WHAT IS CLAIMED IS:

1. A method comprising:

a) determining, by a data processor, that a malicious computer uses a machine learning model with a boundary function for classification of input data in order to determine outcomes;

b) generating, by the data processor, normal data, wherein the normal data includes typical input data corresponding to the type of machine learning model;

c) providing, by the data processor, the normal data to the malicious computer;

d) obtaining, by the data processor, classification data from the malicious computer, wherein the malicious computer uses the machine learning model to generate the classification data based on the normal data;

e) determining, by the data processor, based on the classification data, characteristics of the machine learning model of the malicious computer;

f) generating, by the data processor, based on the characteristics of the machine learning model of the malicious computer, transition data to be used as an input to the machine learning model and intended to shift the boundary function;

g) providing, by the data processor to the malicious computer, the transition data; and

h) repeating, by the data processor, steps f) and g), thereby causing the boundary function to shift over time.

2. The method of claim 1, wherein the characteristics of the machine learning model include separations between clusters of different training data points, an estimate of the boundary function, or a plurality of labels assigned to the normal data.

3. The method of claim 2, wherein the machine learning model is a support vector machine, wherein the boundary function is a hyperplane, and wherein the hyperplane separates a plurality of classifications.

4. The method of claim 3, wherein the plurality of labels are associated with the plurality of classifications.

5. The method of claim 4, wherein the boundary function shifts causing the plurality of classifications to shift.

6. The method of claim 1, wherein generating the normal data further comprises:

retrieving, by the data processor, the normal data from a data source; and

classifying, by the data processor, the normal data.

7. The method of claim 1, wherein the transition data includes data items that share characteristics of data belonging to more than one classification.

8. The method of claim 1, wherein providing the transition data further comprises:

providing, by the data processor, the transition data to a data source, wherein the malicious computer retrieves the transition data from the data source.

9. The method of claim 1, wherein the machine learning model includes linear regression, logistic regression, decision trees, support vector machines, naive Bayes, kNN, K-means, or random forests.

10. A data processor comprising:

a processor;

a memory device; and

a computer-readable medium coupled to the processor, the computer-readable medium comprising code executable by the processor for implementing a method comprising:

a) determining that a malicious computer uses a machine learning model with a boundary function for classification of input data in order to determine outcomes;

b) generating normal data, wherein the normal data includes typical input data corresponding to the type of machine learning model;

c) providing the normal data to the malicious computer;

d) obtaining classification data from the malicious computer, wherein the malicious computer uses the machine learning model to generate the classification data based on the normal data;

e) determining based on the classification data, characteristics of the machine learning model of the malicious computer;

f) generating transition data intended to shift the boundary function;

g) providing, to the malicious computer, the transition data; and

h) repeating steps f) and g), thereby causing the boundary function to shift over time.

11. The data processor of claim 10, wherein the characteristics of the machine learning model include separations between clusters of different training data points, an estimate of the boundary function, or a plurality of labels assigned to the normal data.

12. The data processor of claim 11, wherein the machine learning model is a support vector machine, wherein the boundary function is a hyperplane, and wherein the hyperplane separates a plurality of classifications.

13. The data processor of claim 12, wherein the plurality of labels are associated with the plurality of classifications.

14. The data processor of claim 13, wherein the boundary function shifts causing the plurality of classifications to shift.

15. The data processor of claim 10, wherein generating the normal data further comprises:

retrieving the normal data from a data source; and

classifying the normal data.

16. The data processor of claim 10, wherein the transition data includes data items that share characteristics of data belonging to more than one classification.

17. The data processor of claim 10, wherein providing the transition data further comprises:

providing the transition data to a data source, wherein the malicious computer retrieves the transition data from the data source.

18. The data processor of claim 10, wherein the machine learning model includes linear regression, logistic regression, decision trees, support vector machines, naive Bayes, kNN, K-means, or random forests.