Chapter 44
Strategic Plans and Projects in Language Technology and Artificial Intelligence

Itziar Aldabe, Aritz Farwell, German Rigau, Georg Rehm, and Andy Way

Abstract This chapter on existing strategic plans and projects in Language Technology and Artificial Intelligence is based on an analysis of around 200 documents and is divided into three sections. The first provides a synopsis of international and European reports on Language Technology. The second constitutes a review of existing European Strategic Research Agendas, initiatives, and national plans related to Language Technology. The third contains a SWOT analysis designed to identify the factors that will need to be addressed to help solve the challenge of digital language inequality in Europe. Among the principal conclusions presented is the contention that our continent requires sophisticated multilingual, cross-lingual and monolingual LT for all European languages: LT *for* Europe that is *made in* Europe.[1]

1 Introduction

In varietate concordia (united in diversity) is the official Latin motto of the European Union (EU), adopted in 2000. According to the European Commission, "the motto means that, via the EU, Europeans are united in working together for peace and prosperity, and that the many different cultures, traditions and *languages in Europe* are a positive asset for the continent" [emphasis added].[2] All 24 official EU languages are granted equal status by the EU Charter and the Treaty on the EU. The EU is also home to over 60 regional and minority languages which are protected and promoted under the European Charter for Regional or Minority Languages (ECRML)

Itziar Aldabe · Aritz Farwell · German Rigau
University of the Basque Country, Spain, itziar.aldabe@ehu.eus, aritz.farwell@ehu.eus, german.rigau@ehu.eus

Georg Rehm
Deutsches Forschungszentrum für Künstliche Intelligenz GmbH, Germany, georg.rehm@dfki.de

Andy Way
Dublin City University, ADAPT Centre, Ireland, andy.way@adaptcentre.ie

[1] This chapter is an abridged version of Aldabe et al. (2022).

[2] http://europa.eu/abc/symbols/motto/index_en.htm

since 1992,[3] in addition to migrant languages and various sign languages, spoken by some 50 million people. The Charter of Fundamental Rights of the EU under Article 21[4] states that "any discrimination based on any ground such as sex, race, colour, ethnic or social origin, genetic features, *language*, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation shall be prohibited" [emphasis added].

Multilingualism is a cultural cornerstone of Europe and signifies part of what it means to be and to feel European. However, not only do language barriers still hamper cross-lingual communication and the free flow of knowledge and thought across languages, a dilemma for which no common EU policy has been proposed, many languages themselves are also endangered or on the edge of extinction (even more so from a digital perspective). This is illustrated in the *UNESCO Atlas of the World's Languages in Danger* (Moseley 2010),[5] where a map of Europe shows threatened languages, including black flags that correspond to already extinct languages.

Without a concerted effort to prevent the further deterioration of Europe's linguistic ecosystem, this current snapshot is likely to worsen. And while it may well be that no silver bullet exists to remedy the situation, one approach offers a means to provide immediate support and address the issue of linguistic barriers: Language Technology (LT) and language-centric Artificial Intelligence (AI).

Because natural language is at the heart of human intelligence, it is and must be at the heart of our efforts to develop AI technologies.[6] By the same token, all sophisticated and effective AI-powered tools are impossible without mastery of language.[7] This is why language and LT represent the next great frontier in AI.[8] Already arguably the hottest field in AI, LT also represents one of its fastest growing application areas.[9] In fact, together with vision and robotics, several recent international reports place LT as one of the three core application areas within AI. Its rise to prominence is due to the various methods LT has developed over the years to make the information contained in written and spoken language explicit or to generate written and spoken language. For this reason, it has become the nerve centre of the software that processes unstructured information and exploits the vast amount of data contained in text, audio and video files, including those from the web and social media. Despite the inherent difficulties in many of the tasks performed, current LT support allows for many advanced applications which would have been unthinkable only a few years ago. Among these may be counted speech recognition, speech synthesis, text analytics and machine translation (MT), used by hundreds of millions of people on a daily basis.[10] It is now common to utilise search engines, recommender

---

[3] https://en.m.wikipedia.org/wiki/European_Charter_for_Regional_or_Minority_Languages

[4] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT

[5] http://www.unesco.org/languages-atlas/

[6] https://hbr.org/2022/04/the-power-of-natural-language-processing

[7] https://www.nytimes.com/2022/04/15/magazine/ai-language.html

[8] https://www.forbes.com/sites/robtoews/2022/02/13/language-is-the-next-great-frontier-in-ai

[9] https://analyticsindiamag.com/is-nlp-innovating-faster-than-other-domains-of-ai/

[10] https://www.nimdzi.com/nimdzi-language-technology-atlas-2020/
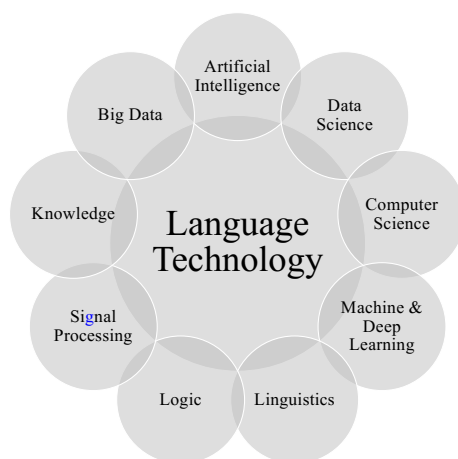
Fig. 1 Language Technology as a multidisciplinary field

systems, virtual assistants, chatbots, text editors, text predictors, MT systems, automatic subtitling, automatic summaries, and inclusive technology, all made possible thanks to LT. The field's rapid development promises even more encouraging results in the near future and its increasing social relevance has been highlighted in national and regional AI and LT strategies both inside and outside of Europe, as well as in prioritised strategic areas for research, development and innovation (R&D&I).

With this in mind, it should not be forgotten that LT is also multidisciplinary in nature, combining knowledge in computer science (and specifically in AI), mathematics, linguistics and psychology, among others. Figure 1 depicts some of the most important disciplines involved in LT. This uniqueness must be weighed in any public or private AI initiative that includes LT, especially given that funding for LT start-ups is booming and only the proper application of LT will allow the enormous volumes of multilingual written and spoken data in sectors as diverse as health, justice, education, or finance to be adequately processed and understood.[11]

Early-stage funding in 2021 amounts to just over USD 1 billion for companies offering solutions that make significant use of NLP, providing a picture of what funders think is innovative.[12] This belief is only reinforced by technology advances such as ChatGPT, whose creator, OpenAI, projects USD 1 billion in revenue by 2024.[13] Similarly, reports from analysts and consulting firms forecast enormous growth in the global LT market based on the explosion of applications observed in recent years and the expected exponential growth in unstructured digital data. For instance, ac-

---

[11] https://www.forbes.com/sites/robtoews/2022/03/27/a-wave-of-billion-dollar-language-ai-star tups-is-coming

[12] https://towardsdatascience.com/nlp-how-to-spend-a-billion-dollars-e0dcdf82ea9f

[13] https://www.reuters.com/business/chatgpt-owner-openai-projects-1-billion-revenue-by-2024 -sources-2022-12-15/

cording to an industry report from 2019,[14] the global NLP market size is expected to grow from USD 10.2 billion in 2019 to USD 26.4 billion by 2024, at a CAGR of 21% is set during the forecast period 2019-2024.[15] A recent report from 2021 estimates that the global NLP market is predicted to grow from USD 20.98 billion in 2021 to USD 127.26 billion in 2028 at a CAGR of 29.4% in the forecasted period.[16] NLP in Europe will witness market growth of 19.7% CAGR and is expected to reach USD 35.1 billion by 2026.[17] As a final example, according to Global Newswire the global NLP market is estimated to reach an expected value of USD 341.7 billion by 2030, growing at a CAGR of 27.6% during the forecast period.[18] These numbers indicate that the return on investment (ROI) will be massive so it is imperative that Europe is at the heart of this growth in future.

