CLAIMS

1. A method for monitoring and detecting bias in an application, the method comprising:

indexing, by a bias monitoring computing system, training data;

obtaining, by the bias monitoring computing system , a plurality of correlation values of one or more features in the indexed training data with a target variable; and

for each of the one or more features:

calculating, by the bias monitoring computing system, a first value of correlation of the plurality of correlation values of the one of the one or more features and a favorable result of a judgement by a trained artificial intelligence model;

calculating, by the bias monitoring computing system, a second value of correlation of the plurality of correlation values of the one of the one or more features and a unfavorable result of a judgement by the trained artificial intelligence model;

calculating, by the bias monitoring computing system, an absolute value of a difference between the calculated first value and the calculated second value;

calculating, by the bias monitoring computing system, a total sum of the calculated absolute value of the plurality of correlation values of the one of the one or more features; and

identifying and monitoring, by the bias monitoring computing system, a bias in the one or more features using the calculated first value, the second value, the absolute value, and the total sum.

2. The method as set forth in claim 1, further comprising: indexing, by the bias monitoring computing system, feedback data received from a client device.

3. The method as set forth in claim 1, further comprising, obtaining, by the bias monitoring computing system, training data from a training data server.

4. The method as set forth in claim 2, further comprising, calculating, by the bias monitoring computing system, a second total sum of the calculated absolute value of the plurality of correlation values of the one of the one or more features for the feedback data.

5. The method as set forth in claim 4, further comprising, presenting, by the bias monitoring computing system, a correlation change value determined by a difference between the calculated total sum of the calculated absolute value associated with the training data and the calculated total sum of the calculated absolute value associated with the feedback data.

6. The method as set forth in claim 5, wherein the correlation change value represents a candidate of a feature to be monitored.

7. A non-transitory machine readable medium having stored thereon instructions comprising machine executable code which when executed by at least one machine causes the machine to:

index training data;

obtain a plurality of correlation values of one or more features in the indexed training data with a target variable; and

for each of the one or more features:

calculate a first value of correlation of the plurality of correlation values of the one of the one or more features and a favorable result of a judgement by a trained artificial intelligence model;

calculate a second value of correlation of the plurality of correlation values of the one of the one or more features and a unfavorable result of a judgement by the trained artificial intelligence model;

calculate an absolute value of a difference between the calculated first value and the calculated second value;

calculate a total sum of the calculated absolute value of the plurality of correlation values of the one of the one or more features; and

identify and monitor a bias in the one or more features using the calculated first value, the second value, the absolute value, and the total sum.

8. The medium as set forth in claim 7, further comprising, indexing feedback data received from a client device.

9. The medium as set forth in claim 7, further comprising, obtaining training data from a training data server.

10. The medium as set forth in claim 8, further comprising, calculating a second total sum of the calculated absolute value of the plurality of correlation values of the one of the one or more features for the feedback data.

11. The medium as set forth in claim 10, further comprising, presenting a correlation change value determined by a difference between the calculated total sum of the calculated absolute value associated with the training data and the calculated total sum of the calculated absolute value associated with the feedback data.

12. The medium as set forth in claim 11, wherein the correlation change value represents a candidate of a feature to be monitored.

13. A bias monitoring computing system, comprising a memory comprising programmed instructions stored in the memory and one or more processors configured to be capable of executing the programmed instructions stored in the memory to:

index training data;

obtain a plurality of correlation values of one or more features in the indexed training data with a target variable; and

for each of the one or more features:

calculate a first value of correlation of the plurality of correlation values of the one of the one or more features and a favorable result of a judgement by a trained artificial intelligence model;

calculate a second value of correlation of the plurality of correlation values of the one of the one or more features and a unfavorable result of a judgement by the trained artificial intelligence model;

calculate an absolute value of a difference between the calculated first value and the calculated second value;

calculate a total sum of the calculated absolute value of the plurality of correlation values of the one of the one or more features; and

identify and monitor a bias in the one or more features using the calculated first value, the second value, the absolute value, and the total sum.

14. The system as set forth in claim 13 wherein the one or more processors are further configured to be capable of executing the programmed instructions stored in the memory to index feedback data received from a client device.

15. The system as set forth in claim 13 wherein the one or more processors are further configured to be capable of executing the programmed instructions stored in the memory to obtain training data from a training data server.

16. The system as set forth in claim 14 wherein the one or more processors are further configured to be capable of executing the programmed instructions stored in the memory to calculate a second total sum of the calculated absolute value of the plurality of correlation values of the one of the one or more features for the feedback data.

17. The system as set forth in claim 16 wherein the one or more processors are further configured to be capable of executing the programmed instructions stored in the memory to present a correlation change value determined by a difference between the calculated total sum of the calculated absolute value associated with the training data and the calculated total sum of the calculated absolute value associated with the feedback data.

18. The system as set forth in claim 17 wherein the correlation change value represents a candidate of a feature to be monitored.

19. The system as set forth in claim 13 wherein the one or more processors are further configured to be capable of executing the programmed instructions stored in the memory to provide the identified and monitored the one or more features with the bias to a client device.

20. The system as set forth in claim 19 wherein the one or more processors are further configured to be capable of executing the programmed instructions stored in the memory to provide a favorable or an unfavorable decision for the provided one or more features with the bias.

21. A method for monitoring and detecting bias in an application, the method comprising:

indexing, by a bias monitoring computing system, training data;

obtaining, by the bias monitoring computing system , a plurality of correlation values of one or more features in the indexed training data with a target variable, wherein for each of the one or more features:

calculating a first value of correlation of the plurality of correlation values of the one of the one or more features and a favorable result of a judgement by a trained artificial intelligence model;

calculating a second value of correlation of the plurality of correlation values of the one of the one or more features and a unfavorable result of a judgement by the trained artificial intelligence model;

calculating, by the bias monitoring computing system, an absolute value of a difference between the calculated first value and the calculated second value; and

calculating, by the bias monitoring computing system, a total sum of the calculated absolute value of the plurality of correlation values of the one of the one or more features;

Identifying, by a bias monitoring computing system, a feature from one or more features of a software application;

determining, by the bias monitoring computing system, if there is a bias in the identified feature from the one or more features of the software application;

determining, by the bias monitoring computing system, a favorable or an unfavorable decision when there is the bias determined in the identified feature; and

providing, by the bias monitoring computing system, the identified feature with the determined bias and the determined favorable or an unfavorable decision.

22. The method as set forth in claim 21 further comprising, identifying and monitoring, by the bias monitoring computing system, a bias in the one or more features using the calculated first value, the second value, the absolute value, and the total sum.