

Service ID S00054

Location Poland, Remote



Experimentation with synthetic data generation and data augmentation techniques

Provider service

Lukasiewicz Poznanski Instytut Technologiczny (L-PIT), Poznan Supercomputing and Networking Center (PSNC)

Link to content

<https://agrifoodtef.eu/catalogue-of-services/experimentation-synthetic-data-generation-and-data-augmentation-techniques>

Type of Sector

Arable farming, Food processing, Greenhouse, Horticulture, Livestock farming, Tree Crops, Viticulture

Accepted type of products

Data, Design / Documentation

Type of service

Collection of test data, Data analysis, Data augmentation, Provision of datasets

Description

This service involves expanding training datasets using synthetic data and data augmentation techniques. This will use algorithms capable of generating synthetic data simulating certain aspects of existing datasets. Such an approach will lead to the enrichment of the existing training material. The service includes the creation of synthetic replicas based on real data, the combining of real and synthetic data, and also, for example, the creation of digital terrain maps based on ground data, which can be further enriched with synthetically generated details. As a result, we offer a comprehensive solution for more reliable development of datasets.

How can the service help you

Our service provides a wide range of possibilities to extend the training datasets held by the contractor for training AI mechanisms. The service is mainly aimed at clients with only limited training datasets for training AI in their designed technical solutions. The service provider will offer, in a first step, to evaluate, in collaboration with the client, a methodology to expand training datasets using synthetic data and data expansion techniques. Dataset augmentation operations will be carried out by methods using synthetic data or based on existing datasets. This will involve algorithms capable of generating synthetic data from fragments of existing datasets. The service offers the possibility of combining real and synthetic data and creating digital terrain maps from ground data, which can be further enhanced with synthetically generated details. Such extensions not only increase the volume and variety of data samples but also improve the generalisation capability of artificial intelligence models, offering a comprehensive solution for more reliable dataset development.

How the service will be delivered

As part of the service, we employ a methodology in line with the guidelines of commonly applicable documents (standards, directives, etc.). Based on this, we will plan and carry out actions leading to the generation of synthetic or augmented datasets. We require a specification of the customer needs as well as a reference dataset. All information is treated confidentially, with the option to sign a Non-Disclosure Agreement (NDA). If the tests related to the datasets' generation processes need to comply with specific industry standards or regulations, kindly inform us in advance. The timeframe and costs are determined individually based on the scope and complexity of the tests. We ensure flexibility and professional support at every stage of the testing process, including both physical and virtual analyses. Please feel free to contact us to discuss the details and tailor the service to your unique needs.

Service customisation

Our service to experiment with synthetic data generation and augmentation techniques is based on a structured workflow, the time, duration, and other details of which can be adjusted according to the characteristics of a problem at hand and specific customer needs. The customer needs to provide the samples of real data sets that should be augmented or serve as a reference for the synthetic data to be generated. The process is flexible and tailored to specific needs, ensuring professionalism and commitment at every stage of our collaboration. As a result, customers will obtain the synthetic or augmented datasets.