

## **From Ship of Theseus to modern elections – deepfakes and their influence on political advertising**



**Imagine a ship made of wood. Imagine we replace one plank to fix a hole. Will it still be the same ship? What if we replace all the planks, is it a new ship then? It is easier to visualize this while dealing with countable planks than with humans. But what if we replace some of the personality traits of a politician. Will he or she still be the same person? These questions are becoming more important in an era of digital manipulation that can intensify the processes observed in politics for years – writes Mateusz Łabuz, expert for disinformation and digital media from the IFSH.**

In an era where artificial intelligence and computer-generated, visually enhanced and filtered content reshapes every aspect of life, politics is no exception. The problem is bigger than photo-filters concealing wrinkles or optimizing body-shapes on our social media posts in accordance to unrealistic beauty ideals. The new technological frontier are deepfakes, and not in the way we have associated with them so far. AI-generated or AI-manipulated visual and audio content that appears to be authentic but in fact is misleading or deceiving has been used as a tool of disinformation, subversion, and cognitive warfare. However, it is increasingly used in other ways, for example to enable politicians to create and optimize their digital avatars and boost their attractiveness. And at the same time replace some specific features, the individual planks of the metaphoric ship that is us. As the digital landscape becomes increasingly

indistinguishable from reality, where do we draw the line between persuasion and deception? Is it possible to set the limits of manipulation at all?

### **Deepfakes: A Game-Changer in Future Political Advertising?**

Political advertising has evolved for decades in parallel with the technologies that allow politicians to communicate with the people. Joseph Stalin's intelligence agencies forged photos to erase political rivals from history books, Franklin D. Roosevelt revolutionized campaigns with radio, John F. Kennedy capitalized on television, and Donald Trump leveraged social media. Although they used various tools that helped amplify their reach and slightly manipulated their image, they were still fundamentally themselves (in some cases – unfortunately). AI-generated media present the next major shift. Deepfakes allow politicians to alter their appearance, tone, message delivery, characteristic marks, or eliminate personal flaws. Creating hyper-realistic avatars that resemble the original, but are never tired, pretend to speak different languages, appear more charismatic, and do not have annoying features is becoming easier than ever.

AI allows for creating copies of reality that are intended to imitate it, thereby misleading recipients. For this reason, scientists have been pointing out for many years that deepfakes undermine the “seeing means believing” paradigm and the epistemic value of audio and visual materials. For most of our history, visual validation was the primary mechanism to proof the validity of something. Seeing with our own eyes, eyewitness accounts, tended to make things to appear more real for humans, than for example hear-say. It is not without reason that we say “pics or didn't happen”. AI-generated media threaten our primary sensemaking tool for an ever more complex reality.

The consensual use of deepfake technology in political campaigns is still relatively new phenomenon and understudied issue. In South Korea's 2022 presidential campaign, Yoon Suk-yeol deployed an AI-generated avatar, so-called “AI Yoon”, to engage with young voters. Yoon's staff produced his deepfakes based on previously recorded materials, which allowed AI Yoon to reach millions of recipients. Politician's digital alter ego was more approachable, as AI Yoon regularly joked and used slang typical for gaming community. This resonated with the targeted audience.

In India, deepfake technology was used by politicians to reach linguistic minorities. In 2024 elections campaign, their speeches were generated by AI in different dialects they normally do not speak and disseminated on social media. New York City Mayor Eric Adams used similar pattern, producing deepfake robocalls in Yiddish and Mandarin. Despite public outcry, he bluntly stated: “I've got to run the city, and I have to be able to speak to people in the languages that they understand, and I'm happy to do so”.

### **The Beginning of the Race**

These are just a few examples where deepfakes have been used to fuel political campaigns, and their applications may seem savvy and harmless on the surface. However, this is just the beginning and the real challenges await us in the future.

Imagine politicians who communicate with their audience primarily through recordings posted in social media. Of course, they already have advantages on their side without using AI at all, as they decide what and how to record. However, supported by AI, they can completely eliminate features that potential recipients might find irritating. A high pitch of voice? Making long pauses while speaking? Poor body language? Or maybe nodding like the real Yoon? AI will take care of this. What's more, it will allow for the generation of instant messages on a large scale, thus creating a false image that can be widely distributed and identified with a true one.