The attention paid to AI and LT in the social, political, and economic spheres reflect the significance of the technology for today's world. This chapter on the existing strategic plans and projects in LT and AI touches on all three of these areas. It is based on an analysis of close to 200 documents (Aldabe et al. 2022) and is divided into three sections. Section 2 provides a synopsis of international and European reports on LT. In addition to trends in innovation, many of these discuss the socioeconomic and political impact of AI and LT from a policy perspective. Section 3 constitutes a review of the existing European Strategic Research Agendas (SRAs), initiatives, and national plans related to LT. A main focus of these is the question of multilingualism and equal technological support for Europe's languages through the application of LT. Section 4 contains a SWOT analysis of the strategic documents and projects, which is designed to identify the factors that will need to be addressed to help solve the pressing issue of digital language inequality in Europe.

2 International Reports on Language Technology

AI capabilities are rapidly evolving and it has become one of the 21st century's most transformative technologies.[19] The growing interest in AI at a global political, scientific and social level has led several international organisations to draft a number of reports and initiatives in recent years. These often focus on the socioeconomic impact of AI technologies and applications with respect to policy.

---

[14] https://www.businesswire.com/news/home/20191230005197/en/Global-Natural-Language-Processing-NLP-Market-Size

[15] https://www.analyticsinsight.net/potentials-of-nlp-techniques-industry-implementation-and-global-market-outline/

[16] https://www.analyticsinsight.net/the-global-nlp-market-is-predicted-to-reach-us127-26-billion-by-2028/

[17] https://www.analyticsinsight.net/nlp-in-europe-is-expected-to-reach-us35-1-billion-by-2026/

[18] https://www.globenewswire.com/en/news-release/2022/09/29/2525379/0/en/Natural-Language-Processing-NLP-Market-Worth-USD-341-7-Billion-with-a-27-6-CAGR-by-2030-Report-by-Market-Research-Future-MRFR.html

[19] https://www.holoniq.com/notes/50-national-ai-strategies-the-2020-ai-strategy-landscape/

2.1 Reports from International Organisations

The Organisation for Economic Co-operation and Development (OECD),[20] a frequent contributor to this discourse, has helped coordinate dialogue on the subject at international fora (notably the G7, G20, EU and UN), offered practical advice to governments on how to actualise AI policy, and stressed the potential that digital technologies demonstrate in responding to societal challenges.[21] Its 2021 report, *State of the implementation of the OECD AI Principles: Insights from national AI policies*, identifies challenges and best practices for the implementation of the five policy recommendations to national governments contained in its OECD AI Principles. These are: 1. invest in AI R&D; 2. foster a digital ecosystem for AI; 3. shape an enabling policy environment for AI; 4. build human capacity and preparation for labour market transformation; and 5. foment international co-operation for trustworthy AI. The report comes on the heels of the OECD's *The Digitalisation of Science, Technology and Innovation*, which emphasises that cutting-edge NLP techniques are opening new analytical possibilities. Among those listed is the ability to recognise victims of sexual exploitation on the internet based on facial detection and social network analysis (Chui et al. 2018). Advances such as this have caught the attention of researchers and policy makers in various countries, who have begun to experiment with NLP to track emerging research topics and technologies. As the report underscores, policy makers use these results to formulate science and innovation policy initiatives, support investments in R&D&I, and evaluate public programmes.[22]

Similar policy guidance and assessments appear elsewhere.[23] The Inter-American Development Bank[24] (IDB), for instance, suggests constructing a shared understanding of AI in order to take better advantage of its opportunities and applications while simultaneously coming to grips with its risks.[25] The World Economic Forum,[26] which provides a framework for governments that wish to develop national AI strategies, assists those responsible for crafting policy in how to ask pertinent questions, follow best practices, identify and involve stakeholders, and create a set of outcome

---

[20] https://www.oecd.org

[21] See, e. g., *Artificial Intelligence in Society* (https://doi.org/10.1787/eedfee77-en); *State of the implementation of the OECD AI Principles: Insights from national AI policies* (https://doi.org/10.1787/1cd40c44-e); *The Digitalisation of Science, Technology and Innovation* (https://doi.org/10.1787/b9e4a2c0-en).

[22] To help policy makers, regulators, legislators and others characterise AI systems deployed in specific contexts, the OECD has developed a user-friendly tool to evaluate AI and LT systems from a policy perspective (https://www.oecd.org/publications/oecd-framework-for-the-classification-of-ai-systems-cb6d9eca-en.htm).

[23] See, e. g., the *Helsinki Initiative on Multilingualism in Scholarly Communication* (https://www.helsinki-initiative.org/en).

[24] https://www.iadb.org

[25] https://publications.iadb.org/en/artificial-intelligence-for-social-good-in-latin-america-and-the-caribbean-the-regional-landscape-and-12-country-snapshots

[26] https://www.weforum.org

indicators.[27] UNESCO[28] extends these considerations to the educational sphere, recommending that governments and other stakeholders, in accordance with their legislation and public policies, respond to education-related opportunities and challenges presented by AI. The *Beijing Consensus on Artificial Intelligence and Education*, an outcome document issued by UNESCO in 2019, stresses the multidisciplinary nature of AI and urges readers to consider the role of AI tools in teaching and learning, highlighting its effectiveness in aiding students with learning impairments or who study in a language other than their mother tongue.[29] In the area of library science, *Responsible Operations: Data Science, Machine Learning, and AI in Libraries*, a position paper from OCLC,[30] notes structural inequalities are perpetuated by data-driven policies (Padilla 2020) and sets an agenda for tackling positive and negative impacts of data science, machine learning, and AI on libraries.[31]

Finally, in early 2022, based on the report *Facilitating the implementation of the European Charter for Regional or Minority Languages through artificial intelligence*, first published in 2020[32] and updated in 2022,[33] the Committee of Experts of the European Charter for Regional or Minority Languages of the Council of Europe (CoE) adopted a statement on the promotion of regional or minority languages through AI.[34] The Committee of Experts encourages states to promote the inclusion of regional or minority languages into research and study on AI with a view to supporting the development of relevant applications as well as to establish, in cooperation with the users of such languages and the private sector, a structured approach to the use of AI applications in the different fields covered by the Charter.

The attention paid to AI and LT in policy reports reflects the social, political, and economic importance that the technology has garnered in today's world; and the same holds true for organisations that trace trends in innovation. In its report, *Technology Trends 2019 Artificial Intelligence*,[35] the World Intellectual Property Organization[36] found that 50% of all AI patents have been published in just the past five years, a striking illustration of how rapidly innovation is advancing in the field. The report, which classifies AI technology trends into techniques, functional applications, and application fields, furthermore points to LT as one of AI's most significant functional applications, attributing over a quarter of all AI-related patents to NLP and speech processing. The number is unsurprising given the current levels of excite-

---

[27] https://www3.weforum.org/docs/WEF_National_AI_Strategy.pdf

[28] https://en.unesco.org

[29] https://unesdoc.unesco.org/ark:/48223/pf0000368303 See also, UNESCO's *Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development*, a 2019 report which, among other breakthroughs, noted a Chinese AI system that is able to correct student essays as a milestone in LT for education (https://unesdoc.unesco.org/ark:/48223/pf0000366994).

