Cymdeithas Seneddol y Gymanwlad
Senedd Cymru

Commonwealth Parliamentary Association
Welsh Parliament

Report by Lee Waters MS

3rd Commonwealth Parliamentary Association Conference:

Artificial Intelligence and Disinformation: 'Democracy in the age of deepfakes'

Singapore

18-20 June 2024.

Quick takes

1. Artificial intelligence is a tool. It has both malign and benign capabilities depending on how it's designed and used. We can't stop it, but we can try and shape it.
2. Sophistication of A.I is changing rapidly and is outpacing regulation or detection technology. But awareness of the pace of change is low.
3. Significant potential to use A.I to boost productivity and create economic growth. Jobs will be replaced by more efficient technology and new jobs and areas of activity will also be created.
4. Active role for Government to manage a 'just transition' in economy, make sure benefits are spread, people are re-skilled and regulation used to ensure fairness and ethics in how A.I is applied - BUT technology is outpacing the ability of the State to respond.
5. Singapore taking a smart strategic approach to lead the field in A.I. They are taking steps to get out in front of it and much we can learn from and collaborate with.
6. There is widespread use of hyper-real fake images for fraud (insurance claims, fake I.D), and in pornography where use of deepfakes without consent is a new form of violence against women.
7. Very difficult to tell whether text has been generated by A.I; only 50/50 chance on noticing whether a picture is real or fake, and even close family members can't tell difference between real or deepfake audio.
8. Deepfake A.I is already being used to manipulate elections and speak misinformation in elections.
9. Estimated that within three years 90% of all online content will be manipulated in whole or in part.
10. EU attempting to regulate (as it did with GDPR) but the five Big Tech firms that dominate are resisting. China leading on baking-in AI standards in goods it produces, e.g autonomous vehicles, to mould the approaches taken in the market.

SIT UP AND TAKE NOTICE!

In the recent Indian elections the daughter of a Tamil Tiger militant chief gave a speech, live streamed on YouTube, urging Tamilians across the world to take up political struggle for their freedom.

So what?

She died 14 years before her 'live' broadcast and is not known to have said any of those things.

Audio messages from Joe Biden were used to discourage supporters from voting in this year's New Hampshire Primary. It was not him, but a hyper-realistic recording using an audio clone of the President to suppress voter turnout to the benefit of his opponents.

Elections in Slovakia were influenced by fake audio of one of the top candidates talking about raising the cost of beer!

And in this year's London's Mayoral election hundreds of thousands of people heard audio of Sadiq Khan making inflammatory remarks before Armistice Day that almost caused "serious disorder". It was fake.

Senior national security officials in the US have been gearing up for "deepfakes" to inject confusion among voters in a way not previously seen. US authorities are involved in contingency planning for a foreign government potentially using AI to interfere in the Presidential election.

Experts predict that within three years 90% of all online content will be manipulated in whole or in part.

SEEING IS BELIEVING? RIGHT?

Um, no.

The sophistication of AI technology is growing exponentially, and in parallel, the costs and barriers to entry are falling.

In just a few years artificial intelligence has leapfrogged from 'narrow AI' that organises existing content, to 'generative' AI that can turn text inputs into an image, turn an image into a song, or turn video into text.

The creation of A.I Deepfakes (a blend of 'deep learning' and 'fake') is allowing ultra-realistic fake videos to be easily made which can depict people doing things they have never done or never said. The idea that 'seeing is believing' is no longer true.

It is no longer possible to tell whether text is generated by artificial intelligence or not. The use of Large Language Models (LLMs), like ChatGBT, are also able to generate student essays that "verged on being undetectable", with 94% not raising concerns with markers.

