CLAIMS

What is claimed is:

1. A computer-implemented method comprising:

extracting, by the 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; and

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.

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 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 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 a similarity threshold:

updating, by the computer, the new voiceprint for the inbound speaker based upon the second inbound 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, 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 a maturity threshold.

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

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

generating, by the computer, an active enrollment prompts the 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 signal; and

updating, by the computer, the new voiceprint according to an additional embedding extracted from the additional inbound 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 a maturity threshold.

13. A system comprising:

a speaker 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.

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 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 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 a similarity threshold:

update the new voiceprint for the inbound speaker based upon the second inbound 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 one or more maturity factors for the new voice print 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 a 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 a maturity threshold:

generate an active enrollment prompts the 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 signal; and

update the new voiceprint according to an additional embedding extracted from the additional inbound 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 a maturity threshold.

~~**25.** A computer-implemented method comprising:~~

~~receiving, by a computer, an inbound audio signal including a plurality of utterances of a plurality of inbound speakers;~~

~~applying, by the computer, a machine-learning architecture to the inbound audio signal to extract a plurality of inbound embeddings corresponding to the plurality of inbound speakers;~~

~~for each inbound speaker of the plurality of inbound speakers:~~

~~generating, by the computer, one or more similarity scores based upon an inbound embedding for the inbound speaker, each similarity score for the inbound speaker indicating a distance between the inbound embedding and one or more voiceprints stored in a speaker profile database; and~~

~~identifying, by the computer, a nearest voiceprint for the inbound speaker from the one or more voiceprints, the nearest voiceprint corresponding to a max similarity score of the one or more similarity scores generated for the inbound speaker; and~~

~~for each max similarity score satisfying one or more similarity score thresholds:~~

updating, by the computer, a speaker profile database to include the inbound embedding for the inbound speaker having the max similarity score satisfying the one or more similarity score thresholds.

26. The method according to claim 25, further comprising identifying, by the computer, the one or more voiceprints stored in the speaker profile database based upon a subscriber identifier received with the inbound audio signal.

27. The method according to claim 25, further comprising:

determining, by the computer, that the max similarity score for the inbound speaker embedding satisfies the one or more similarity scores; and

identifying, by the computer, a speaker profile in the speaker database containing the nearest voiceprint for the inbound speaker, the speaker profile including a speaker identifier.

28. The method according to claim 25, further comprising:

determining, by the computer, that the max similarity score for the inbound speaker satisfies a first similarity threshold and fails a second similarity threshold, the first similarity threshold comparatively lower than a second similarity threshold; and

updating, by the computer, a list of weak embeddings stored in the speaker database to include the inbound embedding of the inbound speaker.

29. The method according to claim 28, further comprising:

executing, by the computer, a re-clustering operation on one or more speaker profiles in the speaker database associated with the subscriber identifier;

updating, by the computer, the max similarity score for the inbound embedding of the list of weak embeddings based upon the re-clustering operation; and

responsive to determining that the max similarity score for the inbound embedding of the list of weak embeddings satisfies one or more similarity score thresholds:

updating, by the computer, the speaker profile database to include the inbound embedding for the inbound speaker having the max similarity score satisfying the one or more similarity score thresholds.

30. The method according to claim 25, further comprising:

detecting, by the computer, a triggering condition for performing a re-clustering operation for the subscriber identifier; and

executing, by the computer, a hierarchical clustering operation on the plurality voiceprints of one or more speaker profiles associated with the subscriber identifier.

31. The method according to claim 30, further comprising updating, by the computer, the speaker identifier associated with a new voiceprint cluster from an existing speaker profile, the new voiceprint cluster generated by applying the hierarchical clustering operation on the one or more speaker profiles associated with the subscriber identifier.

32. The method according to claim 25, further comprising:

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

transmitting, by the computer, the speaker identifier associated with each inbound speaker having the max similarity score satisfying the one or more similarity score thresholds.

33. The method according to claim 25, further comprising receiving, by a computer, the inbound audio signal including a plurality of utterances of a plurality of inbound speakers

34. The method according to claim 25, further comprising:

applying, by the computer, the machine-learning architecture on the one or more features to identify one or more audio events to detect an environmental setting; and

transmitting, by the computer, to an intermediary server an indicator of the environmental setting and each speaker identifier associated with each inbound speaker having the max similarity score satisfying the one or more similarity score thresholds.

