European Language Resource Coordination
Connecting Europe Facility

European Commission

LATEST TRENDS IN ARTIFICIAL INTELLIGENCE

Eileen Marra, ELRC Communications Manager (DFKI)

11th LRB Meeting (26 November 2021)

European Language Resource Coordination
Connecting Europe Facility

European Commission

"AI – Here for good"[1] … but where is Language Technology?

[1] title of the Irish AI Strategy

European Language Resource Coordination
Connecting Europe Facility

European Commission

WHAT WE'LL BE TALKING ABOUT TODAY

- Analysis approach
- National AI Strategies across Europe – Key Findings
- The role of LT in national AI Strategies
- The value of language data
- Conclusions

European Language Resource Coordination
Connecting Europe Facility

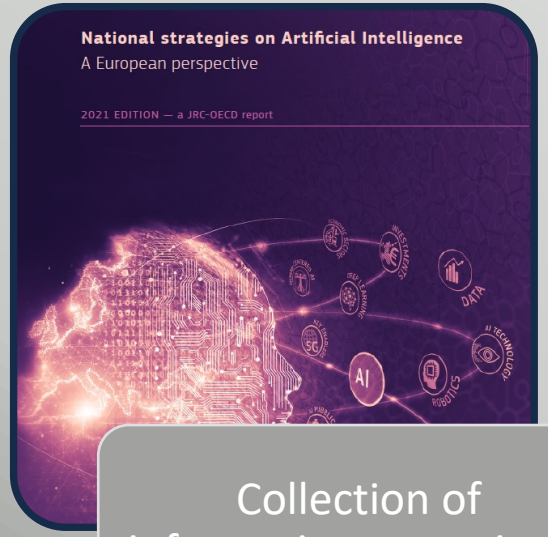
European Commission

Analysis Approach

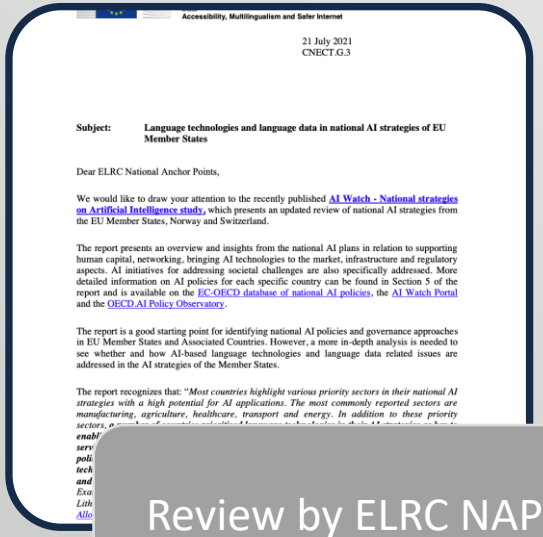
European Language Resource Coordination
Connecting Europe Facility

European Commission

Analysis Approach



Collection of information on national AI Strategies based on AI Watch Report


Accessibility, Multilingualism and Safer Internet

21 July 2021
CNECT.G.3

Subject: **Language technologies and language data in national AI strategies of EU Member States**

Dear ELRC National Anchor Points,

We would like to draw your attention to the recently published **AI Watch - National strategies on Artificial Intelligence study,** which presents an updated review of national AI strategies from the EU Member States, Norway and Switzerland.

The report presents an overview and insights from the national AI plans in relation to supporting human capital, networking, bringing AI technologies to the market, infrastructure and regulatory aspects. AI initiatives for addressing societal challenges are also specifically addressed. More detailed information on AI policies for each specific country can be found in Section 5 of the report and is available on the EC-OECD database of national AI policies, the AI Watch Portal and the OECD AI Policy Observatory.

The report is a good starting point for identifying national AI policies and governance approaches in EU Member States and Associated Countries. However, a more in-depth analysis is needed to see whether and how AI-based language technologies and language data related issues are addressed in the AI strategies of the Member States.

The report recognizes that: "*Most countries highlight various priority sectors in their national AI strategies with a high potential for AI applications. The most commonly reported sectors are manufacturing, agriculture, healthcare, transport and energy. In addition to these priority sectors,*

Review by ELRC NAPs over summer

THANK YOU!