[30] https://www.oclc.org/en/about.html

[31] https://doi.org/10.25333/xk7z-9g97

[32] https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da

[33] https://rm.coe.int/min-lang-2022-4-ai-and-ecrml-en/1680a657c5

[34] https://rm.coe.int/declaration-ai-en/1680a657ff

[35] https://www.wipo.int/publications/en/details.jsp?id=4386

[36] https://www.wipo.int

ment associated with NLP within AI, where the rising star is turning many heads. A case in point is the *State of AI Report* for 2021,[37] issued by UK AI investors with an eye toward stimulating informed conversation on AI and its implications going forward. The report, which considers research, talent, industry, and politics, discusses the emergence of large language models and notes that the latest generation are unlocking new NLP use-cases. Indeed, the arrival of Transformers as a general purpose architecture for ML has been a revelation, beating the state-of-the-art in domains as disparate as computer vision and protein structure prediction.

2.2 Reports from the United States

Reports from the US tell an analogous story to their international counterparts. In its 2021 and 2022 *AI Index Reports*,[38] for example, the Institute for Human-Centered AI (HAI) at Stanford University reviews the growth of research papers and conferences over time and by region, tracks AI accuracy on several benchmarks, focuses on trends in jobs and investment, and examines various national AI strategies. The reports also devote space to data and analysis concerning AI with respect to education, diversity, and ethics. Key takeaways include the observation that 65% of the new PhDs in the US chose jobs in industry over academia compared to 44% the previous year, that there is still little data available on the ethical challenges surrounding AI, and that the AI workforce remains predominantly male. The 2022 report also highlights that while current language models are setting records on technical benchmarks, they are also increasingly reflecting biases from their training data. These findings are accompanied by HAI's Global Vibrancy Tool,[39] which measures performance on various economic, inclusion, and R&D factors across several countries. The tool can create an overall index for the full list of 26 countries and it is of note that none of the top ten is an EU member state. The worrisome nature of the latter data point is compounded in an examination of the global balance and flow of top AI scientists provided by the Paulson Institute's Macro Polo think tank in its Global AI Talent Tracker report.[40] According to this analysis, the US lead in AI is built on attracting international talent, with more than two-thirds of the top-tier AI researchers working in the US having received undergraduate degrees in other countries. Although 18% of the top-tier AI researchers are European, only 10% of them work in Europe.

These final details should sound alarm bells in Europe. As demonstrated in *Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence*, released by the AI100 project in 2021, remarkable progress has been made in AI over the past five years and we may anticipate that its effects will ripple out

---

[37] https://www.stateof.ai

[38] https://aiindex.stanford.edu/report/

[39] https://aiindex.stanford.edu/vibrancy/

[40] https://macropolo.org/digital-projects/the-global-ai-talent-tracker/

for many years to come. Prepared by a panel of experts from around the globe, the report makes clear that the ability of computers to perform sophisticated language- and image-processing tasks has advanced significantly and that more investment of time and resources are required to meet the challenges posed by AI's rapidly evolving technologies. On the one hand, this includes greater government involvement in the areas of regulation and digital education. In an AI-enabled world, citizens young and old must be literate in these new digital technologies. On the other, this means addressing fears that AI technologies will contribute to unemployment in some sectors. A Blumberg Capital survey of 1,000 American adults found that about half are concerned that AI threatens their livelihood. Indeed, despite the fact that 72% agreed that AI would help remove tedious tasks and free up time to concentrate on more creative ones, 81% were reluctant to surrender these tasks to an algorithm for fear of being supplanted.[41] As the authors of *Gathering Strength, Gathering Storms* indicate, AI is leaving the laboratory and entering our lives, having a "real-world impact on people, institutions, and culture."[42]

This perspective is shared by the National Security Commission. In addition to raising concerns that the United States risks falling behind China and other countries in the AI race, its recent 750-page report encourages the federal government to step up investment in the area.[43] Specifically, the commission calls for a *modest down payment* of $40 billion, along with hundreds of billions more in the coming years to galvanise future breakthroughs and help democratise AI research. Moreover, the report provides policy makers with a guide to ensure the US is prepared to defend against AI threats, promote AI innovation, and make responsible use of AI for national security. It is also worth mentioning that the report lists Natural Language Understanding as one of the six uses for deployed AI today. This view, which coincides with the general consensus on LT expressed above, is further reinforced by the Future Today Institute[44] in its 2021 *Tech Trends Report* on AI.[45] The group not only identifies NLP as an area that is experiencing high interest, investment, and growth, but also forecasts that NLP algorithms will do more in the future, including, for example, aid in interpreting genetic changes in viruses.

2.3 Reports from the European Union

Reports from the EU paint an equally upbeat picture about present and future expectations regarding science and technology. A recent Eurobarometer survey on European citizens' knowledge and attitudes towards these shows that 86% believe the overall

---

[41] https://blumbergcapital.com/ai-in-2019/

[42] https://ai100.stanford.edu

[43] https://www.nscai.gov/2021-final-report

[44] https://futuretodayinstitute.com

[45] https://2021techtrends.com/AI-Trends

influence of science and technology is positive.[46] EU citizens expect a range of technologies currently under development, including AI (61%), to improve their way of life over the next 20 years. The case for AI and LT is further laid out by various European Institutions in reports and policy initiatives that highlight their extensive impact on society and what must be done to shepherd this influence. These include, among others, *European Artificial Intelligence (AI) leadership, the path for an integrated vision*;[47] *Strategy on AI*;[48] *Ethics Guidelines for Trustworthy AI*;[49] *Liability for AI and other emerging technologies*;[50] *On Artificial Intelligence: A European approach to excellence and trust*;[51] and *Coordinated Plan on AI*.[52] All agree that AI is an area of strategic importance, a key driver of economic development, and a means to provide solutions to many societal challenges. As such, they concur that the socioeconomic, legal and ethical impact of AI must be carefully measured. For instance, the Joint Research Center (JRC) Science for Policy report, *The Changing Nature of Work and Skills in the Digital Age*,[53] observes that employment opportunities related to the development and maintenance of AI technologies and Big Data infrastructures are expected to grow, whereas jobs that are most vulnerable to automation appear to be those that require relatively low levels of formal education, do not involve complex social interaction, or demand routine manual tasks. Keeping this range in mind is a reminder that digital technologies may not only create or destroy some lines of work, but also fundamentally change what people do on the job and how they do it.

The European Commission's new *Coordinated Plan on AI*, which affirms that NLP is one of the most rapidly advancing fields in AI, is designed to address such potential turbulence.[54] The 2021 plan, in conjunction with the first-ever legal framework for AI,[55] will guarantee the safety and rights of people and businesses, while strengthening AI uptake, investment and innovation across the EU. It is also seen as the EU's next step in fostering global leadership in trustworthy AI, deemed necessary if European AI is to be globally competitive while respecting European values. This is of particular concern given that the EC's 2021 Strategic Foresight Report, *The EU's capacity and freedom to act*,[56] stresses the EU's capabilities in AI, Big

---

[46] https://europa.eu/eurobarometer/surveys/detail/2237

[47] https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU(2018)626074

[48] https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence#Building-Trust-in-Human-Centric-Artificial-Intelligence

[49] https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[50] https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=63199

[51] https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

[52] https://ec.europa.eu/digital-single-market/en/news/coordinated-plan-artificial-intelligence

[53] https://publications.jrc.ec.europa.eu/repository/handle/JRC117505

[54] https://digital-strategy.ec.europa.eu/en/library/new-coordinated-plan-artificial-intelligence

[55] https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence

[56] https://ec.europa.eu/info/strategy/strategic-planning/strategic-foresight/2021-strategic-foresight-report_en

Data and Robotics lag behind the world's leaders, the US and China. To strengthen digital sovereignty and European AI, the report encourages stakeholders to promote values via the finance, development and production of next generation tech.

