine learning practitioners, data scientists and software engineers could achieve much greater sophistication and scale very quickly.



What comes next?

In the near term, we expect AI voice cloning technologies to be actively developed within the creative industries to support content creation and marketing.

In parallel, it is likely that off-the-shelf or modified off-the-shelf technologies will be modified for use in relatively crude 'bulk' influence or disinformation efforts - for example hacktivists running mass marketing campaigns, websites offering cloned celebrity voices to create voicemail recordings, or scammers using new tools to trick people into providing sensitive credentials via phone.

We expect research and development on the underlying voice cloning technology and text-to-speech (TTS) technologies to continue to improve. Research into the underpinning cloning AI will improve the quality, naturalness and complexity of expression for cloned voice content, making it more convincing and harder to detect.

Beyond targeted scams, lifelike voice distributed denial of service (DDoS) attacks and circumvention of voice ID security, improvements in voice cloning AI combined with real-time text-to-speech and speech-to-speech are likely to create new tools for those seeking to engage in targeted influence operations.

Increasingly lifelike cloned voices, combined with the ability to auto-generate content with large scale language models (such as GPT3, Turing-NLG, OPT-175B, PaLM) to create lifelike audio containing hundreds or thousands of variations of dates, times, locations, objects or topics could be used as a form of DDoS attack to flood out real conversations within the noise of many thousands of decoy phone calls.

This combination of language models and text-to-speech cloned voices is just one form of mixed-modality deep fake. Deep fake content combining audio, text, image, social media activities and video will become a primary challenge for those wishing to distinguish what is genuine from convincing machine-generated forgeries.

Industries seeking to create new immersive experiences will push forward with generative AI seeking to create moments of delight and wonder. Despite many positive and legitimate uses, the proliferation of cheap, easy and realistic voice cloning AI – and deep fake technology more broadly – is a deeply unsettling prospect, and could lead to increased fraud, damage to public perceptions and social norms, and an erosion of trust in digital content and sources.

The defence and security community will, as ever, adapt to respond to the changing environment. Organisations will develop new capabilities to detect, deny and disrupt malicious actors leveraging deep fake AI, and support sustained efforts to keep pace with new techniques and applications. Partnerships with the wider research community will be crucial to the success of these efforts.

The UK is already home to several notable research partnerships in this space, such as Edinburgh University's Centre for Speech Technology's ASVspoof programme, a leading research collaboration where voice cloning and clone detection research is conducted in tandem.[xvii]

What comes next?

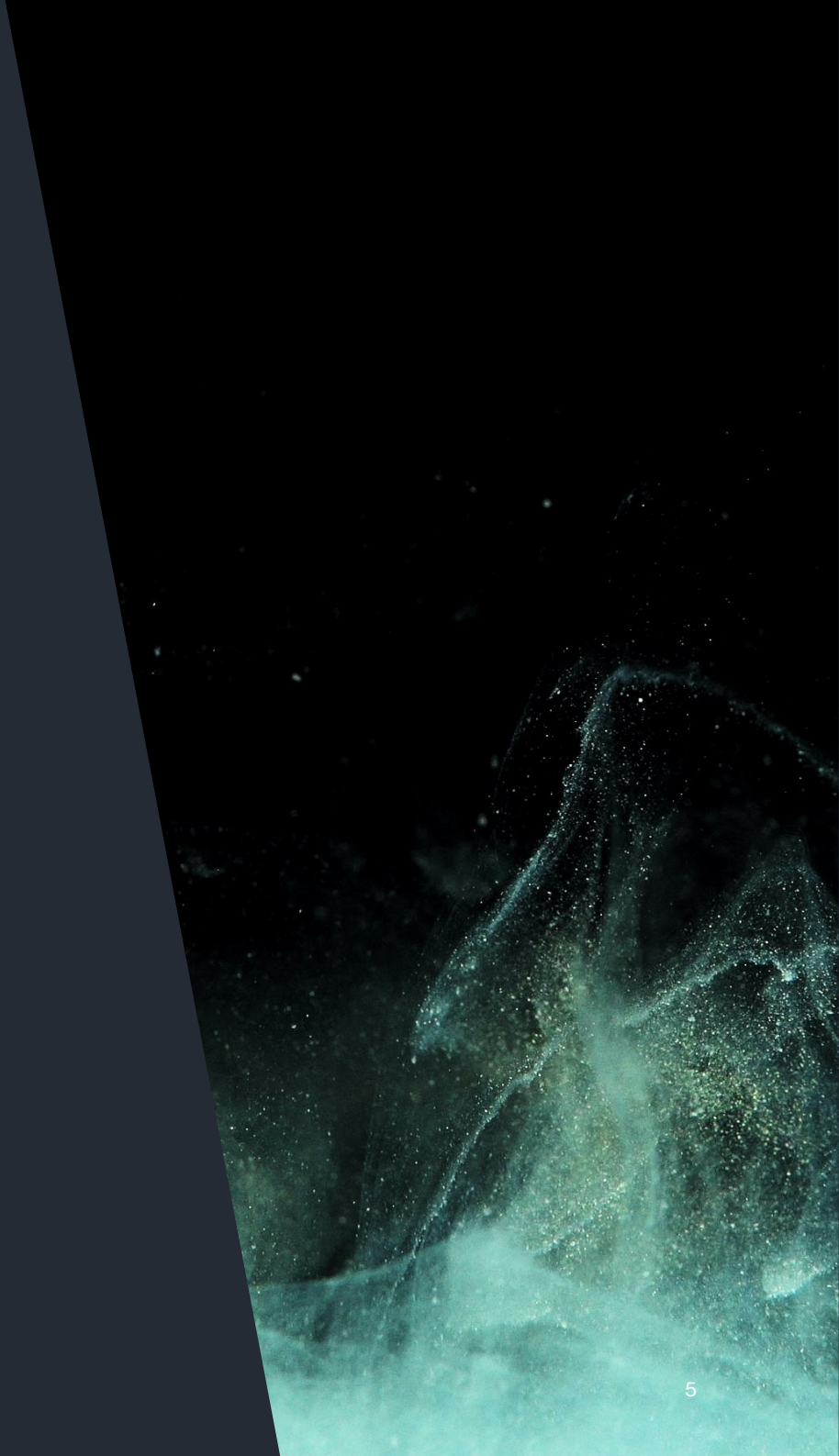
Beyond such operational responses, deeper engagement with communications service providers, creative industries, tech companies, broadcasters, publishers and media organisations will be required to understand and articulate emerging risks and concerns, and assess the requirement for any future policy and legislative interventions.

