BCS' Response to the Speaker's Conference on the Security of Candidates, MPs and Elections Inquiry

Feb 2025

Compiled by Claire Penketh, Senior Policy and Public Affairs Manager

Table of Contents

Introduction

BCS was asked by the Speaker's Conference to contribute to a consultation that considers the factors influencing the threat level against candidates and MPs and the effectiveness of the response to such threats. The Speakers Conference will make recommendations about arrangements necessary to secure free and fair elections and the appropriate protection of candidates at future UK-wide parliamentary elections and of elected representatives thereafter. It has 15 members, including Mr Speaker in the Chair.

Research Method

BCS conducted a survey which resulted in just under 1500 respondents from the IT profession, asking for their reaction to the threat posed by AI-generated deepfakes and mis/dis-information in April 24 during the lead up to the UK General Election in 2024. [1]
We held a policy discussion webinar, convening senior experts in this field, also in April 24. [2]
The panel consisted of Lord Clement-Jones, Liberal Democrat Spokesperson for Science, Innovation and Technology, Hannah Perry, Head of Research (Digital Policy), Demos, Lisa Forte, Partner, Red Goat Cybersecurity, and Tom Bristow, Tech Reporter, Politico.
Our Policy Team also conducted interviews with the following senior BCS experts:

- Professor Andy Phippen, Professor of IT Ethics and Digital Rights at Bournemouth University
- Professor James Davenport, Hebron and Medlock Professor of Information Technology, Department of Computer Science; International Centre for Higher Education Management (ICHEM)

Our response has considered three areas:

i) **How does the prevalence of mis-/disinformation during election periods affect the risks to candidates? And are there sufficient measures in place to identify, tackle, and deter such material?**

ii) **To what extent do foreign state actors influence the nature and level of threats to candidates and MPs? And what steps can be taken to track and mitigate such influence?**

iii) **How are technology and threats likely to evolve, and what more is needed?**

[1] Deepfakes, AI and the General Election https://www.bcs.org/articles-opinion-and-research/deepfakes-a-major-risk-for-the-general-election-according-to-research-with-the-tech-profession/
[2] https://www.bcs.org/articles-opinion-and-research/deep-fakes-and-elections-bcs-policy-jam-april-2024/

BCS AI deepfakes Survey Results

- **65% of tech professionals believed AI-generated audio and video, dis-/mis information could influence the outcome of the poll.**
- **public education and technical solutions, such as watermarking and labelling, would be the two most effective measures for limiting the detrimental impact of deep fakes on democracy.**
- **92% of technologists said political parties should agree to publicise where and how they use AI in their campaigns**.
- Only **8% were optimistic that a pact signed by several major tech companies would be effective**. The firms had agreed in February 2024 to adopt 'reasonable precautions' to prevent AI from being used to disrupt democratic elections around the world in the Tech Accord to Combat Deceptive Use of AI in 2024 Elections.[3]

During the BCS policy discussion webinar Lord Clement-Jones said: "We're in a bit of a crucible now for this kind of technology, and I think that it's unfortunate we weren't able to anticipate it would be in such prolific use before this year started."
He was referring to the explosion of audio and video deepfakes circulating on social media and encrypted messenger apps. In 2018, there were a few thousand doing the rounds— now, it is in the billions. But so far **evidence that it has affected UK elections is lacking** – but there appears to be consensus it **has the potential** to disrupt elections.

A post-election report by the Alan Turing Institute, **AI-Enabled Influence Operations: Threat Analysis of the 2024 UK and European Elections.***"* [2] concluded that while disinformation or deepfake news had limited influence on these elections, the trend has the potential to undermine democracy**:** [4]

"There is no evidence that AI enabled misinformation meaningfully impacted recent UK or European election results However, concerns remain about disinformation damaging the integrity of the democratic system and new risks posed by parody or pornographic deepfakes.

"Researchers from the Centre for Emerging Technology and Security (CETaS) at the Alan Turing Institute identified just 16 confirmed viral cases of AI disinformation or deepfakes during the UK general election, while only 11 viral cases were identified in the EU and French elections combined.

"Despite reassuring findings about the impact of AI on election results in line with previous Turing research, there are emerging concerns about instances of realistic parody or satire

[3] https://www.bbc.co.uk/news/technology-68316683
[4] https://www.turing.ac.uk/news/no-evidence-ai-disinformation-or-deepfakes-impacted-uk-french-or-european-elections-results

deepfakes which, while intended as humour, can include misleading election claims that some voters interpret as factual."

