

## CLAIMS

1. A computer-implemented method for selecting a dataset from given datasets for updating an artificial intelligence module (AI-module), the given datasets comprising each an input dataset and a corresponding output dataset, the method comprising:
  - obtaining values of parameters for defining different clusters of the given datasets;
  - determining a metric of each given dataset, the metric of each given dataset being dependent on a level of membership of the respective given dataset to one of the clusters and a distance of the respective given dataset to a centroid of the same one of the clusters; and
  - selecting at least one of the given datasets from the given datasets for updating the AI-module on the basis of a comparison of the metrics of the given datasets.
2. The computer-implemented method of claim 1, further comprising:
  - determining a metric of each cluster, the metric of each cluster being dependent on a distance of a centroid of the respective cluster to other centroids of the clusters;
  - selecting at least one of the clusters from the clusters on the basis of the metrics of the clusters; and
  - determining the metric of each given dataset, the metric of each given dataset being dependent on the level of membership of the respective given dataset to the selected cluster and the distance of the respective given dataset to the centroid of the selected cluster.
3. The computer-implemented method of claim 1, further comprising determining of the metric for each given dataset based, at least in part, on:
  - determining a set of metrics for each given dataset, each metric of the set of metrics of the respective given dataset corresponding to one cluster of a subset of the clusters, each metric of the set of metrics of the respective given dataset being dependent on the level of membership of the respective given dataset to the corresponding cluster and the distance of the respective given dataset to a centroid of the corresponding cluster; and
  - selecting at least one of the given datasets from the given datasets for updating the AI-module on the basis of a comparison of the set of metrics of the given datasets.

4. The computer-implemented method of claim 1, further comprising:  
generating the values of the parameters for defining the clusters as a function of training datasets, the AI-module being generated using the training datasets.
5. The computer-implemented method of claim 1, further comprising:  
generating the values of the parameters for defining the clusters as a function of the given datasets.
6. The computer-implemented method of claim 1, further comprising:  
generating the values of the parameters for defining the clusters as a function of test datasets, the AI-module being tested using the test datasets.
7. The computer-implemented method of claim 1, further comprising:  
generating the values of the parameters for defining the clusters as a function of an approved or corrected dataset of the given datasets.
8. The computer-implemented method of claim 1, further comprising:  
generating the values of the parameters for defining the clusters as a function of a manually approved or manually corrected dataset of the given datasets.
9. The computer-implemented method of claim 1, further comprising:  
obtaining the values of parameters for defining the clusters performing the Fuzzy-C-Means clustering algorithm.
10. The computer-implemented method of claim 2, further comprising:  
determining the metric of each cluster on the basis of a mean distance of the given datasets to the centroid of the respective cluster.
11. The computer-implemented method of claim 2, further comprising:  
determining the metric of each cluster on the basis of a maximal distance of the given datasets to the centroid of the respective cluster.
12. The computer-implemented method of claim 2, further comprising:  
determining the metric of each cluster on the basis of a mean level of membership of the given datasets to the respective cluster.

13. The computer-implemented method of claim 4, further comprising:  
determining the metric of each cluster on the basis of a mean distance of the training datasets and manually approved or manually corrected datasets of the given datasets to the centroid of the respective cluster.
14. The computer-implemented method of claim 4, further comprising:  
determining the metric of each cluster on the basis of a maximal distance of the training datasets and manually approved or manually corrected datasets of the given datasets to the centroid of the respective cluster.
15. The computer-implemented method of claim 4, further comprising:  
determining the metric of each cluster on the basis of a mean level of membership of the training datasets and manually approved or manually corrected datasets of the given datasets to the respective cluster.
16. The computer-implemented method of claim 4, further comprising:  
determining the metric of each cluster on the basis of a ratio of a first sum of the number of the training datasets being comprised by the respective cluster and a number of manually approved or manually corrected datasets of the given datasets being comprised by the respective cluster and a second sum of a total number of the training datasets and a total number of manually approved or manually corrected datasets of the given datasets.
17. The computer-implemented method of claim 4, further comprising:  
obtaining the values of the parameters for defining the clusters on the basis of the output datasets of the training datasets.
18. The computer-implemented method of claim 1, wherein the input datasets of the given datasets each comprise a value of an identification parameter and the output datasets of the given datasets each comprise a value of a performance indicator.
19. A computer program product for selecting a dataset from given datasets for updating an artificial intelligence module (AI-module), the given datasets each comprising an

input dataset and a corresponding output dataset, the computer program product comprising a computer-readable storage medium having computer-readable program code embodied therewith, the computer-readable program code configured to implement a method comprising:

obtaining values of parameters for defining different clusters of the given datasets;

determining a metric of each given dataset, the metric of each given dataset being dependent on a level of membership of the respective given dataset to one of the clusters and a distance of the respective given dataset to a centroid of the same one of the clusters; and

selecting at least one of the given datasets from the given datasets for updating the AI-module on the basis of a comparison of the metrics of the given datasets.

20. A computer system for selecting a dataset from given datasets for updating an artificial intelligence module (AI-module), the given datasets each comprising an input dataset and a corresponding output dataset, the computer system comprising one or more computer processors, one or more computer readable storage media, and program instructions stored on the one or more computer readable storage media for execution by the one or more computer processors to implement a method comprising:

obtaining values of parameters for defining different clusters of the given datasets;

determining a metric of each given dataset, the metric of each given dataset being dependent on a level of membership of the respective given dataset to one of the clusters and a distance of the respective given dataset to a centroid of the same one of the clusters; and

selecting at least one of the given datasets from the given datasets for updating the AI-module on the basis of a comparison of the metrics of the given datasets.