In addition to formal policy and legislation, organisations can also build partnerships to establish effective ethical standards and codes of conduct across professional bodies, companies and suppliers - providing assurances that their technology will be used for the public good and reducing the risk of misuse. Citizens may need to be engaged more directly through informed messaging and awareness raising campaigns (akin to existing cybersecurity messaging campaigns) to reduce susceptibility of target audiences, increase adoption and use of multi-factor authentication, and increase reporting of scams and fraud.

The UK and its allies cannot afford to simply accept these risks. A collaborative effort is now required between government, industry and academia to accelerate progress to detect, disrupt and mitigate the threats posed by AI voice cloning.

About the Author

Ant Burke is Co-Founder of Camulos.io, where his work focuses on supporting the adoption of data analytics and AI within the UK defence and security community. He previously worked within the Alan Turing Institute's Defence and Security programme in strategy and data science development roles, and at Raytheon UK, where he founded Raytheon's Strategic Research Group. Before this Ant spent over a decade in the UK Ministry of Defence and Dstl, in roles spanning operational analysis, threat and intelligence research and analytics technology development.



References

[i] BBC News. 'Fake Obama created using AI video tool.' Available at: https://www.youtube.com/watch?v=AmUC4m6w1wo/

[ii] Respeacher. 'Voice cloning for content creators.' Available at: https://www.respeecher.com/

[iii] Ministry of Defence. 2018. 'Joint Concept Note Human-Machine Teaming (JCN 1/18).' Available at: https://www.gov.uk/government/publications/human-machine-teaming-jcn-118/

[iv] Quote Investigator. 'Quote Investigator.' Available at: https://quoteinvestigator.com/2020/03/17/own-facts/

[v] GitHub. 'Voice-cloning.' Available at: https://github.com/topics/voice-cloning/

[vi] BBC News. 2021. 'Voice cloning of growing interest to actors and cybercriminals.' Available at: https://www.bbc.co.uk/news/business-57761873/

[vii] Venture Beat. 2020. 'Voice cloning experts cover crime, positive use cases and safeguards.' Available at: https://venturebeat.com/2020/01/29/ftc-voice-cloning-seminar-crime-use-cases-safeguards-ai-machine-learning/

[viii] Resemble. 'AI voice generator that sounds real.' Available at: https://www.resemble.ai/

[ix] Jia, Y., Zhang, Y., Weiss, R. J., Wang, Q., Shen, J., Ren, F., Chen, Z., Nguyen, P., Pang, R., Moreno, I. L., Wu, Y. 2018. 'Transfer learning from speaker verification to multispeaker text-to-speech synthesis.' Available at: https://paperswithcode.com/paper/transfer-learning-from-speaker-verification/

[x] Liao, Q. 2022. 'Try our custom neural voice in 5 minutes with a Lite project.' Available at: https://techcommunity.microsoft.com/t5/ai-cognitive-services-blog/try-out-custom-neural-voice-in-5-minutes-with-a-lite-project/ba-p/3270455/

[xi] GitHub. 'Resemblyzer.' Available at: https://github.com/resemble-ai/Resemblyzer/

[xii] ASV Spoof. 'ASVspoof 2021 workshop.' Available at: https://www.asvspoof.org/workshop/

[xiii] GitHub. 'ASVspoof 2021 baseline systems.' Available at: https://github.com/asvspoof-challenge/2021/

References

[xiv] IDR&D. 'Combat voice spoofing attacks.' Available at: https://www.idrnd.ai/voice-anti-spoofing/

[xv] Amazon Web Services. 'Machine learning on AWS.' Available at: https://aws.amazon.com/ai/

[xvi] Amazon Web Services. 'Amazon chime.' Available at: https://aws.amazon.com/chime/

[xvii] Edinburgh Data Share. 2019. 'ASVspoof 2019: The 3rd Automatic Speaker Verification Spoofing and Countermeasures Challenge database.' Available at: https://datashare.ed.ac.uk/handle/10283/3336