Detailed look at national AI Reports, also searching for key words related to LT and LR

European Language Resource Coordination
Connecting Europe Facility

European Commission

Analysis Approach

Key Topics

- AI regulation/legislative framework for AI
- Societal challenges addressed by AI Strategy
- Major AI Networks and Collaborations
- AI-related LT projects and initiatives
- Available AI Funding for LT
- Major LT players in AI
- LT Policies
- Data Collection Efforts/Repositories

Goal

→ Find out how Language Resources and Language Technology are represented in national AI Strategies

→ Identify missing activities to boost the development of language-centric AI in Europe

European Language Resource Coordination
Connecting Europe Facility

European Commission

National AI Strategies – Key Findings

"AI is not a technology of the future,
it is a technology of the present."

(quote from AI Strategy Ireland, p.2)

National AI Strategies – Key Findings

- Increasing importance of AI all across the CEF-affiliated countries:
  - AI Strategies released for 23 of the 29 countries
  - Remaining Strategies are work in progress (publication planned between 2021 and 2022) → Belgium, Croatia, Greece, Iceland, Italy, Romania
  - Updated version published in 3 countries → Cyprus, Finland and Germany
  - Interesting side fact: 17 of the 23 national AI Strategies available in English

Major Networks and Collaborations

- National AI Competence centres, e.g. Finnish Centre for Artificial Intelligence (FCAI), French interdisciplinary institutes of AI (3IA), Hungarian AI National Laboratory (MILAB), etc.
- Mapping of AI actors and applications to increase innovation community building in e.g. Belgium, Czech Republic, Finland, Germany or Poland
- Digital Innovation Hubs
- International Networks and Partnerships like DARIAH, CLARIN or the Global Partnership on AI (GPAI).

European Language Resource Coordination
Connecting Europe Facility

European Commission

THE ROLE OF LT IN NATIONAL AI STRATEGIES

"In the future, citizens can receive services seamlessly in the language they need (…)."

(quote from Finland's age of AI, p.54)

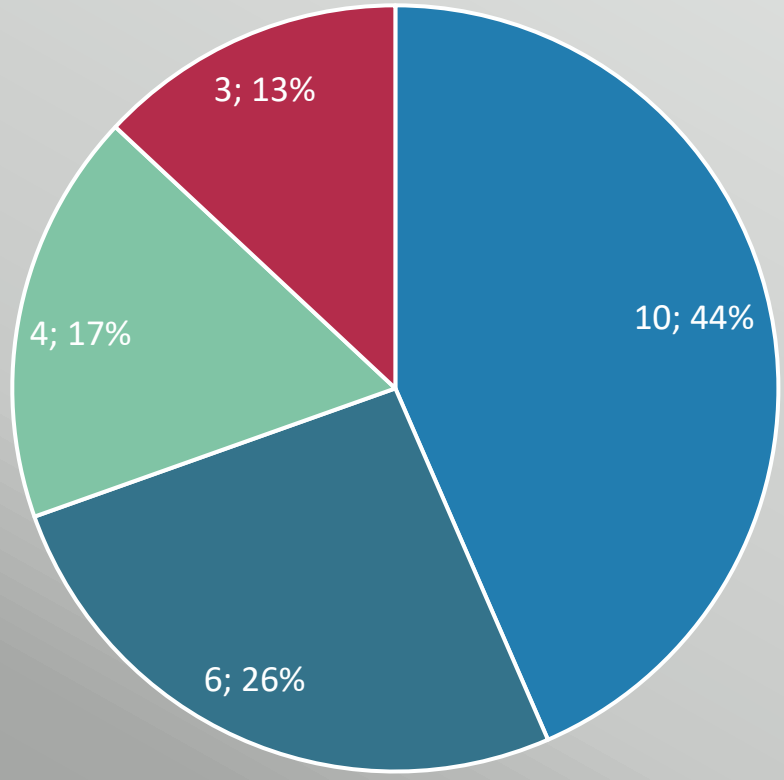
The role of LT in National AI Strategies

- 20 of the published 23 national AI Strategies mention Language Technology

→ Varying emphasis: Full chapters on LT (e.g. Malta) vs. side notes on language-centric AI (e.g. Luxembourg)

