CLAIMS

What is claimed is:

1. A computer-implemented method comprising:

extracting, by a computer, an inbound embedding for an inbound speaker by applying a machine-learning model on an inbound audio signal;

generating, by the computer, a similarity score based upon a distance between the inbound embedding and a voiceprint stored in speaker profile in a speaker profile database;

responsive to the computer determining that the similarity score for the inbound embedding fails to satisfy a similarity threshold:

generating, by the computer, in the speaker profile database a new speaker profile for the inbound speaker containing the inbound embedding, the new speaker profile is database record storing the inbound embedding as a new voiceprint for the inbound speaker, wherein the new voiceprint satisfies a maturity threshold corresponding to one or more maturity factors of the new voiceprint, the one or more maturity factors include at least one of: a number of utterances associated with the new voiceprint, an overall duration of net speech across utterances associated with the new voiceprint, or an audio quality of utterances associated with the new voiceprint.

2. The method according to claim 1, further comprising:

receiving, by the computer, the inbound audio signal from an end-user device via an intermediate server; and

transmitting, by the computer, a new speaker identifier associated with the new speaker profile to the intermediate server.

3. The method according to claim 2, wherein the end-user device is at least one of a smart television, a media device coupled to a television, and an edge device.

4. The method according to claim 1, further comprising extracting, by the computer, one or more features from the inbound audio signal, wherein the computer generates the inbound embedding by applying the machine-learning model on the one or more features the computer extracted from the inbound audio signal.

5. The method according to claim 1, further comprising:

extracting, by the computer, a second inbound embedding from a second inbound audio signal by applying the machine-learning model on the second inbound audio signal;

generating, by the computer, a second similarity score based upon the distance between the second inbound embedding and the new voiceprint stored in the new speaker profile; and

responsive to the computer determining that the second similarity score for the second inbound embedding satisfies the similarity threshold:

updating, by the computer, the new voiceprint for the inbound speaker based upon the second inbound audio signal.

6. The method according to claim 1, further comprising:

receiving, by the computer, a subscriber identifier associated with the inbound audio signal; and

identifying, by the computer, one or more speaker profiles associated with the subscriber identifier stored in the speaker profile database, wherein the computer generates one or more similarity scores for the inbound embedding based upon one or more voiceprints stored in the one or more speaker profiles associated with the subscriber identifier.

7. The method according to claim 6, wherein the subscriber identifier is associated with one or more speaker identifiers, and wherein each speaker profile is associated with a corresponding speaker identifier.

8. The method according to claim 7, wherein at least one of the subscriber identifier and each speaker identifier is an anonymized identifier.

9. The method according to claim 1, wherein the computer generates the new voiceprint based upon one or more inbound embeddings, the method further comprising:

identifying, by the computer, the one or more maturity factors for the new voiceprint based upon the one or more inbound embeddings; and

determining, by the computer, a level of maturity for the new voiceprint based upon the one or more maturity factors.

10. The method according to claim 9, further comprising updating, by the computer, a new similarity threshold of the new speaker profile in response to the computer determining that the level of maturity for the new voiceprint satisfies the maturity threshold.

11. The method according to claim 9, further comprising:

responsive to the computer determining that the level of maturity fails the maturity threshold:

generating, by the computer, an active enrollment prompt comprising a user interface configured to display a request for an additional inbound audio signal;

extracting, by the computer, an additional embedding from the additional inbound audio signal; and

updating, by the computer, the new voiceprint according to the additional embedding extracted from the additional inbound audio signal.

12. The method according to claim 9, further comprising updating, by the computer, the new speaker profile from a temporary profile to a permanent profile in response to the computer determining that the level of maturity satisfies the maturity threshold.

13. A system comprising:

a speaker profile database comprising non-transitory machine-readable storage media configured to store data records containing speaker profiles; and

a computer comprising a processor configured to:

extract an inbound embedding for an inbound speaker by applying a machine-learning model on an inbound audio signal;

generate a similarity score based upon a distance between the inbound embedding and a voiceprint stored in speaker profile in the speaker profile database; and

responsive to the computer determining that the similarity score for the inbound embedding fails to satisfy a similarity threshold:

generate in the speaker profile database a new speaker profile for the inbound speaker containing the inbound embedding, the new speaker profile is database record storing the inbound embedding as a new voiceprint for the inbound speaker, wherein the new voiceprint satisfies a maturity threshold corresponding to one or more maturity factors of the new voiceprint, the one or more maturity factors include at least one of: a number of utterances associated with the new voiceprint, an overall duration of net speech across utterances associated with the new voiceprint, or an audio quality of utterances associated with the new voiceprint.

14. The system according to claim 13, wherein the computer is further configured to:

receive the inbound audio signal from an end-user device via an intermediate server; and

transmit a new speaker identifier associated with the new speaker profile to the intermediate server.

15. The system according to claim 14, wherein the end-user device is at least one of a smart television, a media device coupled to a television, and an edge device.

16. The system according to claim 13, wherein the computer is further configured to extract one or more features from the inbound audio signal, wherein the computer generates the inbound embedding by applying the machine-learning model on the one or more features the computer extracted from the inbound audio signal.

17. The system according to claim 13, wherein the computer is further configured to:

extract a second inbound embedding from a second inbound audio signal by applying the machine-learning model on the second inbound audio signal;

generate a second similarity score based upon the distance between the second inbound embedding and the new voiceprint stored in the new speaker profile; and

responsive to the computer determining that the second similarity score for the second inbound embedding satisfies the similarity threshold:

update the new voiceprint for the inbound speaker based upon the second inbound audio signal.

18. The system according to claim 13, wherein the computer is further configured to:

receive a subscriber identifier associated with the inbound audio signal; and

identify one or more speaker profiles associated with the subscriber identifier stored in the speaker profile database, wherein the computer generates one or more similarity scores for the inbound embedding based upon one or more voiceprints stored in the one or more speaker profiles associated with the subscriber identifier.

19. The system according to claim 18, wherein the subscriber identifier is associated with one or more speaker identifiers, and wherein each speaker profile is associated with a corresponding speaker identifier.

20. The system according to claim 19, wherein at least one of the subscriber identifier and each speaker identifier is an anonymized identifier.

21. The system according to claim 13, wherein the computer generates the new voiceprint based upon one or more inbound embeddings, and wherein the computer is further configured to:

identify the one or more maturity factors for the new voiceprint based upon the one or more inbound embeddings; and

determine a level of maturity for the new voiceprint based upon the one or more maturity factors.

22. The system according to claim 21, wherein the computer is further configured to update a new similarity threshold of the new speaker profile in response to the computer determining that the level of maturity for the new voiceprint satisfies the maturity threshold.

23. The system according to claim 21, wherein the computer is further configured to:

responsive to determining that the level of maturity fails the maturity threshold:

generate an active enrollment prompt comprising a user interface configured to display a request for an additional inbound audio signal;

extract an additional embedding from the additional inbound audio signal; and

update the new voiceprint according to the additional embedding extracted from the additional inbound audio signal.

24. The system according to claim 21, wherein the computer is further configured to update the new speaker profile from a temporary profile to a permanent profile in response to the computer determining that the level of maturity satisfies the maturity threshold.