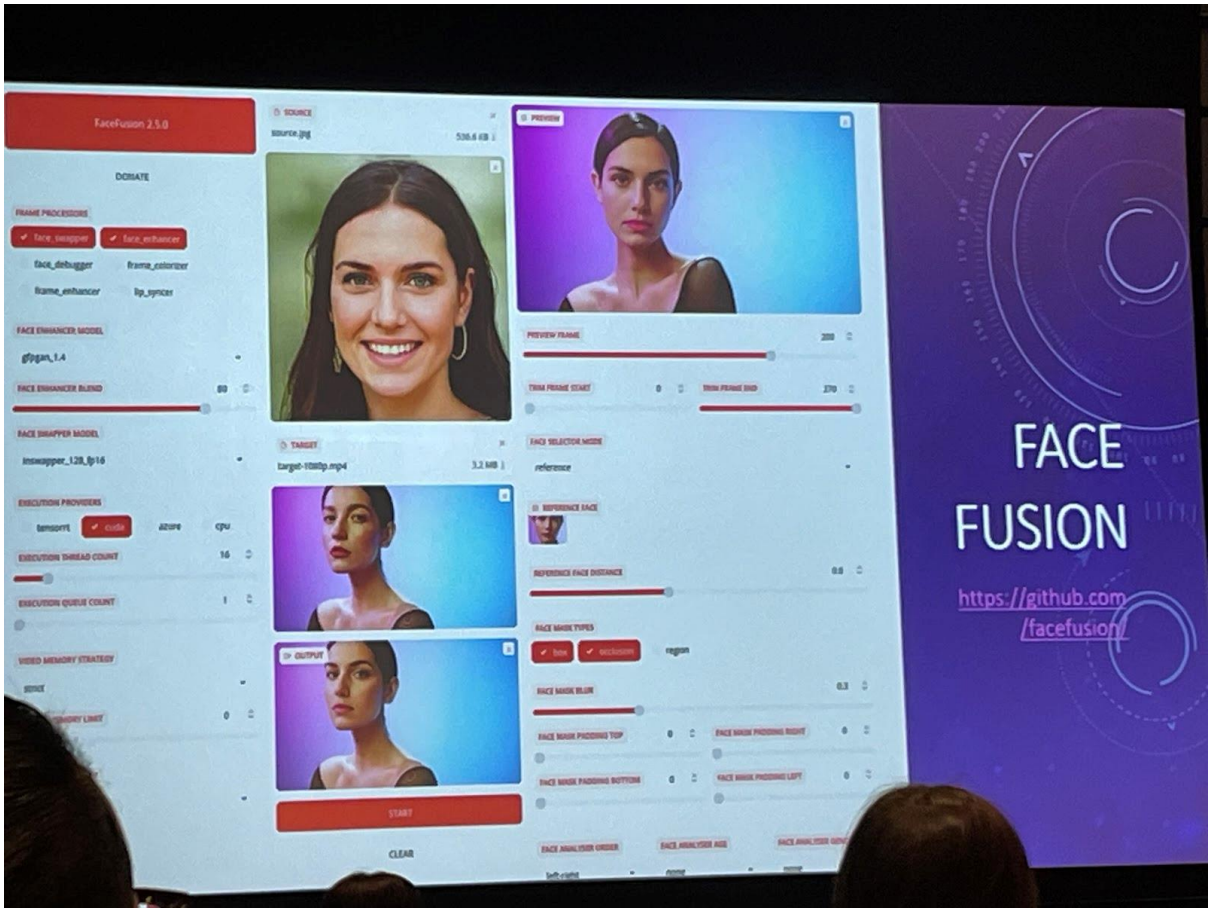
There is just a 50/50 chance whether someone can tell the difference between a real or fake picture.

And even close family members can't detect whether it really is you making a phone call or leaving an audio message

False rumours, misinformation and disinformation are not new in elections. But social media and messaging Apps are giving them an unprecedented fast delivery mechanisms at scale.

Deepfakes are difficult to track, have a wide reach and there is low public awareness of the trickery which allows people to fall for it. And such is the rapid pace of developments in so-called 'Generative AI' that law and regulation are unable to keep up, moreover our own senses and instincts are failing us too.