35. The method according to claim 25, further comprising:

for each max similarity score failing the one or more similarity score thresholds:

generating, by the computer, a new speaker profile associated with the subscriber identifier comprising the inbound embedding; and

updating, by the computer, the speaker profile database to include the new speaker profile and the inbound embedding for the inbound speaker.

36. A system comprising:

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

a computer comprising a processor configured to:

receive an inbound audio signal including a plurality of utterances of a plurality of inbound speakers;

apply a machine learning architecture to the inbound audio signal to extract a plurality of inbound embeddings corresponding to the plurality of inbound speakers;

for each inbound speaker of the plurality of inbound speakers:

generate one or more similarity scores based upon an inbound embedding for the inbound speaker, each similarity score for the inbound speaker indicating a distance between the inbound embedding and one or more voiceprints stored in a speaker profile database; and

identify a nearest voiceprint for the inbound speaker from the one or more voiceprints, the nearest voiceprint corresponding to a max similarity score of the one or more similarity scores generated for the inbound speaker; and

for each max similarity score satisfying one or more similarity score thresholds:

update a speaker profile database to include the inbound embedding for the inbound speaker having the max similarity score satisfying the one or more similarity score thresholds.

37. The system according to claim 36, wherein the computer is further configured to identify the one or more voiceprints stored in the speaker profile database based upon a subscriber identifier received with the inbound audio signal.

38. The system according to claim 36, wherein the computer is further configured to:

determine that the max similarity score for the inbound speaker embedding satisfies the one or more similarity scores; and

identify a speaker profile in the speaker database containing the nearest voiceprint for the inbound speaker, the speaker profile including a speaker identifier.

39. The system according to claim 36, wherein the computer is further configured to:

determine that the max similarity score for the inbound speaker satisfies a first similarity threshold and fails a second similarity threshold, the first similarity threshold comparatively lower than a second similarity threshold; and

update a list of weak embeddings stored in the speaker database to include the inbound embedding of the inbound speaker.

40. The system according to claim 39, wherein the computer is further configured to:

execute a re-clustering operation on one or more speaker profiles in the speaker database associated with the subscriber identifier;

update the max similarity score for the inbound embedding of the list of weak embeddings based upon the re-clustering operation; and

responsive to determining that the max similarity score for the inbound embedding of the list of weak embeddings satisfies one or more similarity score thresholds:

update the speaker profile database to include the inbound embedding for the inbound speaker having the max similarity score satisfying the one or more similarity score thresholds.

41. The system according to claim 36, wherein the computer is further configured to:

detect a triggering condition for performing a re-clustering operation for the subscriber identifier; and

execute a hierarchical clustering operation on the plurality voiceprints of one or more speaker profiles associated with the subscriber identifier.

42. The system according to claim 41, wherein the computer is further configured to update the speaker identifier associated with a new voiceprint cluster from an existing speaker profile, the new voiceprint cluster generated by applying the hierarchical clustering operation on the one or more speaker profiles associated with the subscriber identifier.

43. The system according to claim 36, wherein the computer is further configured to:

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

transmit the speaker identifier associated with each inbound speaker having the max similarity score satisfying the one or more similarity score thresholds.

44. The system according to claim 36, wherein the computer is further configured to receive the inbound audio signal including a plurality of utterances of a plurality of inbound speakers.

45. The system according to claim 36, wherein the computer is further configured to:

apply machine-learning architecture on the one or more features to identify one or more audio events to detect an environmental setting; and

transmit to an intermediary server an indicator of the environmental setting and each speaker identifier associated with each inbound speaker having the max similarity score satisfying the one or more similarity score thresholds.

46. The system according to claim 36, wherein the computer is further configured to:

for each max similarity score failing the one or more similarity score thresholds:

generate a new speaker profile associated with the subscriber identifier comprising the inbound embedding; and

update the speaker profile database to include the new speaker profile and the inbound embedding for the inbound speaker.