A report from Demos, entitled *'***Synthetic Politics: Preparing Democracy for Generative AI'**[5] also recognised the potential threat: *Public facing generative AI tools have the potential to change what and how content is created, and how it enters and spreads around the online world.*

These changes to the information environment have particular implications for democratic integrity: in the effects they have on core democratic ideals of equality, truth and non-violence in political discourse.

How far-reaching these effects will be - and how much policy attention they should capture - is contested.

This report lists a series of actions that should be "*urgently put in place to reduce the acute risks to democratic integrity presented by generative AI tools."* These include AI developers setting clear policies concerning the content that users may and may not generate, especially with respect to content that undermines democratic integrity and watermarking AI-generated content where feasible and warning users about AI inaccuracies.

It recommended social media platforms should double down on the enforcement of rules against harmful speech by removing content that breaches their policies regardless of whether it is generated by human or machine.

The role of media literacy

Returning to the BCS survey of IT professionals, respondents identified the two most effective measures for countering the impact of deep fakes as public education and technical solutions (e.g. watermarking and labelling).
We'll look at the technical solutions shortly – but first turn to the role of public education and digital/media literacy. Ofcom has issued its 'Three-Year Media Literacy Strategy' and BCS Fellow Andy Phippen, Professor of IT Ethics and Digital Rights at Bournemouth University gave this analysis: "While alignment with OFCOM's media literacy strategy is fine, I would call on the regulator to take more of a role in shaping digital literacy nationally. "We know that education related to digital literacy and online safety drops off a cliff in secondary school, and unless pupils chose to study Computer Science, there is nothing in the national curriculum that requires young people to learn about these issues after about aged 13. This needs to change - a more informed population is a more resilient one."

These points were highlighted in a report by Professor Phippen: **Evaluation of the ProjectEVOLVE Database Understanding Online Safety Delivery and Assessment in Schools,** which shows 'a need for increased engagement in secondary schools, especially on more technical and complex topics'. [6]

[5] https://demos.co.uk/wp-content/uploads/2024/03/Synthetic-Politics_Report-1.pdf
[6] https://swgfl.org.uk/research/evaluation-of-the-projectevolve-database-2024/

In our Policy Discussion webinar referenced above, Hannah Perry, from Demos also believed digital literacy education had to start early in education, in Personal, social, health and economic (PSHE) education, which is a non-statutory subject. On this point, Professor James Davenport said media literacy should be strengthen in Citizenship education, a required subject in the National Curriculum at key stage 3 (age 11-14) and 4 (age 11-16).

Technical solutions

The exponential growth of deepfakes has been aided by the accessibility of cheap, easy-to-use AI software to create the material. At the same time, the technical solutions to counter the problem are still trying to catch up.

One of the technical solutions frequently proposed **is watermarking images**. Professor Davenport has provided testimony to suggest that these digital markers are normally thought of as stamps that content has been AI-generated, rather than a guarantee that the photograph or image is genuine and unaltered.

Speaking to BCS on this subject he said: "**There is much talk about 'watermarking AI', but this is probably impractical, and certainly impossible to enforce – the genie is well and truly out of the bottle. What is really needed is watermarking 'guaranteed originals', which is certainly technically possible.**"

There have been moves to do this. The Coalition for Content Provenance and Authenticity (C2PA), formed by Adobe, Arm, Intel, Microsoft, and Truepic, was set up in 2021 to 'address the prevalence of misleading information online through the development of technical standards for certifying the source and history (or provenance) of media content.' Members like Adobe and Microsoft have adopted Content Credentials, embedding metadata—such as the creator's identity and the software used—into images and videos. [7]Google has joined the steering committee of C2PA and says it is 'investing heavily in tools and innovative solutions, like SynthID, to provide it. [8]

But these watermarks are often embedded invisibly, which is good for a copyright detection system – but, can be missed by fallible human eyes viewing the deepfakes[9]. Plus, some forms of digital watermarks are not foolproof: they can be destroyed with fake digital markers inserted, according to a report by the University of Maryland.[10]

In the United States, President Biden issued an Executive Order during his term, calling for a new set of government-led standards on watermarking AI-generated content; however, there was no requirement for tech firms, or the US government, to use them and therefore

[7] https://c2pa.org/
[8] https://blog.google/technology/ai/google-gen-ai-content-transparency-c2pa/
[9] https://www.theverge.com/2024/2/13/24067991/watermark-generative-ai-deepfake-copyright
[10] https://www.cs.umd.edu/article/2023/11/watermarks-aren%E2%80%99t-silver-bullet-ai-misinformation

some critics argued this made them ineffectual. [11] Biden's Executive Order on this was subsequently repealed by President Trump when he came into office. [12]

In September 2024 Meta announced it will be providing 'AI info' labels on its platforms "*so they better reflect the extent of AI used in content*.' [13]

Our members are cynical about 'Big Tech' firms resolving this problem. In February 2024 Google, Meta, Microsoft, OpenAI and TikTok signed a pact to take reasonable precautions to stop AI tools from being used to disrupt democratic elections worldwide. Only 8% of our members surveyed believed the Tech Accord to Combat Deceptive Use of AI in 2024 Elections would be effective.