One important area of focus must be high-value data, a key factor in improving performance and building robust AI models. The EC wants to ensure legal clarity in AI-based applications, especially regarding data. Its proposed regulation on data governance will help by boosting data sharing across sectors and member states, while the General Data Protection Regulation (GDPR) is a major step towards building trust.[57] The member states also recently agreed to a negotiating mandate on a proposal for a Data Governance Act (DGA).[58] The DGA is part of a wider policy to give the EU a competitive edge in the increasingly data-driven economy. The aim is to promote the availability of data that can be utilised to power applications and advanced solutions in AI, personalised medicine, green mobility, smart manufacturing and numerous other areas. While these regulations support the privacy and rights of European citizens, it should be pointed out that significant barriers to the access and re-use of language resources remain, especially with regard to competition with countries that have adopted the "fair use" doctrine, such as the US, Japan or Korea.

Research infrastructures play a role in this regard, including the Common Language Resources and Technology Infrastructure (CLARIN), an ESFRI Landmark and ERIC which offers access to LRs and LTs for researchers in the humanities and social sciences.[59] Not every EU Member State is officially affiliated with it, while others participate only as observers (Belgium joined CLARIN in 2021 and Spain will join in 2023). Additionally, because research funding agencies provide unbalanced resources to the different Member States, European languages are not equally supported by CLARIN (de Jong et al. 2020). This problem has received more attention in the EU project European Language Grid (ELG), which started in 2019 and concluded in June 2022. The ELG cloud platform contains more than 14,000 running services and resources for all European languages (Rehm et al. 2021; Rehm 2023).[60]

Experience with infrastructures such as these has demonstrated that the EU's approach to data infrastructures must be crafted with Big Data technology and LT in mind. The ESFRI roadmap includes a "Landscape Analysis" that provides an advanced analysis of the scientific needs and existing research infrastructure gaps as well as directions for strategic investments in the future that would help maintain Europe's leadership in the global context. According to its findings, research infrastructures in LT are indispensable in breaking new ground because they represent a core aspect of Big Data technology due to the volume and variety of data generated by the accumulation of unstructured text. And as the main task in AI's communication domain, NLP encompasses applications such as text generation, text mining, text classification, MT and speech recognition. Put differently, LT's ability to analyse, understand and generate information expressed in natural language is crucial for im-

---

[57] https://eur-lex.europa.eu/eli/reg/2016/679/oj

[58] https://www.consilium.europa.eu/en/press/press-releases/2021/10/01/eu-looks-to-make-data-sharing-easier-council-agrees-position-on-data-governance-act/

[59] https://www.clarin.eu

[60] https://www.european-language-grid.eu

proving human-computer interaction. This view is confirmed by AI Watch, the EC's knowledge service responsible for monitoring the development, uptake and impact of AI, in three recent reports, *Defining Artificial Intelligence*, *Artificial Intelligence in public services* and *AI Watch, road to the adoption of Artificial Intelligence by the public sector*.[61] By way of example, the latter identified and employed 230 cases of AI usage in public services in order to extract emerging trends in AI, revealing that well over half of the cases are closely related to LT.

Relatedly, the EC's Directorate-General for Communications Networks, Content and Technology (DG CNECT), in collaboration with the Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs (DG GROW), opened a consultation in 2021 that examined use-cases for website translation at small and medium-sized enterprises (SMEs) and surveyed multilingual websites in an effort to analyse language barriers across EU Member States.[62] The inquiry identified specific market needs that could be addressed through public solutions, such as eTranslation,[63] and by European language service providers. Of the over 1,000 SMEs that responded, 75% expressed interest in participating in the EC's subsequent pilot programme to make their website automatically multilingual. When the *European Language Industry Survey* (ELIS)[64] – then known as the *EUATC survey* – was run for the first time in 2013, MT was still primarily seen as a threat and a challenge. Only a few language companies saw it as an opportunity. Today, 65% of language company respondents see the improved quality of neural MT as an opportunity rather than a threat. According to the 2022 survey, 58% of those companies have implemented the technology and an additional 20% are planning to do so. This potential willingness to incorporate LT and AI corresponds with a separate study conducted by Eurostat[65] in 2020. It found that 7% of EU enterprises with at least ten employees used AI applications, 2% utilised ML to analyse big data internally, and 1% evaluated big data internally with the help of LT. Moreover, 2% provided a chat service, where a chatbot or virtual agent generated natural language replies to customers.

3 Major Language Technology Initiatives in Europe

First, we take a closer look at European initiatives (Section 3.1) and then examine national and also regional initiatives (Section 3.2).

---

[61] https://knowledge4policy.ec.europa.eu/ai-watch_en; https://publications.jrc.ec.europa.eu/repository/handle/JRC118163; https://publications.jrc.ec.europa.eu/repository/handle/JRC120399; https://joinup.ec.europa.eu/collection/innovative-public-services/news/ai-watch-road-adoption-artificial-intelligence

[62] https://digital-strategy.ec.europa.eu/en/library/report-sme-survey-multilingual-websites

[63] https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eTranslation; https://ec.europa.eu/education/knowledge-centre-interpretation/eu-initiatives-language-technologies_en

[64] https://elis-survey.org

[65] https://ec.europa.eu/eurostat/

3.1 European Initiatives

The European Parliament recently emphasised that "multilingualism presents one of the greatest assets of cultural diversity in Europe and, at the same time, [is] one of the most significant challenges for the creation of a truly integrated EU." (European Parliament 2018). The belief is reflected in the EU's promotion of multilingualism, which falls within the scope of a variety of EU policy areas. While many of the multifaceted efforts to support Europe's languages are bearing fruit, still greater attention must be paid to removing barriers to intercultural and inter-linguistic dialogue as a means to stimulate mutual understanding. One means to achieve this is through language technology. Nonetheless, although official EU languages are granted equal status politically, they are far from equally supported from a technological perspective (see, e. g., Rehm and Uszkoreit 2012; Rehm et al. 2014; Rehm and Hegele 2018; Rehm et al. 2020b, as well as the chapters in Part I of this book).

Several strategic documents have contributed to the European debate on this subject in the past decade, including *The FLaReNet Strategic Language Resource Agenda* (Soria et al. 2014), *META-NET Strategic Research Agenda for Multilingual Europe 2020* (Rehm and Uszkoreit 2013; Rehm et al. 2016), *Language Technologies for Multilingual Europe: Towards a Human Language Project* (Rehm 2017), and the STOA report, *Language Equality in the digital age: Towards a Human Language Project* (STOA 2018). The latter helped pave the way for the preparation of the European Parliament's joint ITRE/CULT resolution, *Language equality in the digital age* (European Parliament 2018),[66] adopted in a plenary meeting in September 2018 with an overwhelming majority of 592 votes in favour, 45 against and 44 abstentions.