47. A computer-implemented method comprising:

receiving, by a computer, an inbound audio signal for an inbound speaker from an end-user device via a content server;

applying, by the computer, a machine learning model to the inbound audio signal to extract an inbound embedding for the inbound speaker;

generating, by the computer, a similarity score for the inbound embedding based upon a distance between the inbound embedding and a voiceprint stored in a speaker profile in a speaker database, the similarity score satisfying one or more similarity score thresholds;

identifying, by the computer, in the speaker profile one or more speaker characteristics corresponding to one or more content characteristics for the content server; and

transmitting, by the computer, the one or more speaker characteristics associated with the inbound speaker to the media content server.

48. The method according to claim 47, further comprising identifying, by the computer, the speaker profile based upon a subscriber identifier received from the content server.

49. The method according to claim 48, wherein the subscriber identifier received by the computer is a first anonymized identifier, wherein the subscriber identifier is associated with one or more speaker identifiers for one or more speakers, and wherein each speaker identifier is a second anonymized identifier corresponding to each speaker profile.

50. The method according to claim 47, wherein the one or more speaker characteristics includes at least one of an age characteristic and a gender characteristic.

51. The method according to claim 47, further comprising determining, by the computer, at least one of an age characteristic and a gender characteristic of the inbound speaker by applying a second machine-learning model on the inbound audio signal.

52. The method according to claim 47, further comprising determining, by the computer, an age of the inbound speaker based upon an age characteristic stored in the speaker profile for the inbound speaker.

53. The method according to claim 47, further comprising:

receiving, by the computer, speaker information indicating at least one speaker characteristic for the inbound speaker from the content server;

storing, by the computer, the at least one speaker characteristic into the speaker profile for the inbound speaker; and

identifying, by the computer, the speaker profile based, in part, upon the at least one speaker characteristic in the speaker profile.

54. The method according to claim 47, further comprising:

receiving, by the computer, inbound authentication data from content server with the inbound audio signal; and

authenticating, by the computer, the inbound speaker based upon the similarity score satisfying the similarity threshold and the inbound authentication data satisfying expected authentication data stored in the speaker profile.

55. The method according to claim 54, wherein the authentication data comprises at least one of end-user device information, metadata associated with the end-user device, speaker information, and biometric information.

56. The method according to claim 47, further comprising:

extracting, by the computer, one or more features from the inbound audio signal; and

calculating, by the computer, a spoofing score indicating a likelihood that the inbound audio signal includes a spoofing condition based upon the one or more features by applying a second machine learning model.

57. The method according to claim 47, further comprising identifying, by the computer, in a media content database one or more media content files having one or more content characteristics corresponding to the one or more speaker characteristics.

58. The method according to claim 57, further comprising enabling, by the computer, age-restricted content in the content database based upon an age characteristic of the one or more speaker characteristics of the speaker satisfying a corresponding age restriction characteristic in the one or more content characteristics of the age-restricted content.

59. A system comprising:

a speaker database comprising non-transitory machine readable storage media configured to store a plurality of speaker profiles; and

a server comprising a processor configured to:

receive an inbound audio signal for an inbound speaker from an end-user device via a content server;

apply a machine learning model to the inbound audio signal to extract an inbound embedding for the inbound speaker;

generate a similarity score for the inbound embedding based upon a distance between the inbound embedding and a voiceprint stored in a speaker profile in the speaker database, the similarity score satisfying one or more similarity score thresholds;

identify in the speaker profile one or more speaker characteristics corresponding to one or more content characteristics for the content server; and

transmit the one or more speaker characteristics associated with the inbound speaker to the media content server.

60. The system according to claim 59, wherein the computer is further configured to identify the speaker profile based upon a subscriber identifier received from the content server.

61. The system according to claim 60, wherein the subscriber identifier received by the computer is a first anonymized identifier, wherein the subscriber identifier is associated with one or more speaker identifiers for one or more speakers, and wherein each speaker identifier is a second anonymized identifier corresponding to each speaker profile.

62. The system according to claim 59, wherein the one or more speaker characteristics includes at least one of an age characteristic and a gender characteristic.

63. The system according to claim 59, wherein the computer is further configured to determine at least one of an age characteristic and a gender characteristic of the inbound speaker by applying a second machine-learning model on the inbound audio signal.

64. The system according to claim 59, wherein the computer is further configured to determine an age of the inbound speaker based upon an age characteristic stored in the speaker profile for the inbound speaker.

