CLAIMS

What is claimed is:

1. A computer-implemented method comprising:

obtaining, by a computer, an audiovisual data sample containing audiovisual data;

applying, by the computer, a machine-learning architecture to the audiovisual data to generate similarity scores using a voiceprint biometric embedding and faceprint biometric embedding extracted from the audiovisual data and a voiceprint enrollment embedding and a faceprint enrollment embedding, generate a lip-sync score using speech audio and one or more lip-sync embeddings extracted from the audiovisual data, and generate deepfake scores using a speaker spoofprint and a facial spoofprint extracted from the audiovisual data; and

generating, by the computer, a final output score indicating a likelihood that the audiovisual data is genuine using the similarity scores, the lip-sync score and the deepfake scores.

2. The method according to claim 1, further comprising identifying, by the computer, the audiovisual data sample as a genuine data sample in response to determining that the final output scores satisfies a threshold.

3. The method according to claim 1 or 2, further comprising identifying, by the computer, deepfake content in the audiovisual data in response to determining that at least one of the deepfake score satisfies a deepfake detection threshold.

4. The method according to any one of claims 1 to 3, further comprising:

extracting, by a computer, the voiceprint for an audiovisual sample by applying a speaker embedding extraction engine of the machine-learning architecture to an audio signal of the audiovisual data; and

extracting, by the computer, the speaker spoofprint for the audiovisual data by applying an audio spoofprint embedding extraction engine of the machine-learning architecture to the audio signal of the audiovisual data.

5. The method according to any one of claims 1 to 4, further comprising:

extracting, by a computer, the faceprint for the audiovisual data by applying a faceprint embedding extraction engine of the machine-learning architecture to visual media of the audiovisual data; and

extracting, by a computer, the facial spoofprint for the audiovisual data by applying a visual spoofprint embedding extraction engine of the machine-learning architecture to the visual media of the audiovisual data.

6. The method according to any one of claims 1 to 5, further comprising extracting, by the computer a feature for the speaker voiceprint embedding, the feature extracted from audio data of the audiovisual data.

7. The method according to any one of claims 1 to 6, further comprising extracting, by the computer, a feature for the faceprint embedding, the feature extracted from image data of the audiovisual data.

8. The method according to any one of claims 1 to 7, further comprising parsing, by the computer, the audiovisual sample into a plurality of segments having a preconfigured length, wherein the computer generates the biometric embedding and a spoofprint for each segment.

9. The method according to any one of claims 1 to 8, further comprising generating, by the computer, the lip-sync score by applying a lip sync estimation engine of the machine-learning architecture on the audiovisual data, wherein the computer generates final output score using the lip-sync score.

10. A system comprising:

a computer comprising a processor configured to:

obtain an audiovisual data sample containing audiovisual data;

apply a machine-learning architecture to the audiovisual data to generate similarity scores using a voiceprint biometric embedding and faceprint biometric embedding extracted from the audiovisual data and a voiceprint enrollment embedding and a faceprint enrollment embedding, generate a lip-sync score using speech audio and one or more lip-

sync embeddings extracted from the audiovisual data, and generate deepfake scores using a speaker spoofprint and a facial spoofprint extracted from the audiovisual data; and

generate a final output score indicating a likelihood that the audiovisual data is genuine using the similarity scores, the lip-sync score and the deepfake scores.

11. The system according to claim 10, implementing the method of any one of claims 2 to 9.