This raises a fundamental ethical question: if a deepfake version of a politician modifies key traits – body language, charisma, even spoken language – can it still be considered a faithful representation? This dilemma mirrors the ancient conundrum, which can still draw our attention to the sources and nature of identity.

### **The “Ship of Theseus” Paradox and Identity Manipulation**

Imagine the wooden ship Theseus, the Greek hero that slayed the mythological Minotaur, sailed on. Philosophers throughout the ages, starting from Plutarch at the beginning of the 2<sup>nd</sup> century, have asked themselves questions about the nature of identity, creatively developing this riddle. What if we replaced a few planks in this ship? What if we replaced all of them? Would it still be the same ship? Does the identity of the ship change when all its constitutional elements are replaced? Likewise, if every imperfection of a politician is erased or modified through AI, does the digital alter ego still reflect the original? Can it be identified with it? Should it be?

Originally, the conundrum was conceived as an ontological problem that boils down to considerations on identity. One may plausibly argue that it also has an epistemic dimension, related to our perception of the world. Let's suppose I have never seen the original Ship of Theseus, and do not know that some of its elements have been replaced. Or I have never seen a given politician, and I build my image only on interaction with his digital avatar. This is not unusual, considering that a large part of the campaign has moved to the Internet and classic election rallies are losing their importance. I may even know or suppose that something has been changed, but I don't know what or to what extent. Epistemically, I will not be prepared to assess the true features and true identity of any of these objects. What's more, I will develop false cognitive associations, which I will later decode when thinking about a given object. My approach to politics will be marked by a significant cognitive bias.

We already know such phenomena that take advantage of our cognitive weaknesses. After all, politicians have been trying to mislead us for centuries. They carefully choose the narrative, use rhetoric tailored to different target groups, and sometimes influence our unconsciousness through appropriate clothing. Suddenly, it turns out that the color of the tie can be of fundamental importance and influence the decisions we make. As humanity, we also sugarcoat reality. Make-ups, plastic surgeries – these are all useful tools to mask our imperfections and highlight our strengths. Photoshop and the digital revolution have allowed us to manipulate our image even better. Automatic filters on our smartphones make beautifying ourselves easier than ever. We are therefore creating a parallel reality that is supposed to be more beneficial to us.

We all know that politicians manipulate and deceive us. Because they do. Deepfakes political advertising may seem like nothing new, but until now such activities required significant investments of time and energy, and often the efforts of politicians themselves, to skillfully shape our perception of reality. AI takes this phenomenon to a whole new level.

### **Regulation and the Call for Transparency**

In a world where seeing is no longer believing, how do we maintain trust in democratic campaigners? Deepfake technology introduces both ontological and epistemic problems in political advertising. If voters engage with an AI-generated version of the politician rather than the real candidate, their perception of that individual is fundamentally altered. The reality they believe in – the political identity they trust – may be no more than a curated illusion, another digital universe that is an alternative to truth and reality. Let's not delude ourselves, we are already detached from reality, but by looking at a recorded politician we can still determine what pisses us off. The increasing reliance on deepfakes in political campaigns could erode voter confidence and trust even further, making it harder to discern genuine political messaging from AI-generated alternative reality.

Despite its potential, deepfake political advertising remains largely unregulated. The EU AI Act touches on AI-generated content but lacks specific provisions for political advertising. Currently, a key safeguard against deepfakes lies in transparency obligations. To mitigate manipulation, synthetic media will be equipped with mandatory disclaimers. However, the idea behind them refers primarily to the manipulation at the technological level and is based on the generic disclosure of the use of AI to create the content. We can sometimes see this in social media, where AI-generated content is accompanied by a very basic label that is intended to meet current compliance requirements. Do recipients know what "generated with AI" really means? Shouldn't they be automatically informed that they may be manipulated and the video they see may not reflect reality? This would help them make informed choices and address the manipulation on the ontological and epistemic level.