65. The system according to claim 59, wherein the computer is further configured to:

receive speaker information indicating at least one speaker characteristic for the inbound speaker from the content server;

store the at least one speaker characteristic into the speaker profile for the inbound speaker; and

identify the speaker profile based, in part, upon the at least one speaker characteristic in the speaker profile.

66. The system according to claim 59, wherein the computer is further configured to:

receive inbound authentication data from content server with the inbound audio signal; and

authenticate the inbound speaker based upon the similarity score satisfying the similarity threshold and the inbound authentication data satisfying expected authentication data stored in the speaker profile.

67. The system according to claim 66, wherein the authentication data comprises at least one of end-user device information, metadata associated with the end-user device, speaker information, and biometric information.

68. The system according to claim 59, wherein the computer is further configured to:

extract one or more features from the inbound audio signal; and

calculate a spoofing score indicating a likelihood that the inbound audio signal includes a spoofing condition based upon the one or more acoustic features by applying a second machine-learning model.

69. The system according to claim 59, wherein the computer is further configured to identify in a media content database one or more media content files having one or more content characteristics corresponding to the one or more speaker characteristics.

70. The system according to claim 69, wherein the computer is further configured to enable, age restricted content in the content database based upon an age characteristic of the one or more speaker characteristics of the speaker satisfying a corresponding age restriction characteristic in the one or more content characteristics of the age-restricted content.

71. A method comprising:

obtaining, by the computer, a speaker profile associated with a speaker containing one or more embeddings for the speaker;

determining, by the computer, a level of maturity for a voiceprint for the speaker based upon a false acceptance rate and one or more maturity factors for the one or more embeddings; and

updating, by the computer, one or more similarity thresholds of the speaker profile according to the level of maturity and the one or more maturity factors.

72. The method according to claim 71, wherein obtaining the speaker profile comprises:

generating, by the computer, the speaker profile in response to determining that the speaker is a new user of a media content system.

73. The method according to claim 71, wherein obtaining the speaker profile comprises:

extracting, by the computer, an inbound embedding for the speaker by applying an machine learning architecture on an inbound audio signal;

generating, by the computer, a similarity score for the inbound embedding based upon a relative distance between the inbound embedding and the voiceprint; and

identifying, by the computer, the speaker profile stored in a speaker database in response to determining that the similarity score for the inbound embedding satisfies at least one similarity threshold of the one or more similarity thresholds of the speaker profile.

74. The method according to claim 73, wherein the at least one threshold includes a low similarity threshold, and wherein the one or more thresholds includes the low similarity threshold and a high similarity threshold.

75. The method according to claim 73, further comprising updating, by the computer, the voiceprint for the speaker according to the inbound embedding in response to determining that the similarity score for the inbound embedding satisfies a high similarity threshold of the one or more similarity thresholds, wherein the computer determines the level of maturity for the voiceprint after updating the voiceprint.

76. The method according to claim 75, further comprising updating, by the computer, the one or more maturity factors for the voiceprint based upon the inbound embedding used for updating the voiceprint, wherein the computer determines the level of maturity for the voiceprint after updating the one or more maturity factors.

77. The method according to claim 71, further comprising, responsive to the computer determining that the level of maturity of the voiceprint fails to satisfy a maturity threshold:

generating, by the computer, a prompt requesting an additional inbound embedding associated with the speaker; and

for each additional inbound embedding satisfying the one or more similarity thresholds:

updating, by the computer, the voiceprint according to the additional inbound embedding; and

updating, by the computer, the one or more maturity factors for the voiceprint based upon each additional inbound embedding.

78. The method according to claim 71, wherein updating the one or more similarity thresholds includes:

increasing, by the computer, the one or more thresholds of the speaker profile in response to determining that the level of maturity for the voiceprint satisfies a maturity threshold.

79. The method according to claim 71, wherein the one or more maturity factors include at least one of a number of embeddings for the speaker, a duration of net speech occurring in the one or more embeddings, and a quality of audio from which the one or more embeddings.

80. The method according to claim 71, further comprising determining, by the computer, the false acceptance rate based upon a configuration input received from an administrative computer.