Approval of the resolution by such a wide margin demonstrates the importance and relevance of the issue. It includes more than 40 recommendations, structured into four sections: "Improving the institutional framework for language technology policies at EU level", "Recommendations for EU research policies", "Education policies to improve the future of language technologies in Europe" and "Language technologies: benefits for both private companies and public bodies". Among the most salient items are the following (emphases added; some items abbreviated):

- The report "recommends that in order to raise the profile of language technologies in Europe, the Commission *should allocate the area of 'multilingualism and language technology' to the portfolio of a Commissioner*; considers that the Commissioner responsible *should be tasked with promoting linguistic diversity and equality at EU level*, given the importance of linguistic diversity for the future of Europe;" (item 14)
- "suggests *ensuring comprehensive EU-level legal protection for the 60 regional and minority languages*, recognition of the collective rights of national and linguistic minorities in the digital world, and mother-tongue teaching for speakers of official and non-official languages of the EU;" (item 15)

---

[66] https://www.europarl.europa.eu/doceo/document/TA-8-2018-0332_EN.html

- "calls on *the Member States to develop comprehensive language-related policies and to allocate resources and use appropriate tools in order to promote and facilitate linguistic diversity and multilingualism in the digital sphere*; stresses the *shared responsibility of the EU and the Member States* and in developing databases and translation technologies for all EU languages, including languages that are less widely spoken; calls for *coordination between research and industry* with a common objective of enhancing the digital possibilities for language translation and with open access to the data required for technological advancement;" (item 17)
- "calls on the Commission *to establish a large-scale, long-term coordinated funding programme for research, development and innovation in the field of language technologies, at European, national and regional levels*, tailored specifically to Europe's needs and demands; emphasises that the programme should seek to tackle *deep natural language understanding* and increase efficiency by sharing knowledge, infrastructures and resources, with a view to developing innovative technologies and services, in order *to achieve the next scientific breakthrough* in this area and help to reduce the technology gap between European languages; stresses that this should be done with the participation of research centres, academic, enterprises […] and other relevant stakeholders;" (item 25)
- "believes that […] *European education policies should be aimed at retaining talent in Europe*, should analyse the current educational needs related to language technology […] and, based on this, *provide guidelines for the implementation of cohesive joint action at European level* […] including the language-centric artificial intelligence industry;" (item 34)
- "points to the need *to promote the ever-greater participation of women in the field of European studies on language technologies*, as a decisive factor in the development of research and innovation" (item 36)

To these recommendations may be added the remarks made by EC Commissioner Corina Crețu in her closing statement at the hearing on the resolution:

> Ensuring appropriate technological support for all European languages will […] create jobs, growth and opportunities in the DSM [(Digital Single Market)]. It will enhance the quality of public services, and reinforce a stronger sense of unity and belonging throughout Europe. […] [U]nder the next Multiannual Financial Framework (MFF), we will need to reinforce funding, research and education actions. […] [O]vercoming language barriers in the digital environment is essential for an inclusive society, a vibrant DSM and for unity in diversity.

Crețu's statement is in line with previous public appeals voiced in 2016 by former EC Vice President Andrus Ansip and in 2017 by Director General Roberto Viola (DG CNECT) for the need to strengthen multilingualism through technologies.[67]

More recently, the EP's CULT Committee adopted a resolution on AI in the cultural, creative and educational sector in which multilingual and linguistic diversity

---

[67] See *How multilingual is Europe's Digital Single Market?* (https://ec.europa.eu/commission/commissioners/2014-2019/ansip/blog/how-multilingual-europes-digital-single-market_en); *Multilingualism in the Digital Age: a barrier or an opportunity* (https://ec.europa.eu/digital-single-market/en/blog/multilingualism-digital-age-barrier-or-opportunity).

is also taken into account.[68] Regarding the latter, the resolution calls for: 1. AI technologies to be regulated and trained in order to ensure non-discrimination, gender equality, pluralism, as well as cultural and linguistic diversity; 2. specific indicators to measure diversity in order to promote European ventures and prevent algorithm-based recommendations that negatively affect the EU's cultural and linguistic diversity; and 3. an ethical framework for the use of AI technologies in EU media that guarantees access to culturally and linguistically diverse content. Such a framework would also address the misuse of AI to disseminate fake news and disinformation.[69] The resolution goes hand in hand with a study commissioned by the EC that explores the possibilities of applying AI technologies in ten domains that also belong to the cultural, creative and educational sector. The study aims to inspire creative entrepreneurs as well as policy-makers with concrete use cases and recommendations for the application of AI,[70] focusing partly on language-centric AI (NLP, NLU, speech technologies). The resolution also reflects the conclusions of the Education, Youth, Culture and Sport Council held on 4-5 April 2022, which called for the development of an ambitious digital policy for language technologies, translation, and lifelong language learning and teaching. This objective fits with the EU's desire to take advantage of new technologies to foster multilingualism, which it hopes will facilitate access to culture and nurture cultural exchange.[71]

A key commonality in these documents and initiatives is the idea that LT must be *made in* Europe *for* Europe. This approach will not only strengthen Europe's place at the pole position of research excellence, but also contribute to future European cross-border and cross-language communication, economic growth and social stability. The past few years have witnessed a flurry of white papers and SRAs offering roadmaps and recommendations for how best to attain the goal. In 2019, the European Language Resource Coordination (ELRC) white paper, *Sustainable Language Data Sharing to Support Language Equality in Multilingual Europe. Why Language Data Matters*, underscored that the main challenge is a lack of appreciation for the value of language data.[72] To help overcome this perception, the group issued several recommendations aimed at the European and national policy level, including:

- Updating the Open Data Directive (2019/1024/EU) so that it references language data as a high-value data category.[73]
- Conducting a study on language data to identify and quantify the value of language data for citizens, public administrations and businesses.

---

[68] https://www.europarl.europa.eu/news/en/press-room/20210311IPR99709/ai-technologies-must-prevent-discrimination-and-protect-diversity; https://oeil.secure.europarl.europa.eu/oeil/popups/summary.do?id=1663438&t=e&l=en

[69] https://op.europa.eu/en/publication-detail/-/publication/b8722bec-81be-11e9-9f05-01aa75ed71a1

[70] https://digital-strategy.ec.europa.eu/en/library/study-opportunities-and-challenges-artificial-intelligence-ai-technologies-cultural-and-creative

[71] https://www.consilium.europa.eu/en/meetings/eycs/2022/04/04-05/

[72] https://lr-coordination.eu/sites/default/files/Documents/ELRCWhitePaper.pdf

[73] https://digital-strategy.ec.europa.eu/en/policies/legislation-open-data

- Updating national policies (e. g., Open Data policies, digital agenda or strategies for AI) to explicitly support the sharing of language data and LT.
- Including obligatory (language) data management plans in all relevant national funding policies and calls for proposals if not yet included.
- Conducting national surveys to assess translation practices in public administrations at all levels.

These steps will contribute to the development of an inclusive European digital society, a task for which European LT is essential. However, still others are required. The *Report on the Joint Stakeholder Consultation on Research and Innovation in Web Accessibility and Language Technologies*, for instance, highlights that greater work must be done to develop systems capable of adapting and personalising digital content according to individual needs, particularly in terms of accessibility and language.[74] Research into sign languages represents one avenue that merits greater attention, given that sign languages are increasingly becoming recognised as official national languages. Another relevant is the accessibility of information in multimodal contexts with respect to formatting and the understanding of content.

Fortunately, it is evident that the EU is not blind to LT's crucial role in building Europe's digital society and has already begun to dedicate funding and launch initiatives to advance LT and AI. Research, industry, and the public sector have benefitted from these actions. Two prominent examples include the Horizon 2020 Programme and the Connecting Europe Facility (CEF).[75] LT was embedded in the former within research and innovation in the field of information technologies, content technologies, multilingual internet and AI. Through the latter, MT tools (eTranslation) and tools for the management of thesauri and glossaries have been developed (VocBench).[76] There is, however, much left to be done. The *Final study report on CEF Automated Translation value proposition in the context of the European LT market/ecosystem* provides an analysis of the EU's LT market (including Norway and Iceland) and the adoption of LT by public administrations, both at the EU and national levels.[77] The report underscores that EU industry is fragmented and that many small players struggle to compete with the global giants that dominate the market. It further notes that European businesses and the public sector have become dependent on these non-European global companies, which have massive amounts of data at their disposal due to both copyright disparities between the EU (explicit permission required by European entities) and the US (fair use copyright exception), as well as intensive use of their popular systems.

