'Inceptionism' and Balenciaga popes: a brief history of deepfakes

Concern about doctored or manipulative media is always high around election cycles, but 2024 will be different for two reasons: deepfakes made by artificial intelligence (AI) and the sheer number of polls. The term deepfake refers to a hoax that uses AI to create a phoney image, most commonly fake videos of people, with the effect often compounded by a voice component. Combined with the fact that around half the world's population is holding important elections this year – including India, the US, the EU and, most probably, the UK – and there is potential for the technology to be highly disruptive. Here is a guide to some of the most effective deepfakes in recent years, including the first attempts to create hoax images. DeepDream's banana 2015 The banana where it all began. In 2015, Google published a blogpost on what it called "inceptionism", but which rapidly came to be known as "DeepDream". In it, engineers from the company's photo team asked a simple enough question: what happens if you take the AI systems that Google had developed to label images – known as neural networks – and ask them to create images instead? "Neural networks that were trained to discriminate between different kinds of images have quite a bit of the information needed to generate images too," wrote the team. The resulting hallucinatory dreamscapes were hardly high-fidelity, but they showed the promise of the approach. Celebrity face-swaps 2017 Generating images entirely from scratch is hard. But using AI to make changes to pre-existing photos and videos is slightly easier. In 2017, the technology was on the absolute cutting-edge, requiring a powerful computer, a tonne of imagery to learn from, and the time and wherewithal to master tools that weren't user-friendly. But one use-case fell squarely within those limits: face-swapping female celebrities into pornography. By late 2017, such explicit clips were being made and traded, at a remarkable rate, initially on reddit, and then, as word spread and anger rose, on shadier and more hidden forums. These days, according to a briefing note from thinktank Labour Together: "one in every three deepfake tools allow users to create deepfake pornography in under 25 minutes, and at no cost." Jordan Peele/Obama video 2018 Where pornography led, the rest of the media followed, and midway through 2018, face-swapping tools had improved to the extent that they could be used as a creative tool in their own right. One such video, created by BuzzFeed, saw actor and director Jordan Peele doing an impression of Barack Obama – swapped into actual footage of the president himself, ending with an exhortation to "stay woke, bitches". Dall-E's avocado armchair 2021 In 2021, OpenAI released Dall-E, and face-swapping became old news. The first major image generator, Dall-E offered the science-fiction promise of typing a phrase in, and getting a picture out. Sure, those pictures weren't particularly good, but they were images that had never existed before – not simply remixed versions of previous pics, but wholly new things. The first version of Dall-E wasn't great at photorealism, with OpenAI's demo selection

showing one vaguely realistic image set, a number of photos of an eerily pupil-less cat. But for more figurative art, such as this armchair, it already showed promise. Zelensky orders his side to surrender 2022/2023 Within a month of Russia's invasion of Ukraine an amateurish deepfake emerged of President Volodymyr Zelensky calling on his soldiers to lay down their weapons and return to their families. It was poor quality but prompted the real Ukrainian president to hit back on his Instagram account, telling Russian soldiers to return home instead. But then the deepfake Zelensky came back a year later and underlined how the technology had improved. This clip urged Ukrainian soldiers to surrender again and was more convincing. NewsGuard, an organisation that tracks misinformation, said comparing the two videos shows how far the technology has advanced in a short space of time. McKenzie Sadeghi, the AI and foreign influence editor at NewsGuard, said the November 2023 clip "marked a significant leap in deepfake technology since the outbreak of the Russia-Ukraine war in February 2022, representing how manipulated content has become more convincing as technology advances." Sadeghi said the movements of the 2023 deepfake are "fluid and natural" and his mouth movements match the words spoken more closely. The pope in a padded jacket 2023 An image of Pope Francis apparently clad in a Balenciaga quilted jacket was a landmark moment in generative AI and deepfakery. Created by the Midjourney image-making tool, it soon went viral because of its staggering level of realism. "The pope image showed us what generative AI is capable of and showed us how quickly this content can spread online," says Hany Farid, a professor at the University of California, Berkeley, and a specialist in deepfake detection. While the pope image was shared because of its combination of realism and the absurd, it underlined the creative power of now-widely accessible AI systems. Trump with Black voters 2024 This year a faked image emerged of Donald Trump posing with a group of Black men on a door step. There is also an image of Trump posing with a group of Black women. The former president, who will face Joe Biden in the 2024 presidential election, is a popular subject of deepfakes – as is his opponent. "The image of Trump with Black supporters shows us how the visual record can and is being polluted with fake imagery," says Farid. In a blog that compiles political deepfakes, Farid says the image appears to be "an attempt to court Black voters". Farid expresses a general concern that deepfakes will be "weaponised in politics" this year. Joe Biden robocalls 2024 There are numerous examples of the US president's image and voice being used in a manipulative manner. In January, Biden's faked voice was used to encourage Democrats not to vote in a New Hampshire primary, even deploying the Biden-esque phrase of "what a bunch of malarkey". Steve Kramer, a political operative, admitted that he was behind the faked calls. Kramer was working for Biden's challenger, Dean Phillips, whose supporters have experimented with AI technology by creating a short-lived Phillips voice bot. Phillips's campaign said the challenger had nothing to do with the call. Kramer has said he did it to flag the dangers of AI in elections.