81. A system comprising:

a speaker profile database comprising non-transitory machine-readable media configured to store a plurality of speaker profiles; and

a computer comprising a processor configured to:

obtain a speaker profile associated with a speaker containing one or more embeddings for the speaker;

determine a level of maturity for a voiceprint for the speaker based upon a false acceptance rate and one or more maturity factors for the one or more embeddings; and

update one or more similarity thresholds of the speaker profile according to the level of maturity and the one or more maturity factors.

82. The system according to claim 81, wherein to obtain the speaker profile the computer is configured to:

generate the speaker profile in response to determining that the speaker is a new user of a media content system.

83. The system according to claim 81, wherein to obtain the speaker profile the computer is further configured to:

extract an inbound embedding for the speaker by applying an machine learning architecture on an inbound audio signal;

generate a similarity score for the inbound embedding based upon a relative distance between the inbound embedding and the voiceprint; and

identify the speaker profile stored in a speaker database in response to determining that the similarity score for the inbound embedding satisfies at least one similarity threshold of the one or more similarity thresholds of the speaker profile.

84. The system according to claim 83, wherein the at least one threshold includes a low similarity threshold, and wherein the one or more thresholds includes the low similarity threshold and a high similarity threshold.

85. The system according to claim 83, wherein the computer is further configured to:

update the voiceprint for the speaker according to the inbound embedding in response to determining that the similarity score for the inbound embedding satisfies a high similarity threshold of the one or more similarity thresholds, wherein the computer determines the level of maturity for the voiceprint after updating the voiceprint.

86. The system according to claim 85, wherein the computer is further configured to:

update the one or more maturity factors for the voiceprint based upon the inbound embedding used for updating the voiceprint, wherein the computer determines the level of maturity for the voiceprint after updating the one or more maturity factors.

87. The system according to claim 81, wherein the computer is further configured to:

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

generate a prompt requesting an additional inbound embedding associated with the speaker; and

for each additional inbound embedding satisfying the one or more similarity thresholds:

update the voiceprint according to the additional inbound embedding; and

update the one or more maturity factors for the voiceprint based upon each additional inbound embedding.

88. The system according to claim 81, wherein to update the one or more similarity thresholds the device is further configured to:

increase the one or more thresholds of the speaker profile in response to determining that the level of maturity for the voiceprint satisfies a maturity threshold.

89. The system according to claim 81, wherein the one or more maturity factors include at least one of a number of embeddings for the speaker, a duration of net speech occurring in the one or more embeddings, and a quality of audio from which the one or more embeddings.

90. The system according to claim 81, wherein the computer is further configured to determine the false acceptance rate based upon a configuration input received from an administrative computer.

91. A device-implemented method comprising:

receiving, by a device, an inbound audio signal containing an utterance of an inbound speaker;

applying, by the device, an embedding extraction model on the inbound audio signal to extract an inbound embedding for the inbound speaker;

generating, by the device, one or more similarity scores for the inbound embedding upon a relative distance between the inbound embedding and one or more voiceprints stored non-transitory machine readable media;

identifying, by the computer, a speaker identifier associated with a voiceprint for the inbound speaker in response to determining that a similarity score generated using the voiceprint satisfies a similarity threshold; and

transmitting, by the device, the speaker identifier to a content server.

92. The method according to claim 91 wherein the device is at least one of a smart TV, a set-top box, an edge device, a remote control, and a mobile communication device.

93. The method according to claim 91, further comprising receiving from the content server, by the device, media content for display via a display device coupled to the device.

94. The method according to claim 91, wherein the device transmits a request for media content to the content server with the speaker identifier.

95. The method according to claim 91, further comprising extracting, by the device, one or more features from the inbound audio signal, the features having one or more types of biometric features including at least one audio feature.

96. The method according to claim 91, further comprising:

receiving, by a microphone of the device, audio waves containing a plurality of utterances from a plurality of inbound speakers; and

converting, by the device, the audio waves into the inbound audio signal containing the plurality of utterances.

97. The method according to claim 91, wherein the inbound audio signal contains a plurality of utterances of a plurality of inbound speakers, and

wherein the device applies the embedding extraction model on the inbound audio signal to generate a plurality of similarity scores for the plurality of inbound embeddings corresponding to the plurality of inbound speakers, the plurality of similarity scores generated according to a plurality of voiceprints stored in non-transitory machine-readable storage of the device.

