TechScape: Could a Labour ‘nudification’ manifesto bring more safety to AI?

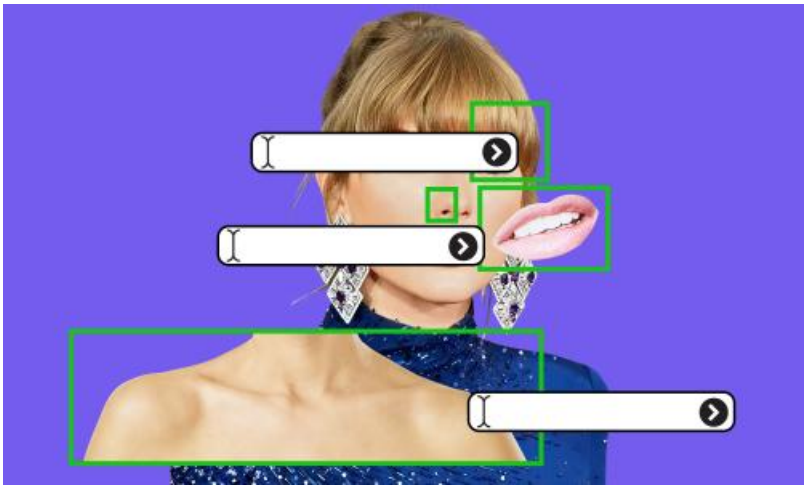
Publication Date: 2024-03-19

Author: Alex Hern

Section: Technology

Tags: Technology, TechScape newsletter, Deepfake, Artificial intelligence (AI), Labour, Conservatives, newsletters

Article URL: https://www.theguardian.com/technology/2024/mar/19/techscape-labour-thinktank-ai-regulation-deepfake-nudification



The politics of AI regulation became a little clearer this weekend, after an influential Labour thinktank laid out its framework for how the party should approach the topic in its manifesto. From our story: The policy paper, produced by the centre-left Labour Together thinktank, proposes a legal ban on dedicated nudification tools that allow users to generate explicit content by uploading images of real people. It would also create an obligation for developers of general-purpose AI tools and web hosting companies to take reasonable steps to ensure they are not involved in the production of such images, or other harmful deepfakes. Labour Together’s suggestions aren’t party policy yet, but they point at the sort of issues Westminster wonks think a campaign can be built on. (If you want to read the tea leaves though, Peter Kyle, the shadow technology minister, said he was “considering proposals carefully”.) For the last few decades, technology has been a curiously apolitical realm in the UK, with all parties agreeing on the vague idea that it’s important to support British technology as a driver of growth and soft power, and little active campaigning beyond that. Even when tech regulation has become a top-level political concern, starting with Theresa May’s government and the introduction of the online safety act, the debates around it have tended towards technocratic rather than principled or partisan. Labour forced some votes over specific amendments to the bill, but when push came to shove it was passed unopposed. In hindsight, that bill’s most consequential battle was within the Tory party itself, as one wing decided to attack the entire process as an attempt to ban “hurt feelings”, in part due to clauses aiming to replace the old “malicious communications” offence with more specific crimes. This time, though, things might be different. If Labour does propose a ban on nudification tools, it seems unlikely to be simply co-opted by the Conservative Party. Instead, it could highlight a divide between the two parties’ concerns around AI, with Rishi Sunak leading the Tories to focus on Silicon Valley-inflected concerns around existential risk, and Labour focusing on misuse in the here and now. ‘MrDeepFakes doesn’t represent tech’ I spoke to the paper’s authors, Kirsty Innes and Laurel Boxall, for the story, and was struck by the extent to which they expected such a divide. “This sort of rapid response has been lacking from the Analogue Tories, who think AI is either a ‘mutant algorithm’ or a toy of Silicon Valley that can be expanded without concern for the impact on working people,” Innes said. “It’s taken them seven years to get the online safety act through parliament, and in the meantime, the world has moved on. “We need to get past the idea that you’re either in favour of innovation, or you’re in favour of protecting the public interest – that it’s government versus business,” Innes added. “The vast majority of tech firms want to see their tools used for good. The tech sector knows this is a problem – MrDeepFakes doesn’t represent them. So I think they’ll want to help us with this.” The policy paper also proposes a softer set of regulations for the wider tech sector

that supports AI. Web hosts, search engines and payment platforms would be obliged to ensure their clients aren't facilitating the creation of "harmful deepfakes", backed up by fines from Ofcom. Critics, in turn, might object that such a policy could have a chilling effect: if "harmful" is in the eye of the beholder, then it may be easier for a platform to ban all deepfake tools entirely. According to polling from Control AI, a non-profit that focuses on AI regulation, the UK public is more supportive of banning deepfakes than almost anywhere else – 86% of people support action. But even Italy, where support was weakest, had a comfortable majority in favour, at 74%. Deepfakes, 'cheapfakes' and AI elections – join us live One other proposal in the paper that seems less likely to come to pass is a suggestion that all major parties commit to not using AI to create misleading content for their campaigning over the next nine months. Call me pessimistic, but I don't think such a commitment would withstand the acrimony we're about to see ramping up across Britain – nor the value of plausibly deniable social media campaigning that smears your rivals. Coincidentally, I'll be hosting a Guardian Live event on that very topic next month. A panel of experts, including Katie Harbath of the technology policy firm Anchor Change and Imran Ahmed of the Center for Countering Digital Hate, will join me to talk about what the next year could look like, as two billion people vote in the first wave of elections that could be plausibly affected by generative AI. It feels a given that we'll see deepfakes and other AI-generated misinformation be used as a campaigning tool, but it's less clear whether that will work. Are fake images and videos a step change in the misinformation game, or are they just an evolution of text-based lies and "cheapfakes", a real image with a fake or misleading caption? I'm more worried about the effect of the new technology on the already weakened public realm. Twitter is a shell of its former self, Reddit is just about to go public on the back of AI deals, Threads explicitly suppresses political conversation and Google search is full of AI-generated SEO spam. Where is the conversation going to actually happen? And how does campaigning work in that brave new world? Robots I don't normally drop YouTube videos here, but Figure's latest demonstration is so extraordinary it's worth sharing this video. We're firmly out of prediction season, but if I had to make one about the next 12 months, it would be this: what 2022 was for chatbots, 2024 will be for robots. Robotics has traditionally been a hard, slow and expensive area to work in. But lessons drawn from the AI breakthroughs of the last few years are starting to change that. If you can train systems in simulated worlds, command them with natural language, and then successfully give them control of physical bodies, you can start seeing the same rate of improvement that we've had with large language models over the past five years. And, as I understand it, that's what has been happening. If you want to read the complete version of the newsletter please subscribe to receive TechScape in your inbox every Tuesday.