---

[74] https://ec.europa.eu/digital-single-market/en/news/report-joint-stakeholder-consultation-research-and-innovation-web-accessibility-and-language-0. See also the New European Media's SRIA: https://nem-initiative.org/wp-content/uploads/2020/06/nem-strategic-research-and-innovation-agenda-2020.pdf?x98588

[75] https://ec.europa.eu/programmes/horizon2020/en/h2020-section/information-and-communication-technologies; https://ec.europa.eu/digital-single-market/en/connecting-europe-facility; https://ec.europa.eu/digital-single-market/en/language-technologies

[76] https://ec.europa.eu/isa2/solutions/vocbench3_en

[77] https://op.europa.eu/en/publication-detail/-/publication/8494e56d-ef0b-11e9-a32c-01aa75ed7 1a1/language-en/format-PDF/source-106906783

Nonetheless, the dependency on American or Chinese systems and the torrent of data flowing out of Europe mask areas in which European initiatives may make real the ideal of LT made *in* Europe *for* Europe. Several large international tech companies, by way of example, provide MT services free of charge. EU industry, by contrast, is experienced in navigating through Europe's many languages and European MT developers have successfully deployed services for the public sector through the support of EU-funded programmes. LT made for Europe means harnessing this know-how to support MT for all its languages and create domain-specific and application-specific MT while simultaneously being attentive to security and privacy issues. Moreover, as stated in *My Europe. My language: With language technologies made in the EU*,[78] LT offers opportunities to reduce language barriers across Europe and in the DSM at the intersection of Big Data, AI and HPC. Indeed, the European High Performance Computing Joint Undertaking[79] (EuroHPC JU), a joint initiative between the EU, European countries and private partners, is developing a world-class supercomputing ecosystem in Europe.[80] The Language Data Space EU project, a platform and marketplace for the collection, creation, sharing and re-use of multilingual and multimodal language data, was launched in January 2023.[81]

The EC has also established public-private partnerships (PPPs) in the area of AI.[82] As detailed by Curry et al. (2021), the Big Data Value PPP, created by the EC and the BDVA in 2014, represented a substantial collective effort on the part of the European data community to formulate a set of technical research priorities for Big Data. According to the report, Europe's multilingualism presents a particular challenge when it comes to data:

> Large amounts of data are being made available in a variety of formats ranging from unstructured to semi-structured to structured formats […] A great deal of this data is created or converted and further processed as text. Algorithms or machines are not able to process the data sources due to the lack of explicit semantics. In Europe, text-based data resources occur in many different languages, since customers and citizens create content in their local language. This multilingualism of data sources means that it is often impossible to align them using existing tools because they are generally available only in the English language. Thus, the seamless aligning of data sources for data analysis or business intelligence applications is hindered by the lack of language support and gaps in the availability of appropriate resources.[83]

---

[78] https://digital-strategy.ec.europa.eu/en/library/my-europe-my-language-language-technologies-made-eu-brochure

[79] https://eurohpc-ju.europa.eu

[80] https://digital-strategy.ec.europa.eu/en/activities/work-programmes-digital

[81] https://digital-strategy.ec.europa.eu/en/funding/language-data-space-call-tenders

[82] https://adr-association.eu

[83] https://elements-of-big-data-value.eu/research-priorities-for-big-data-value

The Big Data Value PPP's successor, the AI, Data and Robotics Partnership (formed in 2020 along with BDVA,[84] euRobotics,[85] ELLIS,[86] CLAIRE,[87] and EurAI[88]) expanded on this issue and zeroed in on NLP's importance in its Strategic Research, Innovation and Deployment Agenda,[89] "Natural Language Processing has particular resonance within Europe's multi-lingual landscape and offers the potential to harmonise human interaction." Unfortunately, although the PPP includes LT experts, research groups and companies via some the groups involved, currently no European LT association or network is represented in the PPP.

The initiative, however, complements the Coordinated Plan on Artificial Intelligence (CPAI) proposed by the European Commission for the period 2021-2027. The plan, which considers AI an area of strategic importance and aims to propel Europe to the forefront in terms of developing and exploiting AI technologies, calls for the EU to provide a minimum one billion euro annual investment in Horizon Europe and Digital Europe, although the objective is to reach twenty billion euros a year between public and private investments.[90] The focus is on four key areas: increasing investment in AI; the availability of data; the promotion of talent; and ensuring security, ethics and trust in AI. Success in these domains leans on the belief that member states must develop and coordinate their own national AI strategies, of which an analysis and comparison is provided in the report *AI Watch: National strategies on Artificial Intelligence: A European perspective in 2019*.[91]

3.2 National and Regional Initiatives

The perspective that the EU Member States should be responsible for their individual AI strategies stems partly from the observation that each country or region is best placed to address their own particular needs. The response by European countries to the CPAI has been largely positive and the number of states with an AI strategy (29 out of 30; only Croatia has no official strategy as of yet) demonstrates its success. Moreover, it is in the national plans that currently exist where many of the initiatives concerning LT and language-centric AI reside, although this is not to say that dedicated LT programmes are widespread in Europe. And in comparison to non-EU national AI initiatives, Europe's member states lag behind when LT is taken into account. Since Canada published the world's first national AI strategy in 2017, more

---

[84] https://www.bdva.eu

[85] https://www.eu-robotics.net

[86] https://ellis.eu

[87] https://claire-ai.org

[88] https://eurai.org

[89] https://adr-association.eu/wp-content/uploads/2020/09/AI-Data-Robotics-Partnership-SRIDA-V3.0-1.pdf

[90] https://knowledge4policy.ec.europa.eu/ai-watch/coordinated-action-plan-ai_en

[91] https://ec.europa.eu/jrc/en/publication/ai-watch-national-strategies-artificial-intelligence-european-perspective-2019

than 30 other countries and regions have published similar documents as of December 2020.[92] Several non-EU nations merit brief consideration here due to the explicit inclusion of NLP in their plans. China's AI strategy, one of the most comprehensive in the world, singles out NLU technology as a decisive area to promote university AI curricula and in its pursuit of AI talent (Zhang et al. 2021). The UK, which emphasises a strong partnership between business, academia, and government, created a pilot programme for under-18-year-olds to encourage careers in the AI sector, explicitly mentioning NLP. India's approach to AI considers the multilingual reality of the country a means to achieve technological leadership in AI and cites the development of an advanced NLP infrastructure for its languages as a stepping stone in that direction.[93] Finally, the US emphasises the crucial role LT plays in AI and NLU appears as one of the six "Uses for Deployed AI Today" in the National Security Commission on Artificial Intelligence's *Final Report*, published in 2021.[94]

In Europe, only a handful of dedicated national programmes funded projects related to LT before 2018.[95] Instead, financial support for the development of LT was generally provided through generic R&D&I calls in most member states. The Spanish case is one of those notable exceptions. The Spanish government has recently announced a new strategic plan for economic recovery and transformation (PERTE) called "The New Economics of Language."[96] The PERTE is presented as an opportunity to take advantage of the potential of Spanish and co-official languages for economic growth and international competitiveness in areas such as AI, translation, learning, cultural dissemination, audiovisual production, research and science. It has a budget of 1.1 billion euros in public funds and aims to mobilise another billion in private investment. Additionally, following the lines of the Spanish Plan for the Advancement of LT,[97] several regional governments have also launched LT initiatives, including AINA (Catalonia),[98] Nós (Galicia)[99] and GAITU (the Basque Country).[100]

At the European level, LT received better support through calls in various programmes: FP7, Horizon 2020, CEF Telecom, CIP ICT-PSP, EUREKA and EU-

---

[92] https://aiindex.stanford.edu/report/

[93] *AI in India: A Policy Agenda*. The report also highlights natural language voice recognition as a way to to account for the diversity in languages and digital skills in the Indian context and recommends the creation of annotated data sets for their languages to add incremental value to existing services ranging from e-commerce to agriculture.