98. The method according to claim 97, further comprising:

responsive to the device determining that a second similarity score of the plurality of similarity scores for a second inbound embedding of the plurality of embeddings fails to satisfy at least one similarity threshold:

generating, by the device, a second speaker profile for the second inbound speaker containing the second inbound embedding and a second speaker identifier in the non-transitory machine-readable storage.

99. The method according to claim 91, further comprising:

generating, by the device, one or more enrollment prompts for display to the inbound speaker requesting one or more enrollment audio signals for the inbound speaker;

generating, by the device, the voiceprint for the inbound speaker based upon one or more enrollment embeddings extracted from the one or more enrollment audio signals;

responsive to the device determining that a level of maturity of the voiceprint fails to satisfy a maturity threshold:

generating, by the device, a second enrollment prompt requesting a second enrollment audio signal; and

updating, by the device, the voiceprint for the inbound speaker based upon a second enrollment embedding extracted from the second enrollment audio signal.

100. The method according to claim 91, further comprising:

extracting, by the computer, one or more acoustic features from the inbound audio signal; and

calculating, by the computer, a spoofing score indicating a likelihood that the inbound audio signal includes a spoofing condition based upon the one or more acoustic features by applying a second machine-learning model.

101. A system comprising:

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

a device comprising a processor configured to:

receive an inbound audio signal containing an utterance of an inbound speaker;

apply an embedding extraction model on the inbound audio signal to extract an inbound embedding for the inbound speaker;

generate one or more similarity scores for the inbound embedding upon a relative distance between the inbound embedding and one or more voiceprints stored the speaker database;

identifying, by the computer, a speaker identifier associated with a voiceprint for the inbound speaker in response to determining that a similarity score generated using the voiceprint satisfies a similarity threshold; and

transmit the speaker identifier to a content server.

102. The system according to claim 101, wherein the device is at least one of a smart TV, a set-top box, an edge device, a remote control, and mobile communication device.

103. The system according to claim 101, wherein the device is further configured to receive media content from the content server for display via a display device coupled to the device.

104. The system according to claim 101, wherein the device is configured to transmit a request for media content to the content server with the speaker identifier.

105. The system according to claim 101, wherein the device is further configured to:

extract one or more features from the inbound audio signal, the features having one or more types of biometric features including at least one audio feature.

106. The system according to claim 101, further comprising a microphone configured to receive audio waves containing a plurality of utterances from a plurality of inbound speakers; and

wherein the device is further configured to convert the audio waves into the inbound audio signal containing the plurality of utterances.

107. The system according to claim 101, wherein the inbound audio signal contains a plurality of utterances of a plurality of inbound speakers,

wherein the device is configured to apply the embedding extraction model on the inbound audio signal to generate a plurality of similarity scores for the plurality of inbound embeddings corresponding to the plurality of inbound speakers, and wherein the device generates the plurality of similarity scores according to a plurality of voiceprints stored in the speaker database.

108. The system according to claim 107, wherein the device is further configured to:

determine that a second similarity score of the plurality of similarity scores for a second inbound embedding of the plurality of embeddings fails to satisfy at least one similarity threshold; and

generate a second speaker profile for the second inbound speaker containing the second inbound embedding and a second speaker identifier in the non-transitory machine-readable storage.

109. The system according to claim 101, wherein the device is further configured to:

generate one or more enrollment prompts for display to the inbound speaker requesting one or more enrollment audio signals for the inbound speaker;

generate the voiceprint for the inbound speaker based upon one or more enrollment embeddings extracted from the one or more enrollment audio signals;

responsive to the device determining that a level of maturity of the voiceprint fails to satisfy a maturity threshold:

generate a second enrollment prompt requesting a second enrollment audio signal; and

update the voiceprint for the inbound speaker based upon a second enrollment embedding extracted from the second enrollment audio signal.

~~**110.** The system according to claim 101, wherein the device is further configured to:~~

~~extract one or more acoustic features from the inbound audio signal; and~~

~~calculate a spoofing score indicating a likelihood that the inbound audio signal includes a spoofing condition based upon the one or more acoustic features by applying a second machine-learning model.~~