္မား

Empowering Cultural Heritage through Artificial Intelligence

The AI4Culture project aims to develop an on-line capacity building hub for AI technologies in the cultural heritage (CH) sector. This hub takes the form of an innovative platform making CH data and tools more accessible and understandable in today's multilingual digital era, by facilitating data sharing, providing tool documentation, and promoting cultural content reuse. AI4Culture answers the DIGITAL-2022-CULTURAL-02 call, under the topic *Data space for cultural heritage (deployment)*, and fits under the Digital Europe Programme. The objective of the call is the creation of the European common CH data space, which provides support to the digital transformation of Europe's cultural sector and fosters the creation and reuse of content in cultural and creative sectors. The company <u>CrossLang</u> decided to participate in a project proposal for this call in order to apply its yearlong expertise in automated translation to a sector with very challenging textual material.

The Al4Culture project started in April 2023 and will run until March 2025, for a total of 24 months. It involves 12 partners and is coordinated by the AILS Laboratory of the National Technical University of Athens (NTUA). The consortium showcases the following technologies on the platform: transcription and translation of scanned printed and handwritten documents (CrossLang and Austrian Institute of Technology), generation and validation of multilingual subtitles (Italian research institution Fondazione Bruno Kessler and company Translated), enrichment of metadata of CH objects with person names, colors and object types (NTUA, its spin-off Datoptron, and Belgian company Datable), and translation of the metadata of objects (Spanish company Pangeanic). Non-technical aspects of the project related to capacity building and dissemination are mainly taken care of by Europeana Foundation (Netherlands), which stewards the common CH data space, the European Fashion Heritage Association (Italy), the DigitGLAM unit at University of Leuven, and the Institute for Sound and Vision (Netherlands).



The platform will launch its first version in the second half of 2024, offering access to openly labelled datasets for training and testing AI models, to deployable and reusable tools, and to capacity building materials on the use of these tools and datasets for training, testing, and evaluation. The project partners are converting existing datasets (such as manual transcriptions of scanned documents) into the format required for training models and are creating new datasets semi-automatically. While the creation of datasets useful for AI applications holds great potential in terms of digital transformation, it can also prove highly challenging. To illustrate this, the performance of systems for automatic

transcription of scanned documents ranges from very high (in case of modern printed documents), to moderate or even very low (in case of handwritten documents).



The platform targets CH students and professionals, data providers, researchers and AI model developers, amongst others. The audience will also be able to contribute resources itself. Towards the end of 2024, several workshops will be organised. This will allow the technical partners in the project to get in contact with stakeholders from the CH sector and hence to explore potential new collaborations.



For more information, visit the project website or contact ai4culture@ails.ece.ntua.gr.

AI4Culture is funded by the European Commission's DIGITAL programme (project 101100683).