[94] https://www.nscai.gov/2021-final-report. See also, the *American AI Initiative*.

[95] Spanish *Plan for the Advancement of Language Technology*: https://plantl.mineco.gob.es/tecnologias-lenguaje/actividades/estudios/Paginas/tecnologias-del-lenguaje-en-Europa.aspx

[96] https://planderecuperacion.gob.es/como-acceder-a-los-fondos/pertes/perte-nueva-economia-de-la-lengua

[97] https://plantl.mineco.gob.es/Paginas/index.aspx

[98] https://politiquesdigitals.gencat.cat/ca/tic/aina-el-projecte-per-garantir-el-catala-en-lera-digital/

[99] https://www.xunta.gal/hemeroteca/-/nova/134792/xunta-usc-ponen-marcha-lsquo-proxecto-nosrsquo-que-permitira-incorporar-galego

[100] https://www.irekia.euskadi.eus/es/news/76846-gobierno-vasco-presentado-gaitu-plan-accion-las-tecnologias-lengua-2021-2024-cual-tiene-objetivo-integrar-euskera-las-tecnologias-linguisticas

	LT-related funding			Artificial Intelligence	
	None at all	Some funding	Dedicated LT programme	AI strategy	LT funding through AI
Austria	●			●	
Belgium		●		D	●
Bulgaria		●		●	
Croatia	●				
Cyprus				●	
Czechia		●		●	
Denmark			●	●	●
Estonia			●	●	●
Finland		●		●	
France		●		●	●
Germany		●		●	●
Greece		●		D	
Hungary		●		●	
Iceland			●	●	
Ireland		●		●	
Italy		●		●	
Latvia		●		●	
Lithuania		●		●	
Luxembourg		●		●	
Malta		●		●	●
Netherlands		●		●	
Norway		●		●	
Poland		●		●	
Portugal		●		●	
Romania		●		D	
Serbia	●			●	
Slovakia	●			●	
Slovenia		●		●	
Spain			●	●	
Sweden		●		●	

Table 1 The Language Technology funding situation in Europe (2019/2021), extracted from Rehm et al. (2020b) and updated with the newest AI strategies (D: Draft)

ROSTARS, among others. However, in these most funding for LT projects gradually reduced as well. If these findings are compared to those presented by Rehm et al. (2020b), we observe a slight increase in the number of language-centric AI initiatives over the next couple of years (see Table 1 and Figure 2).[101] It is noteworthy that only 12 European countries out of the 30 studied explicitly consider LT within their national policy initiatives. This is significant because the successful development of the next generation of innovative AI technology relies on setting aside funding

---

[101] According to Rehm et al. (2020b), only four of the 30 surveyed countries do not have some level of LT funding. Four countries have programmes dedicated to LT (Denmark, Estonia, Iceland, Spain), six provide funding for LT-related topics through AI (Belgium, Denmark, Estonia, France, Germany, Malta) and two (Ireland, Latvia) that do not have LT programmes, but rather a language strategy defined by their governments. See also Rehm et al. (2016, 2020a, 2021).
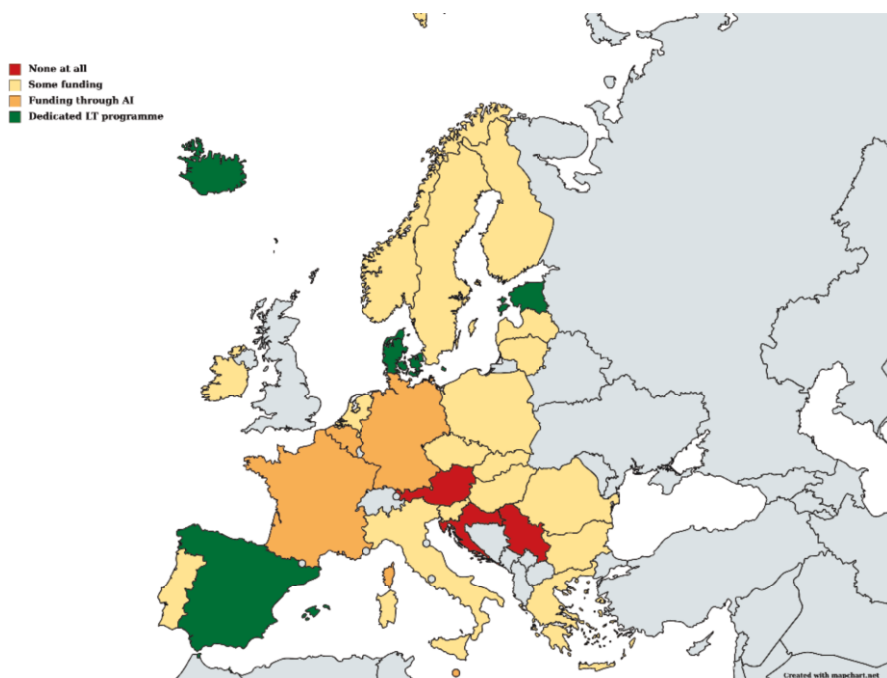
Fig. 2 The Language Technology funding situation in Europe

exclusively for LT. The same holds true for European countries that hope to incorporate LT-based AI applications, such as interactive dialogue systems and personal virtual assistants, into public services.[102]

4 SWOT Analysis

This section summarises, in the form of a SWOT analysis, the most relevant findings of the reports, documents and initiatives that were reviewed for this chapter. It attempts to identify the most significant favourable and unfavourable factors that must be addressed to make digital language equality a reality in Europe by 2030.

---

[102] https://digital-strategy.ec.europa.eu/en/news/new-report-looks-ai-national-strategies-progress-and-future-steps

4.1 Strengths

- Emergence of powerful new deep learning techniques, tools that are revolutionising LT.
- Important basic LT has been developed and applications that are used on a daily basis by hundreds of millions of users for speech recognition, speech synthesis, text analytics and MT are available.
- Existence of multiple national and European LT research networks, associations, communities and other relevant stakeholders whose objective is to promote all kinds of activities related to research, development, education and industry in the field of LT, both nationally and internationally.
- Existence of unique, valuable and potentially extremely useful data resources that can be exploited by current LT. An enormous amount of data is expressed in human language.
- Increasing number of companies in LT and good level of readiness for the implementation of LT in production environments.
- LT contributes to the development of inclusive digital societies, and is critical for responding to social challenges (accessibility, transparency, equity).

4.2 Weaknesses

- Deep learning LT and large pre-trained language models have shortcomings. Language models have limited real-world knowledge, can generate biased and factually incorrect text, may contain personal information, etc. They are also expensive to train and have a heavy carbon footprint. It is important to understand the limitations of large pre-trained language models and put their success in context.
- The LT markets are currently dominated by large non-EU actors, which do not address the specific needs of a multilingual Europe; Europe remains far behind due to market fragmentation, insufficient funding and legal barriers, thus hindering online commerce and communication. Europe does not fully exploit its enormous potential in LT.
- LT currently only plays a rather subordinate role in the political agenda and public debate of the EU and most of its Member States.
- There is a general misconception and over-hyping of actual AI and LT capabilities. AI is often perceived in a polarised fashion as either "magical" technology that can solve any problem or as a threat to jobs and workers, who will be replaced by machines.
- No EU policy has been proposed to address the problem of language barriers.
- GDPR/Copyright is a major barrier to the access and re-use of language resources, in competition with countries that adopt the "fair use" doctrine.

