STATE OF CALIFORNIA
FAIR POLITICAL PRACTICES COMMISSION
1102 Q Street • Suite 3050 • Sacramento, CA 95811
(916) 322-5660

To: Chair Silver and Commissioners Brandt, Ortiz, and Wilson

From: Eva Hartman, Executive Fellow

Subject: **Detecting AI in Campaign Advertisements Panel Report**

Date: June 19, 2025

I. Executive Summary

The Fair Political Practices Commission held the first of three panels in an informational series titled "Effectively Leveraging and Regulating AI in the Political Process" on April 22, 2025. This panel, "Detecting AI in Campaign Advertisements," featured presentations from academic and private sector experts who offered various technological options for enforcing AB 2355: Artificial Intelligence Disclosure. The discussion concluded that there is no perfect method of Artificial Intelligence ("AI") detection, and a combination of technological solutions, human investigation and legislation would be needed for the FPPC to fully enforce this legislation.

II. Background: AB 2355 Artificial Intelligence Disclosure (2024)

AB 2355 by Assemblymembers Carillo and Cervantes created an additional disclosure requirement for political advertisers: they must label their advertisements with a disclaimer stating that the photos, videos or audio included were generated or substantially altered by artificial intelligence. The bill specifies that determinations on whether an advertisement has been "substantially altered" by artificial intelligence rest on whether the media would incorrectly appear authentic to a reasonable person, or whether a person would have a fundamentally altered understanding of the media had it not been altered.

The Fair Political Practices Commission has been tasked with enforcing these disclosures and currently faces the challenge of accurately evaluating political advertisements for signs of AI use. Without a tool to reliably identify artificial intelligence alterations in audio, videos and photos, the FPPC may be unable to effectively hold violators accountable. This technological challenge is shared by governmental ethics agencies across the country who are also confronting the rise of artificial intelligence use in their elections. This panel sought to open a conversation on this issue and identify potential options for detecting AI generated or altered content.

III. Presentations

Dr. James F. O'Brien, Professor of Computer Science, University of California, Berkeley

Professor O'Brien is an expert on detection and analysis of fake images and video. He has frequently worked with news organizations on exposing fake or altered photographs, as well as images created by generative artificial intelligence software. His methods have been used to expose fabrication of medical research records, validate evidentiary videos, and rule out conspiracy theories relating to photographs of the Moon landing.

Summary:

- *Detection Software:* Dr. O'Brien advocated for a view of generative AI and detection software as being in an arms race for technological superiority, which will inevitably be won by generative AI. He stated that it will soon be impossible for humans and machines to distinguish between an AI generated image and an authentic image.
- *Human Expertise:* Dr. O'Brien stated that human experts on AI image detection, including those who testify in courts about the reliability of images, are more trustworthy than detection software. However, the human detection process is time intensive and expensive, making it an ineffective method considering the high volume of AI generated images.
- *Watermarks:* Dr. O'Brien expressed his opinion that watermarking AI generated or altered images is ineffective because the watermark can easily be cropped out or removed using further AI alterations.
- *Provenance:* Dr. O'Brien believes that provenance or Content Credentials may be the most useful form of image labelling and alteration tracking. However, in order for it to be reliable, the use of content credentials must be mandated, and the image must be handled through a secure chain of custody, which Dr. O'Brien likened to that of criminal evidence.

Kevin Guo, CEO, Hive

Hive is an artificial intelligence company that offers visual, text and audio moderation, AI-generated content classification, text, image and video generation, searches, likeness detection, and more. Kevin Guo co-founded Hive in 2015 and currently serves as CEO.

Summary:

- *Efficacy:* Mr. Guo stated that he believes that his product is, and will continue to be, a highly effective means of identifying AI generated and altered images.
- *Technology:* Hive's AI detection software focuses on the origination processes of images instead of the pixelation, which is the target of many other detection companies. Because of this, Mr. Guo believes that Hive's technology will continue to be effective against increasingly sophisticated AI content generators.
- *Accuracy:* Images are labelled with confidence scores to represent the model's confidence that the image was made using artificial intelligence, which could be incorporated into the investigative process. The model also predicts which AI technology was used to create the image.

Katie Brooke, Senior Manager, Public Policy & Government Relations, Adobe, and Andrew Kaback, Lead Product Manager (CAI), Adobe

In 2019, Adobe founded the Content Authenticity Initiative, a technological community committed to developing open-source tools for verifiably recording the provenance of any digital media, including content made with generative AI. They created Content Credentials, which contain information about who produced a piece of content, when they produced it, and which tools and editing processes they used.

Summary:

- *Efficacy:* Ms. Brooke and Mr. Kaback stated that when credentials are mandated, they are a highly effective way to track the origin and editing of images, particularly images in political ads which will be run through multiple different types of editing software.
- *Accuracy:* Content Credentials cannot be faked or mimicked. Credential detection software will not recognize them.
- *Drawbacks:* Content Credentials will not include information from editing or artificial intelligence platforms when the platform has not chosen to adopt them. They are not an AI detection software, though they can sometimes serve this purpose, rather, they track the chain of custody and editing the photo undergoes.
- *Investigative Use:* Ms. Brooke and Mr. Kaback stated that Content Credentials may be most effective if their use is mandated by the Legislature. However, because many political advertisements are already being edited using Adobe tools, they likely have Content Credentials embedded without the creator explicitly intending to do so. Therefore, they could aid current investigations.

IV. Conclusion: Themes, Insights, and Implications for the FPPC

- The use of artificial intelligence to generate and alter images, video, and audio will continue to evolve and become more difficult to detect. Whether detection methods and tools will advance quickly enough to enable the FPPC to accurately identify AI-generated or modified content remains uncertain.
- The FPPC is unlikely to find a method or tool that guarantees perfect detection accuracy. Regardless of what future action is taken, the FPPC will need to incorporate human investigation into case inquiries.
- Further issues with the law, including defining what alterations count as "significant," and what qualifies as a "significantly altered understanding" of content, can be resolved through regulation.
- A comprehensive legislative solution defining AI generated content should be labeled or identified may help resolve some identification issues for the public seeing the advertisement but would leave open the question of enforcement of such a law. The Legislature's openness to, and the legality of, adopting mandatory Content Credentials should be studied.