

# Language Data for European AI: Striking on Two Fronts

**Christos Ellinides**

**Director-General**

Directorate-General for Translation (DGT)  
and AI-based multilingual services

*8 May 2025*

*University of Cyprus*

# Lay of the land

MULTILINGUALISM  
IS MY  
SUPERPOWER

*Euroscepticism*



# Striking on Two Fronts

1. Modern and technologically revamped Public Administrations

2. EU digital and technical sovereignty

# Introduction to DGT

**Over 70-year history**

**DGT translates into and out of 24 official EU languages**

**Current output: ±2.5 million pages per year**



***People are at the heart of translation***

# Linguistic data assets: (Euramis and DGT-TM)

## ○ Euramis data repository

Launched in 1995

## ○ 2 billion sentences

> 300 000 a day

## ○ DGT-TM

Launched in 2007

→ used by research, business,  
freelance translators  
and more



# Expanding into language technologies

- **Machine translation**

MT@EC -> eTranslation (*from a statistical model to neural technology and GenAI*)

- **ELRC Data Collection (2016-2020)**

better support for low-resource languages

- **Broadening reach**

expansion into non-translation applications: AI-based services





## Digital Europe AI-Based Multilingual Services

### AI-Based Multilingual Services

These services are made available in the EU under the Digital Europe Programme for use by EU institutions, public administrations, academia, SMEs, NGOs, Digital Europe Programme projects and EPSO candidates.

They offer both web pages and APIs for machine-to-machine access.



#### eTranslation

Neural machine translation built on the EU's history of professional translation.



#### WebText

Use AI to refine English, French, or German text for your website.



#### Multilingual Post

Short translations in multiple languages in one shot for X.



#### eBriefing

Generate reports from sets of documents in official or general styles.



#### Accessible Text

Rewrite your English text to make it more accessible.



#### Speech-to-Text

Upload your media and get full transcriptions or subtitles back.



#### eReply

Have AI help you prepare replies to correspondence, queries, and other requests.



#### eSummary

Quickly find out the main content of long documents.



#### Natural Language Processing Tools (NLP)

Anonymisation, Classification and Named-Entity Recognition.

### Access a

Access to the  
EU staff is pre-re

- Who can
- Registrat
- API regis

Contact point:  
[DGT-AI-Lang](#)

### Additional useful resources



#### Video guides and help

Simple explanations of our Services



#### Developer's corner

Guidelines and procedures on how to integrate



#### Web translation tool

Free plugin for your website



#### Language Data Space

European platform for exchanging languages resources

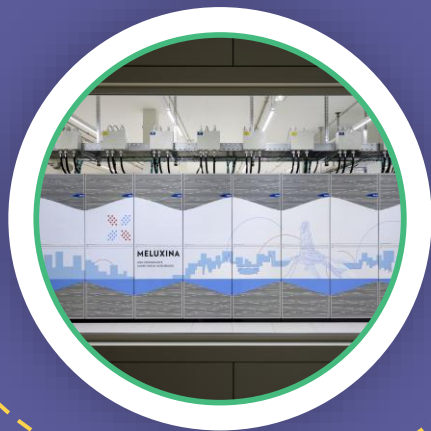


#### IATE

Interactive Terminology for Europe



# Roadmap for the future



## European HPC Joint Undertaking

→ 'MeluXina' and 'Leonardo' supercomputers

### **EU institutional LLM**

- ✓ made in Europe
- ✓ open source
- ✓ institutional, curated data

## European Language Data Space (ELDS)

- European marketplace for language data
- supporting Europe's digital sovereignty







# European Language Data Space

The European Marketplace for Language Data.

Discover the LDS Beta >

© European Union

## Unleash the potential of data – for EU businesses and citizens.

*The European Language Data Space aims at building a trustworthy and effective data market for the exchange of language resources in the public and – even more importantly – in the private sector, in line with the [EU Data Strategy](#).*



### Share

You will be able to share and monetise your language data, language models and other language resources through a single platform, taking European values and compliance with EU regulations fully into account.



### Connect

You will be able to connect and exchange with other stakeholders through the European Language Data Space.



### Benefit

You will be supported by the European Language Data Space in the development of multilingual and multimodal language technologies and language-centric AI.



# The political context

## ● EU AI Act

## ● Political Guidelines of the European Commission 2024-2029

→ Bolster supercomputing capacity

→ Support industrial uses for AI

→ Step up deployment of AI within the EU

## ● AI Action summit (Paris, February 2025)

AI Action summit | Paris, February 2025

*'Action is exactly what we need right now. The time has come for us to formulate a vision of where we want AI to take us, as society and as humanity. And then we need to act and accelerate Europe in getting there.'*



**Ursula von der Leyen**  
President of the European  
Commission

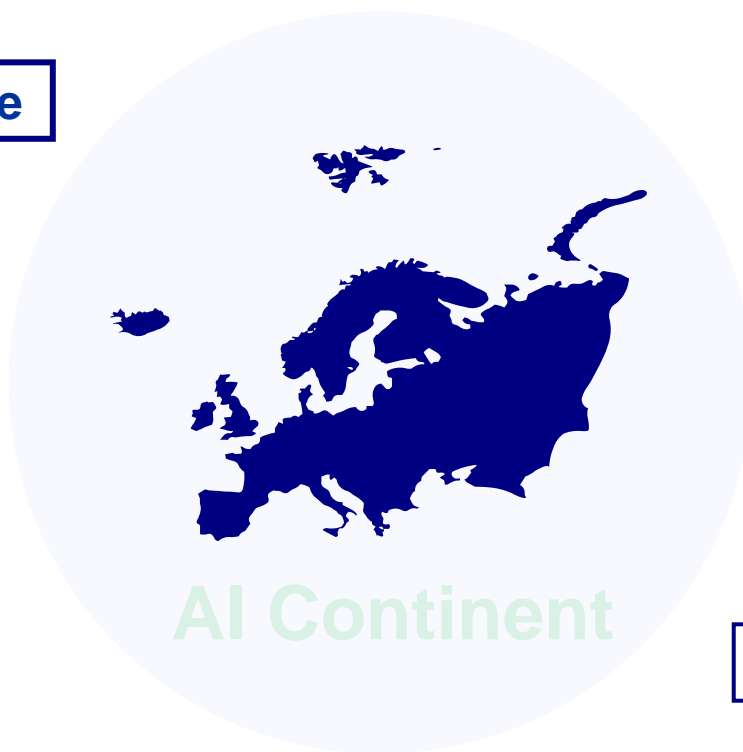
# Towards the creation of an AI continent



Computing infrastructure



Skills



Data



Regulatory simplification



Development of algorithms and adoption



# From Translations to AI-based services

## DGT's Role in AI & Language Data

Machine  
translation  
and AI  
based  
multilingual  
services



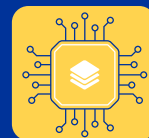
European AI  
development

DGT's  
translations  
throughout the  
years



### 25+ Years of Data

→ Translation archives



### AI-based services and Machine Translation

→ Evolution of data usage



### Multilingual Dataset

→ Includes low-resource languages



### EU Collaboration on AI

→ European sovereignty & wealth of languages

# Thank you

© European Union 2025

Unless otherwise noted the reuse of this presentation is authorised under the [CC BY 4.0](#) license. For any use or reproduction of elements that are not owned by the EU, permission may need to be sought directly from the respective right holders.

Slide 05: photo of computer, source: AdobeStock.com