- Countries where LT is not explicitly mentioned: Sweden, Estonia (but: included LT in draft strategy of Estonian language), Netherlands (but: mentions chatbots as a useful NLP application)
- Examples:
  - Bulgaria: Use of LT to support foreign language learning
    "In practice, any formalized set of grammar rules can be considered as a resource for automatic testing of knowledge of the relevant aspects of the language, which is built into specially designed tests for verification. It would be useful for Bulgarians abroad to provide a public online interface for learning Bulgarian grammar. " (Concept for the development of AI in Bulgaria until 2030, p. 44)
  - Hungary: *"The application and further development of existing technologies to the Hungarian language is of significant national interest." (Hungary's Artificial Intelligence Strategy, p. 26)*

The role of LT in National Strategies: Emphasis on LT



European Language Resource Coordination
Connecting Europe Facility

European Commission

The role of LT in National AI Strategies: LT Funding

Interesting facts – a selection

Austria's digital roadmap does not foresee specific funding to LT. The same applies to e.g. Cyprus, Ireland or Portugal

In Iceland, the National LT Programme is funded by the Government through the non-profit organisation Almannarómur

In Denmark, 2.6 million € have been allocated to the initiative sprokteknologi.dk between 2019 and 2026 to support the development of Danish LT.

In Romania, most of the AI funding for LT comes from EC.

In Estonia, the LT Programme foresees 1.2 € per year for developing basic resources for LT and practical applications in institutions engaged in research and development activities.

In Germany, the Federation will be able to make available approx. €3000 million for implementing the AI strategy in the period from 2018 to 2025. This includes LT, but no specific share is indicated.

In Sweden, the national research council has annual fundings of LT projects of around 5-10 million crowns/year (± 0.5 to 1 million €)

European Language Resource Coordination
Connecting Europe Facility

European Commission

THE VALUE OF LANGUAGE DATA

"There is reason to believe that the public sector possesses far more data that could be used in developing language technology than it realises."

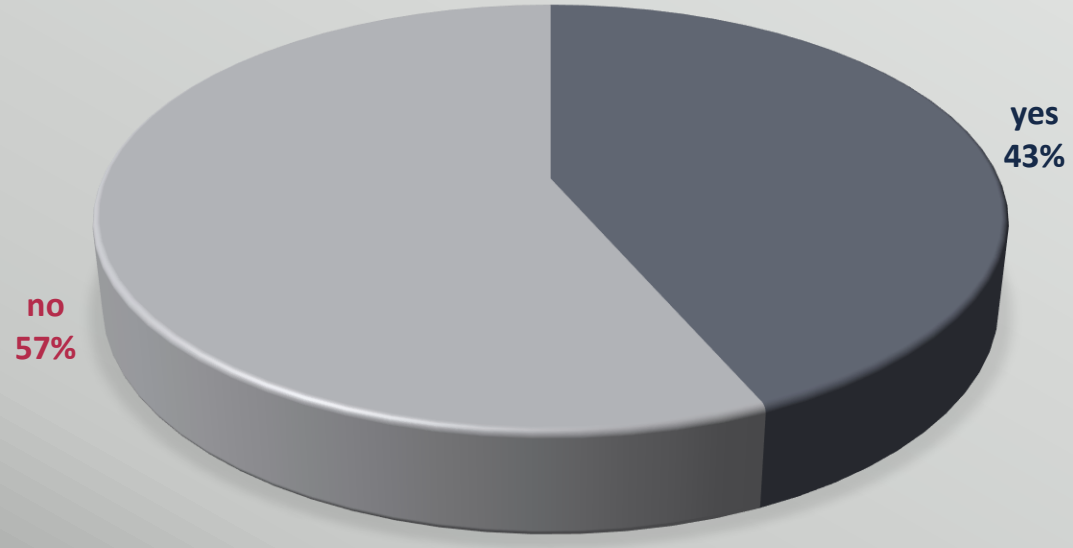
(quote from Norwegian AI Strategy, p.20)

European Language Resource Coordination
Connecting Europe Facility

European Commission

The Role of Language Data

Is the National AI Strategy referring to the importance of language data?



