CLAIMS

What is claimed is:

1. A computer-implemented method, comprising:

receiving data associated with an electronic communication;

determining one or more attributes associated with the received data;

determining, based on the one or more attributes, one or more classification labels;

creating a machine learning model based on the one or more attributes associated with the received data and the one or more classification labels, wherein

the machine learning model identifies one or more patterns associated with a first violation condition,

the machine learning model comprises a key indicator (KI) model, and

the machine learning model is pre trained for analysis of text data associated with one or more electronic communications;

creating a lexicon comprising one or more terms associated with a second violation condition;

creating a scenario by joining the first violation condition and the second violation condition with one or more Boolean operators; and

configuring a computer system to detect violations in a target dataset using the scenario, wherein the target dataset comprises electronic communications.

2. The computer-implemented method of claim 1, wherein the scenario includes a filter configured to exclude a portion of the target dataset.

3. The computer-implemented method of claim 2, wherein the filter is configured to exclude portions of the target dataset with certain types of electronic communications.

4. The computer-implemented method of any one of claims 1 to 3, wherein the method includes storing the scenario in a computer readable medium.

5. The computer-implemented method of claim 4, wherein the method includes comparing the stored scenario to a second stored scenario, and based on the comparison, outputting data representing the differences between the stored scenario and the second stored scenario.

6. The computer-implemented method of any one of claims 1 to 5, wherein the received data comprises at least one of text data and metadata associated with the electronic communications.

7. The computer implemented method of any one of claims 1 to 5, wherein the target dataset comprises at least one of text data and metadata associated with the electronic communications.

8. The computer-implemented method of any one of claims 1 to 7, further comprising determining whether the received data includes a segment of target language.

9. A non-transitory, computer-readable medium comprising executable instructions which, when executed by a processor, perform functions that comprise:

receiving data associated with an electronic communication;

determining one or more attributes associated with the received data;

determining, based on the one or more attributes, one or more classification labels;

creating a machine learning model based on the one or more attributes associated with the received data and the one or more classification labels, wherein

the machine learning model identifies one or more patterns associated with a first violation condition,

the machine learning model comprises a key indicator (KI) model, and

the machine learning model is pre trained for analysis of text data associated with one or more electronic communications;

creating a lexicon comprising one or more terms associated with a second violation condition;

creating a scenario by joining the first violation condition and the second violation condition with one or more Boolean operators; and

configuring a computer system to detect violations in a target dataset using the scenario, wherein the target dataset comprises electronic communications.

10. The computer-readable medium of claim 9, wherein the executable instructions further comprise a scenario that includes a filter configured to exclude a portion of the target dataset.

11. The computer-readable medium of claim 10, wherein the filter is configured to exclude portions of the target dataset with certain types of electronic communications.

12. The computer-readable medium of any one of claims 9 to 11, wherein the executable instructions further comprise, when executed, storing the scenario in a computer readable medium.

13. The computer-readable medium of claim 12, wherein the executable instructions comprise, when executed, comparing the stored scenario to a second stored scenario, and based on the comparison, outputting data representing the differences between the stored scenario and the second stored scenario.

14. The computer-readable medium of any one of claims 9 to 13, wherein the received data comprises at least one of text data and metadata associated with the electronic communications.

15. The computer-readable medium of any one of claims 9 to 13, wherein the target dataset comprises at least one of text data and metadata associated with the electronic communications.

16. The computer-readable medium of any one of claims 9 to 15, wherein the executable instructions comprise, when executed, determining whether the received data includes a segment of target language.

17. A system comprising:

one or more processors; and

memory coupled with the one or more processors, the memory storing executable instructions that when executed by the one or more processors cause the one or more processors to effectuate operations comprising:

receiving data associated with an electronic communication;

determining one or more attributes associated with the received data;

determining, based on the one or more attributes, one or more classification labels;

creating a machine learning model based on the one or more attributes associated with the received data and the one or more classification labels, wherein

the machine learning model identifies one or more patterns associated with a first violation condition,

the machine learning model comprises a key indicator (KI) model, and

the machine learning model is pre trained for analysis of text data associated with one or more electronic communications;

creating a lexicon comprising one or more terms associated with a second violation condition;

creating a scenario by joining the first violation condition and the second violation condition with one or more Boolean operators; and

configuring a computer system to detect violations in a target dataset using the scenario, wherein the target dataset comprises electronic communications.