However, the UK government could show its support for watermarking standards of original content and for the labelling of AI generated content.

Social media, the role of algorithms and friction

The April 2024 BCS Policy Discussion addressed the a tech solution to **slow the flow of fake news,** which, in effect, introduces 'friction'.
Lord Clement-Jones provided the following testimony: "We talked about this when we did the Online Safety Joint Committee scrutiny of the draft Bill because one of the real difficulties is the amplification of content. But I haven't actually seen a tool that does that, and I'd be really interested to see something along those lines.
"I don't think there's any silver bullet, and I don't think regulation is necessarily a silver bullet, but the combination of all this might slow down people's ability to spread this kind of information."

There was some consensus from the panellists that a label identifying when a shared post had not been read by the author could be useful in preventing the spread of misinformation. Social Media companies and their desire for viral content was identified as a likely blocker to this policy.

The anger factor

On Jan 30 2025 Politico published an article identifying that, despite making up only 3.5%of the total number of posts sent by Westminster parliamentarians, each Reform post has on average 6,300 engagements, compared to just 500 for most other UK MPs. As well as exposure, there is a financial incentive, claims Politico, because Elon Musk has recently

[11] https://www.theguardian.com/technology/2023/oct/30/biden-orders-tech-firms-to-share-ai-safety-test-results-with-us-government
[12] https://www.reuters.com/technology/artificial-intelligence/trump-revokes-biden-executive-order-addressing-ai-risks-2025-01-21
[13] https://about.fb.com/news/2024/04/metas-approach-to-labeling-ai-generated-content-and-manipulated-media/

given premium users — who meet certain requirements — the ability to monetise their accounts. That could, the article argues, lead to an incentive to post incendiary material. [14] BCS Fellow and Cybersecurity expert Professor Victoria Baines is well placed to comment on this as the former Trust and Safety Manager at Facebook.

She warned that Musk's decision to replace the platform's human-led trust and safety team with AI moderation has undermined the system. She was quoted as saying inflammatory post are common, despite the guidelines that are still in place: "They still include hateful activity, harassment and insulting content related to race, ethnicity, gender and sexual orientation, but that is exactly the kind of content being shared by U.K. and U.S. politicians right now, and by some of these paid amplifiers."

It's not difficult to see the very real-world impact that misinformation on social media can have on politicians, and not necessarily during an election campaign. For instance, Jess Philips MP, Minister for Safeguarding was called a "rape genocide apologist" by Elon Musk on his social media platform X following the UK government's decision not to hold another National Inquiry into grooming gangs. The consequences for Ms Philips were threats which resulted in her personal protection having to be increased.

Other examples include how the use of deepfakes can affect anyone in the political sphere. A recent case where a volunteer Labour Party canvasser, who was a teacher, was forced into hiding following fake claims she was racist.[15]

The Online Safety Act is due to come into force shortly and in December 2024, Ofcom published its first codes of practice and guidance and further iterations are promised. There are calls that the government should consider looking at is strengthening Ofcom's powers to directly tackling the role of deepfakes, and mis/disinformation in elections as part of the ongoing updating of the Act.

For instance, Full Fact, independent fact-checkers, have campaigned for a move "*towards a regime whereby internet companies have a legal duty to tackle the full range of misleading and harmful information spreading on their platforms.*" [16]

However, Professor Phippen says that could be problematic: "Including deep fakes and disinformation in a strengthened Online Safety Act would encounter the same issues as other "legal but harmful" content. **If you can't define it in law, you can't expect an algorithm to spot it."**

Tougher penalties to protect politicians

[14] https://www.politico.eu/newsletter/politico-london-influence/the-age-of-the-shtposter/

[15] https://www.theguardian.com/education/2025/jan/19/teacher-was-forced-into-hiding-after-fake-video-appeared-to-show-her-making-racist-slur

[16] https://fullfact.org/blog/2024/oct/online-safety-act-should-help-fact-checkers-on-misinformation/

With more than half of political candidates reporting abuse, the Electoral Commission wants to see tougher penalties for criminal acts, tighter political party membership rules, including potential deselection.
In the above BCS survey, respondents were overwhelming in favour**, at 92% , of a recommendation that political parties should agree to publicise where and how they use AI in their campaigns**. This shows the respondents felt political parties, organisations and independent candidates have a duty to play their part.

