CLAIMS

1. A computer-implemented method comprising:

segmenting a data input into sequential n-gram chunks based on a predetermined rule, wherein the data input is received through speech recognition;

receiving metadata regarding characteristics of the data input;

generating a language model based on the metadata;

generating a first set of language model variations of the data input;

generating, in response to generating the first set of language model variations, a second set of language model variations;

generating a plurality of template sentences for the subset of the first set and the second set of language model variations;

training the language model based on at least the first set of language model variations;

generating one or more alternatives for the data input using the trained language model; and

transmitting an output comprising the one or more alternatives for the data input.

2. The method of claim 1, wherein the generating the first set of language model variations is based on at least the metadata and the predetermined rule.

3. The method of claim 2, wherein the generating the plurality of template sentences is included in a process of creating prepended text and appended text for a subset of the first set and the second set of language model variations.

4. The method of claim 3, wherein the training the language model is further based on the second set of language model variations, the prepended text, and the appended text.

5. The method of claim 3, wherein creating the prepended text and appended text further comprises:

analyzing the common user metadata; and

determining, based on the analysis, common prepended phrases and appended phrases used when speech is inputted.

6. The method of claim 1, wherein generating the first set of language model variations comprises:

determining a number of possible variations for the data input.

7. The method of claim 6, further comprising:

determining that the number of possible variations for the data input is below a threshold number of variations; and

generating, in response to the determining, all possible language model variations of the data input.

8. The method of claim 6, further comprising:

determining that the number of possible variations for the data input is above a threshold number of variations; and

generating, in response to the determining, a reduced number of variations for the data input based on the n-gram chunks.

9. The method of claim 6, wherein generating the first set of language model variations of the data input comprises:

generating alternate words that may represent a number based on common digit patterns.

10. The method of claim 1, wherein the n-gram chunks are one of 4-gram chunks and 5-gram chunks.

11. The method of claim 1, wherein the data input is alphanumeric.

12. A system having one or more computer processors, the system configured to:

segment a data input into a sequential n-gram chunks based on a predetermined rule, wherein the data input is received through speech recognition;

receive metadata regarding characteristics of the data input;

generate a language model based on the metadata;

generate a first set of language model variations of the data input;

generate, in response to generating the first set of language model variations, a second set of language model variations;

generate a plurality of template sentences for the subset of the first set and the second set of language model variations;

train the language model based on at least the first set of language model variations;

generate one or more alternatives for the data input using the trained language model; and

transmit an output comprising the one or more alternatives for the data input.

13. The system of claim 12, wherein the generating the first set of language model variations is based on at least the metadata and the predetermined rule.

14. The system of claim 13, wherein the generating the plurality of template sentences is included in a process of creating prepended text and appended text for a subset of the first set and the second set of language model variations.

15. The system of claim 14, wherein the training the language model is further based on the second set of language model variations, the prepended text, and the appended text.

16. The system of claim 14, wherein creating the prepended text and appended text further comprises:

analyzing the common user metadata; and

determining, based on the analysis, common prepended phrases and appended phrases used when speech is inputted.

17. A computer program product comprising a computer readable storage medium having program instructions embodied therewith, the program instructions executable by a server to cause the server to perform a method, the method comprising:

segmenting a data input into a sequential n-gram chunks based on a predetermined rule, wherein the data input is received through speech recognition;

receiving metadata regarding characteristics of the data input;

generating a language model based on the metadata;

generating a first set of language model variations of the data input;

generating, in response to generating the first set of language model variations, a second set of language model variations;

generating a plurality of template sentences for the subset of the first set and the second set of language model variations;

training the language model based on at least the first set of language model variations;

generating one or more alternatives for the data input using the trained language model; and

transmitting an output comprising the one or more alternatives for the data input.

18. The computer program product of claim 17, wherein the generating the first set of language model variations is based on at least the metadata and the predetermined rule.

19. The computer program product of claim 18, wherein the generating the plurality of template sentences is included in a process of creating prepended text and appended text for a subset of the first set and the second set of language model variations.

20. The computer program product of claim 19, wherein creating the prepended text and appended text further comprises:

analyzing the common user metadata; and

determining, based on the analysis, common prepended phrases and appended phrases used when speech is inputted.