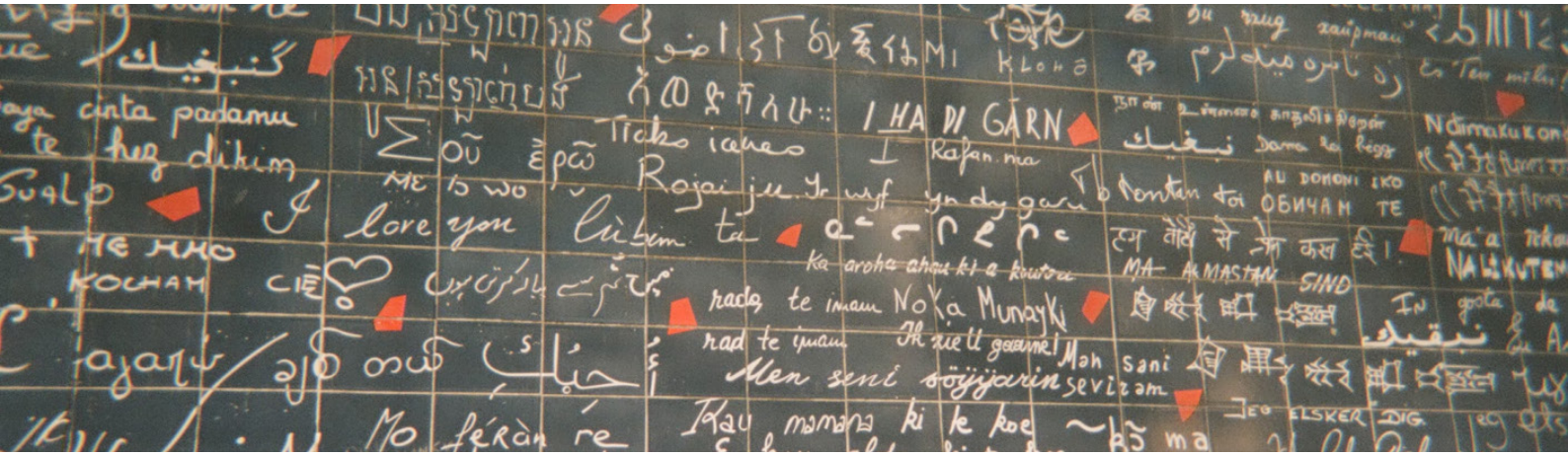
LORYA

MINISTÈRE
DE L'EUROPE
ET DES AFFAIRES
ÉTRANGÈRES
Liberté
Égalité
Fraternité

UNDP



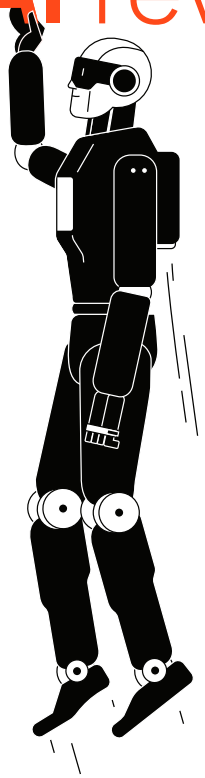
Including underrepresented languages in the AI revolution

Donor:
Government of France - Ministry for Europe and Foreign Affairs

Funding:
95,000 USD

Implemented by:
United Nations Development Programme (UNDP)

Duration:
February 2025 – February 2026



What is Lorya?

- It is a digital tool that turns written cultural heritage from print - into clean, machine-readable text that can be used to train AI language tools in local languages.
- It helps Serbian and other under-represented language speakers to leverage existing cultural and historical resources to take part in the global AI revolution.
- It makes information of cultural and historical significance, including old books, newspapers, and manuscripts - more accessible to researchers, historians, students and the public – for study, research and creating new products and services.

Our vision

To help include more languages in the AI technology revolution by developing a digital platform that local teams worldwide can easily and affordably adapt to ensure their own written cultural heritage is incorporated and accessible to all.

Why do we need it?

Today's AI language technology excels in English and only a few other major languages because most under-resourced languages lack sufficient clean, machine-readable text data to train AI systems effectively. As a result, large volumes of cultural and historical materials remain inaccessible, locked in images and scans. Traditional optical character recognition (OCR) tools rarely perform well on complex scripts, older typographies, or mixed orthographies, leaving many communities unable to digitize their heritage at scale. Recognizing the need for a solution that could be successfully used globally, UNDP is building on the solution initially designed for the Serbian language.

Developing Lorya - Key Activities:

1. **User-Centered Design and Localization**

   Lorya is designed for librarians, archivists, and cultural heritage professionals with limited technical expertise. Its clear interface, guided workflows, and localized options make it easy to digitize cultural materials both manually and with AI support. By streamlining editing and correction, Lorya reduces time and effort, making large-scale digitization more practical and affordable.

2. **Infrastructure, Deployment, and Adaptability**

   Lorya is built in a way that makes it easy to set up quickly, whether on local servers or the cloud. It works well with both commercial and free AI tools, so teams can easily adjust it to fit their own needs and keep improving it over time.

Why This Matters:

Digitizing cultural heritage takes a lot of time and money, especially in places with fewer resources. Lorya reduces these barriers by combining human expertise with AI tools that people can adjust themselves, lowering costs while improving accessibility. This empowers local institutions to preserve their written heritage and strengthen the presence of underrepresented languages in global AI systems.

Future use of Lorya

Lorya will be released as a digital public good, allowing partners to adopt it freely, and help contribute improvements. Designed with an editor at the center, Lorya will allow local teams to manually correct results, apply the most advanced open models, and train those models over time, so quality improves, and costs decrease. This will result in reliable text and reusable data, ready for AI technologies of today and the future.