Foreign Influence

A significant number of disinformation campaigns originate from foreign state and non-state actors. These actors leverage messaging platforms like WhatsApp and social media to spread deepfakes and mis and disinformation, as seen in elections including the US,[17] India, [18] Pakistan, Slovakia[19], and notably in Romania,[20] where an election was cancelled amid allegations of Russian interference. Such tactics aim not only to sway voter opinion but also to destabilise trust in democratic institutions.

Challenges in Mitigation

Tracking foreign influence is difficult due to the decentralised nature of these campaigns and the use of encrypted communication platforms. Efforts to enforce global cooperation on watermarking or labelling content face resistance from non-compliant states and actors.

Here are the key recommendations for the government based on the BCS survey of our members, and further desk-top research:

Addressing Misinformation and Disinformation During Elections and Mitigating Technological Evolution

- **Enhance Public Education & Digital Literacy**
    - **Integrate and strengthen media literacy into school curricula** in Citizenship and Personal, Social, Health and Economic Education lessons to help individuals identify misinformation, as supported by Ofcom's Media Literacy Strategy.
    - Launch nationwide awareness campaigns to **educate the public on deepfakes** and AI-generated disinformation.
- **Strengthen Technical Solutions**

[17] https://www.brookings.edu/articles/how-disinformation-defined-the-2024-election-narrative/
[18] https://www.weforum.org/stories/2024/08/deepfakes-india-tackling-ai-generated-misinformation-elections/
[19] https://ec.europa.eu/newsroom/edmo/newsletter-archives/52231
[20] https://www.journalofdemocracy.org/online-exclusive/why-romania-just-canceled-its-presidential-election/

  - Encourage the adoption of watermarking and AI detection tools for digital content verification.
  - Promote transparency by requiring **political parties to disclose the use of AI** in campaigns (**supported by 92% of BCS survey respondents**).

- **Regulation & Platform Accountability**
  - Support friction measures on social media platforms to reduce the spread of fake news, slowing virality of suspect content.
  - Consider strengthening the Online Safety Act to specifically address deepfakes and disinformation.
  - Look at giving Ofcom increased regulatory power to enforce platform accountability on the issue of mis/disinformation.
- **Legal & Law Enforcement Measures**
  - Consider increasing penalties for those who deliberately spread harmful political disinformation.
  - Consider legal requirements for political parties to disclose AI-generated campaign materials.

Tackling Foreign State Influence on Election Integrity

- **Enhanced Detection & Monitoring**
  - Bolster intelligence-sharing mechanisms to track and address threats from state actors.
  - Invest in AI-powered monitoring tools to track foreign disinformation campaigns.
- **Legislative & Diplomatic Actions**
  - Consider introducing sanctions or penalties for foreign entities found to be spreading election-related disinformation.
  - Establish an international framework for combatting election interference through collaboration with democratic allies.
- **Platform Cooperation & Data Transparency**
  - Support mandatory content provenance and watermarking standards for AI-generated media.
  - Work with tech companies to identify and limit the spread of state-sponsored misinformation campaigns.

Conclusion

Following extensive original research, complemented with in-depth interviews with technological experts, BCS has formed the conclusion that a **multi-layered approac**h that involves **legislation, technology, international cooperation, and public education** is necessary to safeguard democracy from AI-driven misinformation and foreign interference. Proactive measures will strengthen electoral integrity and public trust in democratic institutions and help bolster the safety of politicians.

BCS would like to thank the Speaker's Conference for the opportunity to engage on this important issue and would be happy to contribute further.

Who we are

BCS is the UK's Chartered Institute for Information Technology. The purpose of BCS as defined by its Royal Charter is to promote and advance the education and practice of computing for the benefit of the public.

We bring together industry, academics, practitioners, and government to share knowledge, promote new thinking, inform the design of new curricula, shape public policy and inform the public.

As the professional membership and accreditation body for Information Technology we serve over 70,000 members including practitioners, businesses, academics, and students, in the UK and internationally. We also have over fifty specialist groups

We also accredit the computing degree courses in over ninety universities around the UK. As a leading information technology qualification body, we offer a range of widely recognised professional and end-user qualifications.

BCS
The Chartered Institute for IT
3 Newbridge House,
Newbridge Square,
Swindon SN1 1BY
BCS is a registered charity: No 292786