All you need is a sample of just three seconds of audio or a single image and Microsoft's VASA "is capable of not only producing lip movements that are exquisitely synchronized with the audio, but also capturing a large spectrum of facial nuances and natural head motions that contribute to the perception of authenticity and liveliness".



‘BAD ACTORS’

The scale and sophistication of the tech has reached the point where it can overwhelm ‘fact-checking’ capabilities.

Current technology can't detect deepfakes that have been created by techniques it wasn’t trained on, and struggles to distinguish between real and fake content; the emerging detection technology generates a high number of false alarms.

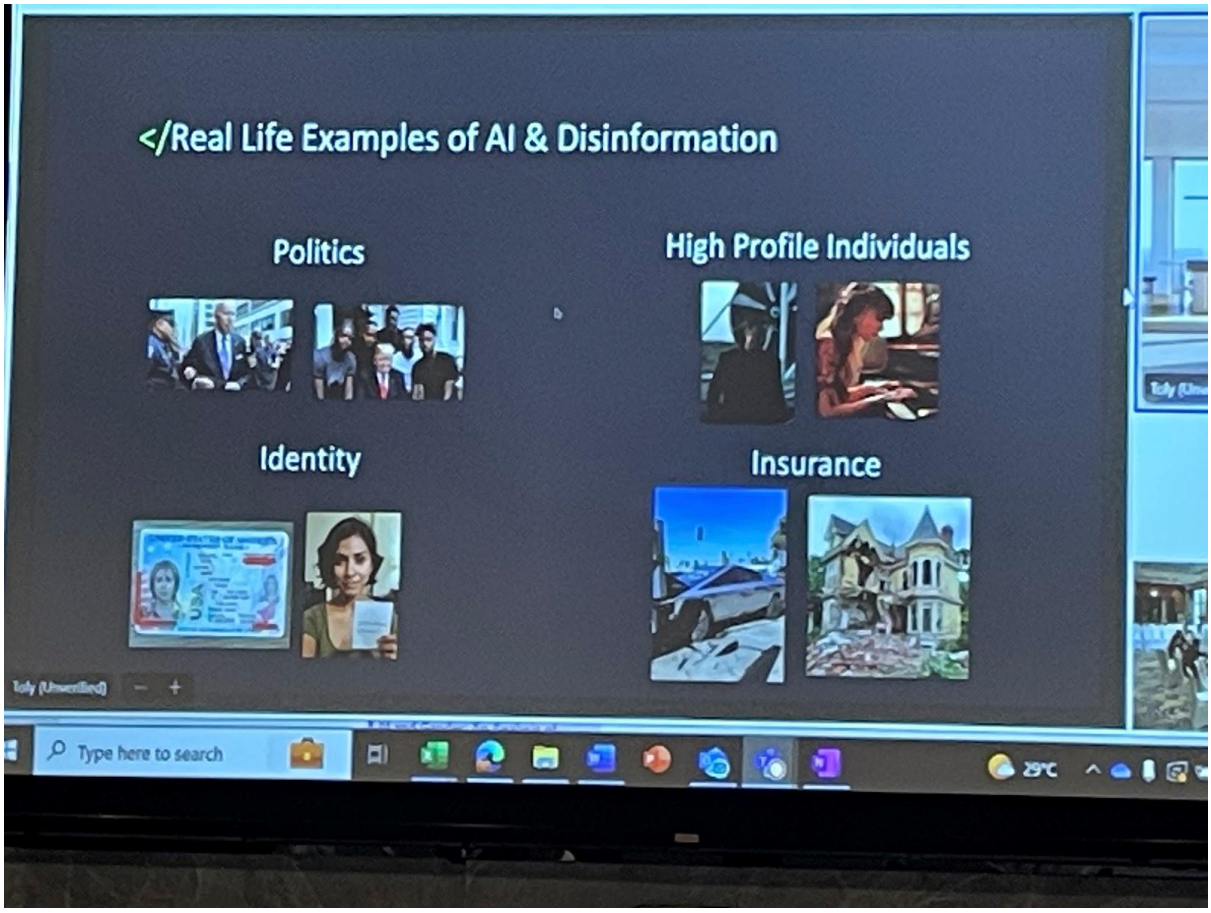
It also has clear criminal applications; Deepfake images can be used to trick ‘on-boarding’ processes such as passport applications, online bank anti-fraud measures, and false insurance claims.