This is of particular importance in the area of political advertising, which is one of the elements that co-shapes the essence of democracy – free and fair elections – and is therefore subject to the enhanced legal regime. Perhaps strengthened disclaimers could become a part of the political advertising ecosystem, precisely because of its importance and impact on democracy. Maybe it's worth considering similar disclaimers for other forms of visual alterations? After all, Photoshopping a photo comes down to the same thing – manipulating the recipient as to the authenticity. In future this might be a focus of the relatively new EU *Regulation on the transparency and targeting of political advertising* adopted in 2024, and the legislation introduced at the national level.

### **The Future of AI in Politics: Where Do We Go from Here?**

Deepfake technology, while controversial, has also positive applications. It could help political candidates connect with marginalized groups, improve accessibility in multilingual societies, and make political engagement more inclusive. Adams' robocalls or the strategy used in India could be seen as a chance to overcome linguistic exclusion, and in some respects can even be labeled as pro-democratic. Moreover, if AI could break down language barriers through real-

time translation, it could enable global direct democracy and create an entirely new communication ecosystem in the EU. However, if there is no disclaimer that AI was involved in producing content, this form of communication should be assessed negatively and disqualified as legitimate. When such disclaimers are put in place, the matter is more complicated, because they do not exclude manipulation that reaches deeper levels of our consciousness.

I do believe that regulations for deepfakes political advertising should go beyond disclosure and place restrictions on the extent to which AI can be consensually used to modify political figures. Legislators must address these issues before deepfake technology becomes a standard tool in political campaigns. Without ethical guidelines, the risks might outweigh the benefits, and the future will inevitably bring further technological progress. The 2024 campaign in India showed that politicians are already willing to go further – some AI-generated robocalls addressed citizens by their first names. It's not hard for me to imagine that as technology advances, politicians' avatars will engage in direct conversations with voters. Without clear rules, we will find ourselves, at best, in a deeply faked democracy.

One of the major concerns of the growing analytical and creative capabilities of generative AI is the ability to automatically produce content tailored to individual audiences (personalization). Deepfake avatars fed by data analysis, able to select a narrative and direct it to specific subgroups or even individuals, would open a new chapter of political manipulation and deny the idea of choices based on facts and logic even further, partly regardless of the markings due to the above-mentioned weaknesses of human psychology.

Policymakers must balance innovation with accountability, ensuring that voters are informed participants in democratic processes, not unconscious subjects of AI-generated illusions. Therefore, apart from regulation, media literacy programs must be adapted to the age of AI-generated content. Citizens should be equipped with tools and skills to critically assess political messaging and understand at least basics of cognitive manipulations. This is, after all, an element of building social resilience, also in the area of disinformation coming from adversaries. Educational campaigns and digital literacy initiatives can be crucial in safeguarding democratic integrity on many fields.

### **Conclusion: A Fork in the Road**

A few years ago, Mark Zuckerberg promised us that soon we would all move to the metaverse and spend long hours in virtual reality. Perhaps today we are much closer to creating our own small multiverses that imitate reality, but change its essential elements. Chatbots impersonating popular figures that interact with users, synthetic celebrities on social media, and soon synthetic politicians.

Every day as a society we create our own reflections of digital identities, our own "Ships of Theseus" in which we eagerly replace the planks. This trend is inevitable, and technology allows us to do it on an ever-increasing scale. Ancient dilemma is suddenly more alive than ever, and it seems to have far more real consequences.

As deepfakes start to redefine political advertising, society faces an urgent question: how do we safeguard democratic integrity in an age where everything can be artificially generated? I

don't want to prophesize any political dystopia, but I am deeply concerned that in the era of disinformation and post-truth narratives, where society has already trouble agreeing on basic facts, new AI-driven incarnations of AI Yoon and Co will manipulate our cognitive processes even more effectively. And if they can be also driven by microtargeting and the analytical power of AI, the political reality will sail away from us forever.

**Mateusz Łabuz** – researcher at the Institute for Peace Research and Security Policy at the University of Hamburg (IFSH) and PhD researcher at the Chemnitz University of Technology. Non-resident fellow at the Future Shift Labs and CEE Digital Democracy Watch. He was a career diplomat at the Polish Ministry of Foreign Affairs. He lectures on cybersecurity, artificial intelligence, and disinformation at the University of the National Education Commission and the Pontifical University of John Paul II in Cracow. His interest comprises the synthetic media, deepfakes, as well as emerging trends in the disinformation sphere, including social resilience and cognitive warfare.