European Language Resource Coordination
Connecting Europe Facility

European Commission

The Value of Language Data: Good Practice Examples

Ireland

- "Many of the language datasets currently used for training AI systems originate from US-based sources and may not contain common everyday terms used by people in Ireland. To render AI systems accessible to a wider range of our population, as well as to develop services in Irish based on AI for Irish language-speakers, good language technology resources need to be developed." (AI Strategy "AI – Here for Good", p.42)

Norway

- Special chapter about LR & LT:
  "There is reason to believe that the public sector possesses far more data that could be used in developing language technology than it realises. The Government will therefore promote awareness of language data and language resources in the public sector" (National Strategy for Artificial Intelligence, p. 20)

Spain

- One of the action items:

"Boosting the National Language Technology Plan and the creation of resources in the Spanish Language in AI initiative" (ES: Impulso al Plan Nacional de Tecnologías del Lenguaje y la creación de recursos en la iniciativa de Lengua Española en la IA")

European Language Resource Coordination
Connecting Europe Facility

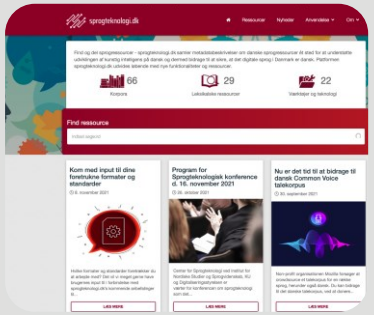
European Commission

National Data collection Efforts: Examples

National Open Data Portals

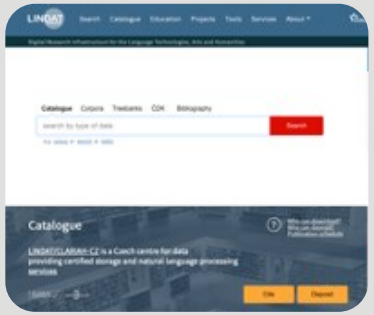
(often limited number of language data)

Across Europe



The portal sprogteknologi.dk stores metadata about existing LR

Denmark



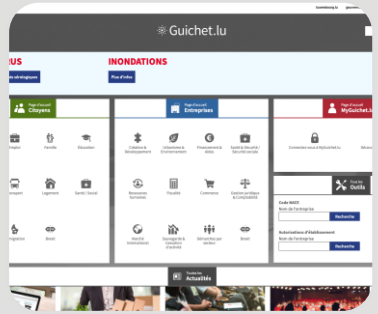
In Czech Republic, the largest repository for open language resources is the LINDAT/CLARIAH-CZ infrastructure (https://lindat.cz)

Czech Republic

In Germany, the Data Usage Act was introduced to facilitate data accessibility

LEAM.AI fosters the development of large AI models

Germany



Guichet.lu exchanges glossaries with the Information and Press Service (SIP) and the national data portal.
Plus: the exchange of TMs between ministries and Guichet.lu is to be extended!

Luxembourg

European Language Resource Coordination
Connecting Europe Facility

European Commission

CONCLUSIONS

CONCLUSIONS

Goal 1: Find out how Language Resources and Language Technology are represented in national AI Strategies

- Importance and potential of AI: Recognised all across Europe, but in most cases LT could/should be more prominent
- Major differences between the national strategies: Not mentioning LT at all vs. dedicating whole chapters/key pillars to the development of LT → LT should become a priority in all countries
- Value of language data has not been mentioned explicitly in most of the national AI Strategies → However, many countries are starting initiatives to build big data models in LT (e.g. the Spanish project LEIA or LEAM.AI in Germany)
- Many national AI Strategies highlight the need for cross-border communication to build on recent achievements
  → Maximum success through cooperation

European Language Resource Coordination
Connecting Europe Facility

European Commission

CONCLUSIONS:

Goal 2: Identify missing activities to boost the development of language-centric AI in Europe - Examples:

- Lack of employees with the right skills: At European level, there is an estimated demand of 600,000 specialists within IT programming, and this is expected to increase in the years to come.[1] → Reinforcement of human skills in AI at all educational levels (training and lifelong learning)
- Need to encourage partnerships with leading international organisations to increase the level of research and innovation in AI [2]
- Further information campaigns and awareness-raising acitivities to promote LT and LR!

… additional activities may be identified in the course of today!

[1]Source: Danish AI Strategy
[2] Source: Cypriot AI Strategy

European Language Resource Coordination
Connecting Europe Facility

European Commission

LET'S KEEP ON JOINING FORCES TO DRIVE THE CHANGE!

THANK YOU FOR YOUR ATTENTION