It can also be used on Zoom / Teams calls to enable someone to pretend to be someone else!

This will lead to people questioning what they see and what they can trust. This can ‘poison the well’ and result in what’s called the ‘liar’s dividend’, where scepticism around deepfakes is used to cast doubt on genuine evidence. For instance, two defendants on trial for the January 6 attack on the U.S capital attempted to argue a video showing them at the Capitol on the basis that it could have been AI-generated.

This will foster cynicism and undermine a collective sense of reality. What is real? What is fake? Who can be trusted?

Experts warn that the persuasiveness of the hyper-real - but fake - content, and the fact it can be personalised and micro-targeted, is likely to be used to ‘preach to the converted’ and deepen the pre-existing tendency towards confirmation bias. This could reinforce the growing sense of polarisation and ideological divides.



DEEPENING BIAS

There is a particular gendered element to the development of A.I too. Only 12% of AI researchers and 6% of professional software developers in AI are women. The lack of diversity in AI development teams contributes to biased algorithms. A study analyzing 133 AI systems across different industries found that about 44% of them exhibited gender bias, and 25% showed both gender and racial bias.

Men are more likely to use A.I in their professional or personal lives, (54% of men use AI only 35% of women do so); and women may be more reluctant to use AI tools due to concerns about trust, accuracy, and plagiarism.

Researchers found that many women choose to limit their online visibility or withdraw entirely to protect themselves, which can hinder career advancement, networking opportunities, and professional growth. Such abuse also causes many

women to avoid the public sphere, reducing women's voices in political discussions and leading to less representative and inclusive decision-making.

AI algorithms learn from patterns in the data, AI models tend to use the majority as the reference point to the disadvantage of minority groups. Work is underway to research effective ways of addressing the issue of 'Algorithmic fairness', but academics warn is way involve trade offs between accuracy and fairness.

There's a growing concern about the use of deepfakes beng used as a form of violence against women. It is estimated that of all deepfake videos, 98% were pornographic and of those 99% were of women. This 'non-consensual porn' causes as much harm as the non-consensual distribution of intimate images.

WE NEED RULES

There are efforts underway to regulate the use of artificial intelligence.

Some of the tech giants are pushing back and prefer self-regulation. This is critical because just 5 big tech firms dominate the landscape - Apple, Meta (Facebook), Alphabet (Google), Amazon, Microsoft).

The EU have passed a sweeping piece of AI regulation to try and influence global standards, just as it did successfully with GDPR. China has made significant inroads in moulding approaches. Chinese innovation on audit and disclosure around AI, as well as the standards baked into its AI-enabled technology exports like autonomous vehicles or digital tutors, are already globally significant.

To help democratic institutions respond the Organization of American States (OAS) and the Commonwealth Parliamentary Association have developed a Parliamentary Handbook on Disinformation, AI and Synthetic Media The handbook contains strategies for combating disinformation and guidance on how Parliamentarians can work with civil society, the media and technology companies, to develop regulatory/legislative frameworks to address the challenges of disinformation, as well as how they can take steps to safeguard their own online profiles and communication channels.

ITS NOT ALL BAD

Microsoft acknowledges its Deepfake A.I - VASA - can be misused, but says it also has 'substantial positive potential':

"The benefits – such as enhancing educational equity, improving accessibility for individuals with communication challenges, offering companionship or therapeutic support to those in need, among many others – underscore the importance of our research and other related explorations".

In the recent Indian elections as well as spreading fake information A.I bots also helped reduce linguistic barriers by allowing candidates to reach more voters who speak one of India's many regional languages.

