CLAIMS

1. A system, comprising:

a computing device comprising a processor and a memory;

a training dataset stored in the memory, the training dataset comprising a plurality of records; and

a first machine-learning model stored in the memory that, when executed by the processor, cause the computing device to at least:

analyze the training dataset to select a probability distribution function (PDF) associated with the plurality of records based least in part on a metric associated with the first machine learning model; and

generate a new record based at least in part on the PDF that has been selected; and

a second machine-learning model stored in the memory that, when executed by the processor, cause the computing device to at least:

analyze the training dataset to identify similarities between the plurality of records;

evaluate the new record generated by the first machine-learning model to determine whether the new record is indistinguishable from at least a subset of the plurality of records in the training data set based at least in part on a predefined error rate;

update the first machine-learning model based at least in part on the evaluation of the new record; and

update the second machine-learning model based at least in part on the evaluation of the new record.

2. The system of claim 1, wherein:

the first machine-learning model causes the computing device to generate a plurality of new records; and

the system further comprises a third machine-learning model stored in the memory that is trained using the plurality of new records generated by the first machine-learning model.

3. The system of claim 1 or 2, wherein the plurality of new records are generated in response to a determination that the second machine-learning model is unable to distinguish between the new record generated by the first machine-learning model and individual ones of the plurality of records in the training dataset.

4. The system of claims 1-3, wherein the plurality of new records are generated from a random sample of a predefined number of points in a sample space defined by the PDF selected by the first machine-learning model.

5. The system of claims 1-4, wherein the first machine-learning model repeatedly generates the new record until the second machine-learning model is unable to distinguish the new record from the plurality of records in the training data set at a predefined rate.

6. The system of claims 1-5, wherein the predefined rate is fifty percent when equal size new records are created.

7. The system of claims 1-6, wherein the first machine-learning model causes the computing device to generate the new record at least twice and the second machine-learning model causes the computing device to evaluate the new record at least twice, update the first machine-learning model at least twice, and update the second machine-learning model at least twice.

8. A computer-implemented method, comprising:

analyzing a plurality of original records to select a probability distribution function (PDF) based at least in part on machine learning model metrics, wherein the PDF comprises a sample space, and the sample space comprises the plurality of original records;

generating a plurality of new records using the PDF;

creating an augmented dataset that comprises the plurality of new records; and

training a machine-learning model using the augmented dataset.

9. The computer-implemented method of claim 8, wherein analyzing the plurality of original records to identify the probability distribution function further comprises:

training a generator machine-learning model to create a new record that is similar to individual ones of the plurality of original records;

training a discriminator machine-learning model to distinguish between the new record and the individual ones of the plurality of original records; and

identifying the probability distribution function in response to the new record created by the generator machine-learning model being mistaken by the discriminator machine-learning model at a predefined rate.

10. The computer-implemented method of claim 9, wherein the predefined rate is approximately fifty percent of comparisons performed by the discriminator between the new record and the plurality of original records.

11. The computer-implemented method of claim 9 or 10, wherein the generator machine-learning model is a first generator machine-learning model, the first generator machine-learning model and at least a second generator machine-learning model are included in a plurality of generator machine-learning models and the method further comprises:

training at least the second generator machine-learning model to create the new record that is similar to individual ones of the plurality of original records; and

selecting the first generator machine-learning model from the plurality of generator machine-learning models based at least in part on:

a run length associated with each generator machine-learning model and the discriminator machine-learning model,

a generator loss rank associated with each generator machine-learning model and the discriminator machine-learning model,

a discriminator loss rank associated with each generator machine-learning model and the discriminator machine-learning model,

a different rank associated with each generator machine-learning model and the discriminator machine-learning model, or

at least one result of a Kolmogorov-Smirnov (KS) test that includes a first probability distribution function associated with the plurality of original records and a second probability distribution function associated with the plurality of new records

wherein the identifying of the probability distribution function further occurs in response to selecting the first generator machine-learning model from the plurality of generator machine-learning models.

12. The computer-implemented method of claims 8-11, wherein generating the plurality of new records using the probability distribution function further comprises randomly selecting a predefined number of points in the sample space defined by the probability distribution function.

13. The computer-implemented method of claims 8-12, further comprising adding the plurality of original records to the augmented dataset.

14. The computer-implemented method of claims 8-13, wherein the machine-learning model comprises a neural network.

15. A system, comprising:

a computing device comprising a processor and a memory; and

machine-readable instructions stored in the memory that, when executed by the processor, cause the computing device to at least:

analyze a plurality of original records to select a probability distribution function (PDF) based least in part on a machine learning model metric, wherein the PDF comprises a sample space, and the sample space comprises the plurality of original records;

generate a plurality of new records using the PDF;

create an augmented dataset that comprises the plurality of new records; and

train a machine-learning model using the augmented dataset.

16. The system of claim 15, wherein the machine-readable instructions that cause the computing device to analyze the plurality of original records to identify the probability distribution function further cause the computing device to at least:

train a generator machine-learning model to create a new record that is similar to individual ones of the plurality of original records;

train a discriminator machine-learning model to distinguish between the new record and the individual ones of the plurality of original records; and

identify the probability distribution function in response to the new record created by the generator machine-learning model being mistaken by the discriminator machine-learning model at a predefined rate.

17. The system of claim 16, wherein the predefined rate is approximately fifty percent of comparisons performed by the discriminator between the new record and the plurality of original records.

18. The system of claim 16 or 17, wherein the generator machine-learning model is a first generator machine-learning model, the first generator machine-learning model and at least a second generator machine-learning model are included in a plurality of generator machine-learning models and the machine-readable instructions further cause the computing device to at least:

train at least the second generator machine-learning model to create the new record that is similar to individual ones of the plurality of original records; and

select the first generator machine-learning model from the plurality of generator machine-learning models based at least in part on:

a run length associated with each generator machine-learning model and the discriminator machine-learning model,

a generator loss rank associated with each generator machine-learning model and the discriminator machine-learning model,

a discriminator loss rank associated with each generator machine-learning model and the discriminator machine-learning model,

a different rank associated with each generator machine-learning model and the discriminator machine-learning model, or

at least one result of a Kolmogorov-Smirnov (KS) test that includes a first probability distribution function associated with the plurality of original records and a second probability distribution function associated with the plurality of new records

wherein the identification of the probability distribution function further occurs in response to selecting the first generator machine-learning model from the plurality of generator machine-learning models.

19. The system of claims 15-18, wherein the machine-readable instructions that cause the computing device to generate the plurality of new records using the probability distribution function further cause the computing device to randomly select a predefined number of points in the sample space defined by the probability distribution function.

20. The system of claims 15-19, wherein the machine-readable instructions, when executed by the processor, further cause the computing device to at least add the plurality of original records to the augmented dataset.