

Case study for the 2025 EEB – Semi-final B

TRUTH IN THE AGE OF AI – ETHICS AT THE EDGE OF REALITY

In the digital age, artificial intelligence (AI) has revolutionised content creation and information dissemination. Technologies such as deepfakes, AI-generated synthetic media that convincingly mimic real people, have blurred the line between truth and fiction. Simultaneously, algorithmic recommendation systems on platforms like YouTube and TikTok filter and amplify content, shaping public perception and potentially distorting reality.

In early 2026, a video featuring a prominent European political leader making inflammatory remarks about immigration policy went viral. The video was later revealed to be a deepfake. Despite its falsehood, the video was algorithmically promoted by platforms like YouTube and TikTok, reaching millions of viewers within hours. The platforms' recommendation systems, designed to maximise engagement, identified the video as "high-interest" content and pushed it to users already interacting with political material.

The video remained online for 48 hours before being flagged and removed. During that time, it sparked widespread public outrage, protests in several cities, and even influenced political discourse. Content moderators had flagged the video internally, but the platform's automated systems overrode manual review due to its high engagement metrics.

This incident raises a complex ethical dilemma:

- Who is responsible when synthetic media is algorithmically amplified before its authenticity can be verified?
- Is it the creators of the deepfake, who intentionally manipulated public perception?
- The platforms, whose algorithms prioritised virality over truth?
- Or the users, who shared and reacted to the content without verifying its source?

This dilemma sits at the intersection of AI ethics, data security, and information integrity, challenging the boundaries of responsibility in the digital age.