Prime Minister Modi for instance, used Bhashini, the government's AI-powered tool, to ensure that Tamil-speaking audiences could hear his speech, which was delivered in Hindi and translated into Tamil in real-time. His speeches have also been translated into Kannada, Bengali, Telugu, Odia, and Malayalam, among other languages, using AI. The prime minister's official app — NaMo — has launched a

feature designed to promote the government's policy successes more widely through AI-powered chatbots.

Like any tool it has both benign and malign uses.

ITS THE ECONOMY, STUPID

Singapore is taking a strategic view about the economic potential of embracing the AI revolution. It launched a national R&D programme 2017 and an arms-length institute - AI Singapore - which is the first national level approach to build deep national capabilities in AI. They are focus on six strategic areas:

- Research - developing local talent and applying AI to real-world problems
- Technology - supporting high-impact projects that meet national challenges
- Innovation - Spur and support widespread industry adoption of AI
- Products - Build practical applications in key development areas
- Governance - Research in governance, ethics and accountability of AI systems

The approach focuses heavily on applied research alongside the private sector to real-world business challenges, and developing a pipeline of talent in the country - including an intensive paid Masters course to grow the talent pool. The R&D effort is focused around a small number of strategic challenges.

Some practical examples of A.I projects they have worked on with SMEs:

- Developed AI model to assist dentists by automating x-ray charting process, improving accuracy and efficiency. This innovation allows dentists to focus more time on patient care.
- Co-produced a model to enhance delivery efficiency for a same-day parcel startup. The AI model improved route efficiency by 20% and streamlined operations, enhancing service quality and reducing costs.
- Speeding up cell analysis for hamessing MicroAlgae in plant-based food innovation. Counting cells manually under a microscope to check for harvest readiness is very time consuming (30-40 mins per slide). The collaboration led to the development of an AI tool to slash analysis time by 30x.
- Speech Recognition Model for emergency calls which helps operators by transcribing in real-time, in multiple languages to optimise resource allocation and reducing response times.

- Collaboration with pathologists to enhance breast tumour diagnosis, resulting in a bespoke 2-stage CV model trained on tissue images. Achieving 87.5% accuracy, the tool assists pathologists, cuts costs and surgical interventions while easing patient anxiety.

A full account of the approach Singapore has followed has followed has been published as an e-book: AI-First Nation by Laurence Liew.

Appropriately he used Gemini, Google's version of Chat GBT, to ask him questions which then formed the basis of the book - AI has biographer, as he put it.

LETS TALK

There are those who think we should stop further development of A.I.

Richard Heinberg, a leading thinker on sustainability, has argued that 'Artificial Intelligence Must Be Stopped Now':

Whenever a new technology is introduced, the usual practice is to wait and see its positive and negative outcomes before implementing regulations. But if we wait until AI has developed further, we will no longer be in charge. We may find it impossible to regain control of the technology we have created.

I think that is fanciful.

