



Funded by the European Union's Horizon  
2020 programme under grand agreement  
No. 959879

# AI DATA SCHEMA TRANSFORMER

ENABLING INTEROPERABILITY

[www.glass-h2020.eu](http://www.glass-h2020.eu)  
[info@glass-h2020.eu](mailto:info@glass-h2020.eu)

## GLASS ON SOCIAL MEDIA

Follow us and stay updated on our  
latest developments!



[www.linkedin.com/  
showcase/glass-eu/](https://www.linkedin.com/showcase/glass-eu/)



[www.twitter.com/  
GLASS\\_H2020\\_EU](https://www.twitter.com/GLASS_H2020_EU)



[www.youtube.com/  
@GlassH2020](https://www.youtube.com/@GlassH2020)

# THE AI DATA SCHEMA TRANSFORMER

## TOWARD MODERNIZED PUBLIC SERVICES

Towards the modernization and ease of public services and the reduction of the associated bureaucracy, public administrations need to transform their back-offices, upgrade their existing internal processes and services, and provide privacy-preserving and secure solutions. To achieve this, it is necessary to leverage key digital enablers such as blockchain and artificial intelligence as well as capture and establish common interoperability practices. GLASS, in alignment with these principles, aims to establish a distributed solution that, through dedicated components, aims to provide an intercommunication layer trained to identify the underlying data structures from external systems, recognise and map their internal data schemas and interpret the information exchanged in a system understandable format.

The GLASS project introduces a middleware component namely “The AI Data Schema Transformer” that specialises in identifying and mapping data schemas within different types of systems and architectures and interprets them accordingly among, either centralised or distributed, heterogeneous and non-standardized systems. This approach enables a capacity for public organizations and private entities such as businesses and initiatives to integrate their operational environment within the GLASS network and leverage the services offered towards enabling interoperability between users, cross-border and cross-sector organisations.

The AI Data Schema Transformer is driven by the motivation to establish a harmonised EU interoperability framework. Specifically, it aims to act as an intercommunication layer trained to identify the underlying data structures from external systems, recognise and map their internal data schemas and interpret the information exchanged in a system understandable format. The AI data schema transformer enables interoperability among cross-border organisations, systems, modules, databases and data to dynamically adapt to the interconnected functionalities and enable information sharing among the stakeholders.

# DEVELOPMENT METHODOLOGY

## FOR HARMONIZED EU INTEROPERABILITY

The development process outlines an AI-augmented mechanism capable of identifying supported data structures from external systems, recognizing and mapping their internal data schemas, and dynamically adopting data structures and semantics for system-interpretable information exchange. The AI system processes json files with unknown tuples, inferring meaning through value classification. The pre-classifier and group classification algorithms significantly reduce the need for running classifiers, enhancing result quality and efficiency. The calibration process eliminates the constant need for running classifiers, potentially reducing costs significantly. The prototype's current stage handles identification for three evidence types across three different languages of interest (Greek, Turkish and Portuguese), with plans for standardization, finalization of file formats, additional evidence types, and integration within the GLASS ecosystem.

# CLASSIFICATION & CALIBRATION




## THE MODULES

- ✓ **Pre-classifier module:** The purpose of a pre-classifier is to break down a larger set of attributes into smaller groups according to a set of arbitrary criteria.
- ✓ **Calibration:** The calibration process changes the way that we approach this class of classification problems so that we can achieve absolute accuracy with minimal processing.
- ✓ **Grouped classification:** The Grouped Classification method efficiently processes unknown attribute-value pairs by retrieving dedicated classifiers, recording predictions with confidence levels, and dynamically filtering results based on a specified threshold, ultimately producing a list of values with recursive handling of temporary list conditions for optimal accuracy.
- ✓ **Post-calibration transformation:** The post-calibration transformation process involves presenting evidence to the system, checking if the (Country, Institution, Evidence Type) triplet has been calibrated, accessing the mapping for the triplet, mapping each attribute descriptor to its target, and finally outputting the updated evidence.

# MACHINE LEARNING MODELS

## THE METHODOLOGY

For the development of the three Machine Learning (ML) classification models (Greek, Portuguese and Turkish), which are able to classify a text into a category, we utilised the following procedure, which is composed of three phases:

	<p><b>Data partitioning</b></p> <p>Initially, the available dataset is pre-processed (data cleaning and calculation of text embeddings) and then is partitioned into two distinct datasets, namely “Training dataset” and “Evaluation dataset”.</p>
	<p><b>Hyperparameter optimization</b></p> <p>The training dataset is used for hyper-parameters optimization of the selected XGBoost algorithm. It is worth mentioning that XGBoost is an ensemble-based algorithm, which constitutes one of the most efficient and widely used ML algorithms. For the hyper-parameter optimization stratified 10-cross-validation procedure was used.</p>
	<p><b>Classification model training and evaluation</b></p> <p>After the optimization set of hyper-parameters is complete, the XGBoost is set up with an identified optimal set of parameters and trained with the “Train Dataset” for developing the classification model. Then, the developed classification model is then evaluated on the “Evaluation Dataset”.</p>



Single Sign-on eGovernment paradigm based on a distributed file exchange network for Security, transparency, cost effectiveness and trust

## GLASS puts the individual at the heart of controlling their own relationship with government & business

GLASS introduces an advanced e-government solution aligned fully with the EU eGovernment Action Plan 2016-2020 and the EU Digital Single Market strategy.

[www.glass-h2020.eu](http://www.glass-h2020.eu)  
[info@glass-h2020.eu](mailto:info@glass-h2020.eu)

### GLASS ON SOCIAL MEDIA

Follow us and stay updated on our latest developments!



[www.linkedin.com/showcase/glass-eu/](http://www.linkedin.com/showcase/glass-eu)



[www.twitter.com/GLASS\\_H2020\\_EU](http://www.twitter.com/GLASS_H2020_EU)



[www.youtube.com/@GlassH2020](http://www.youtube.com/@GlassH2020)



## GLASS SUPPORTS THE FREE MOVEMENT OF CITIZENS & BUSINESS

Free movement of EU citizens is a fundamental principle of the Article 45 of the Treaty on the Functioning of the European Union and developed by EU secondary legislation and the Case law of the Court of Justice.

### Subscribe to our Newsletter!

Stay informed about the latest developments and updates about Glass.

[www.glass-h2020.eu/#comp-ljdebrk5](http://www.glass-h2020.eu/#comp-ljdebrk5)



Funded by the European Union's Horizon 2020 programme under grand agreement No. 959879