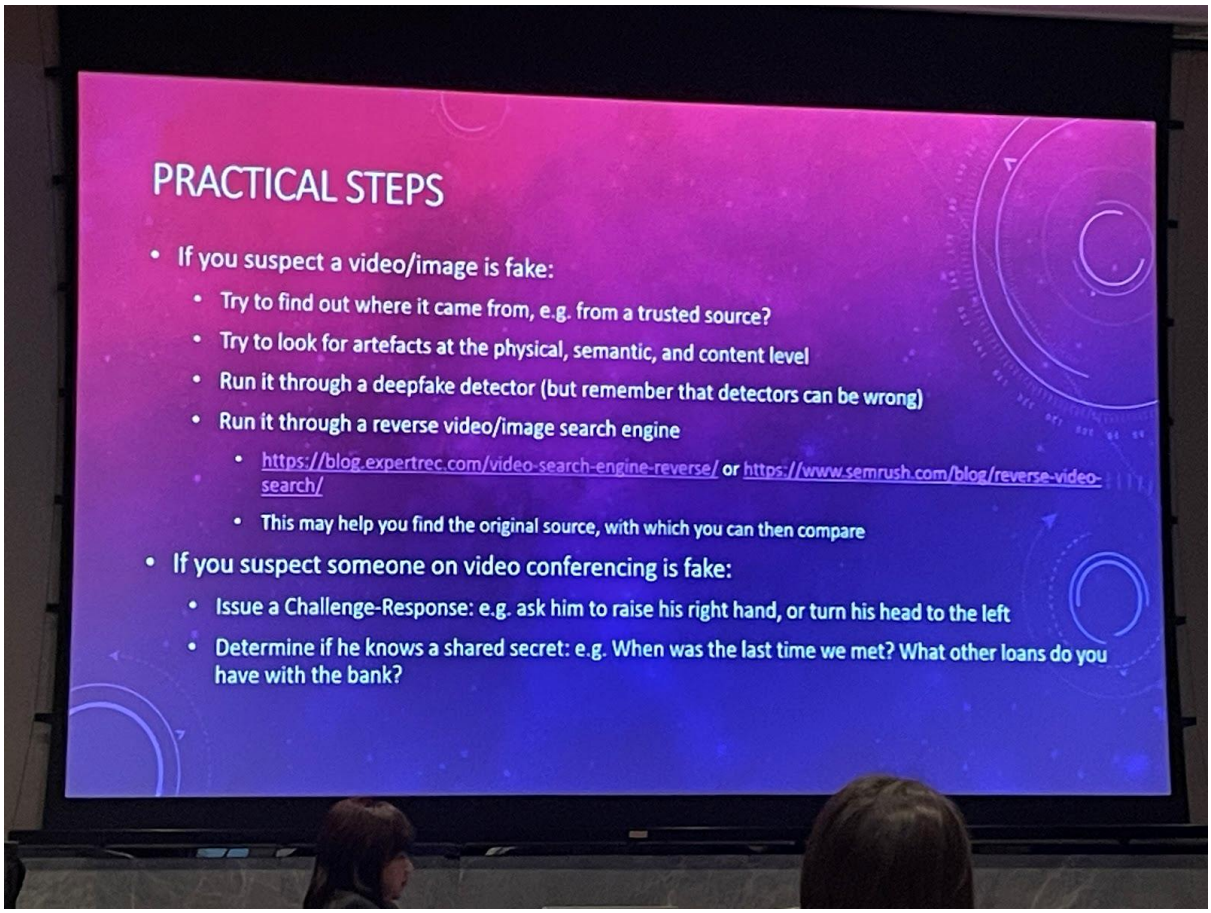
My take from the expert testimony from experts at the conference from Singapore, the UNDP, Australia and the UK is that these leaps in technology represent both threats and opportunities. Both technology and regulation will struggle to keep it in check in the short-term, and one of the most important things we can do to raise awareness about what AI is capable of, and how it is manifesting itself in our lives and in our politics.

Technical approaches such as authenticating content with watermarks, or developing technology that can check the provenance of content is all capable of being tricked. As that continues to develop we all need to socialise voters to start questioning what they see and hear.

The consensus was that:

- Democratic institutions need to evolve as technology evolves.
- There must be a public conversation about how societies can respond to transformative artificial intelligence.

- We need global collaboration to ensure AI is safe, and to achieve equitable benefits.



The 3rd Commonwealth Parliamentary Association Conference on Artificial Intelligence and Disinformation: 'Democracy in the age of deepfakes' took place in Singapore between 18-20 June 2024.

Delegates attended from Borno (Nigeria); British Virgin Islands; Cameroon; Ghana; Jamaica; Kenya; Malawi; Namibia; New South Wales (Australia); Nigeria; Penang (Malaysia); Queensland (Australia); Singapore; Sri Lanka; United Kingdom; Victoria (Australia); Wales; Western Australia; Zanzibar.

Lee Waters MS,
Senedd CPA Branch delegate.
July 2024