- The Open Data Directive (2019/1024/EU) does not include language data as a "high-value data category". Most of the data require extensive IPR clearing (to address Copyright and GDPR).
- There is a lack of adequate LT policies and sustainability plans at the European and national levels to properly support European languages through LT. Only four of the 30 European countries studied have a dedicated LT national programme, only six have included LT funding through national AI strategies.
- There is scarce and limited LT support for non-official EU languages.
- No European LT association is represented in the new Data, AI and Robotics public-private partnership.
- There is a lack of necessary resources (experts, HPC capabilities, etc.) compared to large US and Chinese enterprises that lead the development of new LT systems. In particular, the "computing divide" between large firms and non-elite universities increases concerns around bias and fairness within AI technology, and presents an obstacle towards democratising AI.
- Compared to English, there are far fewer LT resources and tools including language resources, annotated corpora, pre-trained language models, benchmark datasets, software libraries, etc.
- There is an uneven distribution of resources (funding, open data, language resources, scientists, experts, computing facilities, IT companies, etc.) by country, region and language.
- There is a weak open data sharing culture for many public stakeholders and SMEs.
- The investment in AI does not reflect the real importance of LT.
- There is a fragmented European market with an extremely large and varied base of about 1,000 SME companies that develop LT. Small to medium national technology companies have little capital and investment in LT capabilities. The markets are small for low-resource language speakers.
- In many countries, there are weak links between academia and industry and insufficient effective mechanisms for knowledge transfer.
- There is weak internationalisation of R&D&I and innovation.

4.3 Opportunities

- Many new powerful monolingual, multilingual and cross-lingual deep learning LT capabilities are available.
- LT is key for the realisation and support of European multilingualism.
- LT is used in practically all everyday digital products and services, since most use language to some extent, especially all internet-related products such as search engines, social networks and e-commerce services.
- LT can impact on sectors of fundamental importance to the well-being of all European citizens, such as health, administration, justice, education, culture, tourism, etc.

- LT offers effective solutions to facilitate monolingual and multilingual communication, including for the deaf and hard of hearing, the blind and visually impaired and those with language-related disabilities or impairments.
- LT is one of the most important AI application areas with a fast growing economic impact. Enormous growth is expected in the global LT market based on the explosion of applications observed in recent years and the expected exponential growth in unstructured digital data.
- Europe can play an economic leading role with its neighbouring countries through good partnerships based on the use of LT customised to other languages.
- Growing trend for the LT market and industry in Europe regarding the exploitation of digital resources and data of linguistic interest. Digitisation is one of the key means to generate new economic growth.
- Consolidation of a competitive LT industry that harnesses the potential of research and academia both in educating well-trained LT professionals and in transferring research results to industry and public administrations.
- Increasing awareness about the possibilities of AI and LT and the necessity to invest and coordinate efforts.
- Substantial breakthroughs and fast development of LT offer new opportunities for digital communication; current multilingual and cross-lingual deep learning LT allows for the creation of new multilingual pre-trained language models and systems that can leverage and balance LT across all European languages.
- Ensure openness of infrastructures for data and technologies.

4.4 Threats

- In comparison to 2012, the results of the European Language Equality project in 2022 show that the gap between English and all other languages appears to be getting *bigger* instead of smaller.
- Development of non-explainable techniques and deep learning models without any commonsense or up-to-date knowledge, with social biases, containing personal and private data, with a heavy impact on carbon footprint, etc.
- AI is a broad area, which overshadows and dwarfs the importance, benefits and contributions of LT, especially in Europe.
- Loss of LT skills and human capital trained in Europe due to the lack of sufficient research, transfer and funding opportunities.
- Inability to retain in, or attract to, the EU researchers and workers skilled in LT and AI.
- Growing development of the sector in US and China that will eventually penetrate the European application market, limiting the Digital Language Equality opportunities as described in this report.
- The complexity of copyright, GDPR, Open Data directives etc. makes access to language resources too costly, unclear and risky.

- Fear of many jobs becoming redundant due to the deployment of AI-powered technologies.

5 Conclusions

Europe's multilingual nature is also one of the main obstacles to a truly connected, cross-lingual communication and information space. Moreover, while language diversity is at the core of European identity, many of our languages are in danger of digital extinction because they are not sufficiently supported through Language Technologies (Moseley 2010; Rehm and Uszkoreit 2012; STOA 2018; European Parliament 2018).[103] Sophisticated multilingual, cross-lingual and monolingual LT for all European languages would future-proof our languages as cornerstones of our cultural heritage and richness. In recent years, European research in LT has faced increased competition from other continents, especially with respect to breakthroughs in AI. These scientific advancements have led to global commercial successes, from which the respective regions benefit especially. As a consequence, many European scientists, including young high-potential researchers, are leaving Europe to continue their work abroad. Europe must invest in retaining and attracting these researchers. Our continent is in need of powerful LT made *in* Europe *for* all European citizens, tailored to our unique cultures, societies and economic requirements so that a linguistically fragmented Europe may become a truly unified and inclusive one. This ambitious but worthy effort involves supporting its rich and diverse linguistic cultural heritage, from broadly spoken languages to minority and regional languages, as well as the languages of immigrants and important trade partners, benefiting European citizens, European industry and European society.

References

Aldabe, Itziar, Georg Rehm, German Rigau, and Andy Way (2022). *Deliverable D3.1 Report on existing strategic documents and projects in LT/AI (second revision)*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/LT-strategic-documents-v3.pdf.

Chui, Michael, Martin Harryson, James Manyika, Roger Roberts, Rita Chung, Ashley van Heteren, and Pieter Nel (2018). "Notes from the AI frontier: Applying AI for social good". In: *McKinsey Global Institute*.

Curry, Edward, Andreas Metzger, Sonja Zillner, Jean-Christophe Pazzaglia, and Ana García Robles, eds. (2021). *The Elements of Big Data Value: Foundations of the Research and Innovation Ecosystem*. Cham: Springer.

de Jong, Franciska, Bente Maegaard, Darja Fišer, Dieter van Uytvanck, and Andreas Witt (2020). "Interoperability in an Infrastructure Enabling Multidisciplinary Research: The case of CLARIN". In: *Proceedings of the Twelfth Language Resources and Evaluation Conference*. Mar-

---

[103] http://www.unesco.org/languages-atlas/index.php?hl=en&page=atlasmap

seille, France: European Language Resources Association, pp. 3406–3413. https://aclantholog y.org/2020.lrec-1.417.

European Parliament (2018). *Language Equality in the Digital Age. European Parliament resolution of 11 September 2018 on Language Equality in the Digital Age (2018/2028(INI)*. http://w ww.europarl.europa.eu/doceo/document/TA-8-2018-0332_EN.pdf.

Moseley, Christopher (2010). *Atlas of the World's Languages in Danger*. http://www.unesco.org/c ulture/en/endangeredlanguages/atlas.

Padilla, Thomas (2020). "Responsible Operations: Data Science, Machine Learning, and AI in Libraries". In: *American Archivist* 83, pp. 483–487.

Rehm, Georg, ed. (2017). *Language Technologies for Multilingual Europe: Towards a Human Language Project. Strategic Research and Innovation Agenda*. CRACKER and Cracking the Language Barrier federation. http://cracker-project.eu/sria/.

Rehm, Georg, ed. (2023). *European Language Grid: A Language Technology Platform for Multilingual Europe*. Cognitive Technologies. Cham, Switzerland: Springer.

Rehm, Georg, Maria Berger, Ela Elsholz, Stefanie Hegele, Florian Kintzel, Katrin Marheinecke, Stelios Piperidis, Miltos Deligiannis, Dimitris Galanis, Katerina Gkirtzou, Penny Labropoulou, Kalina Bontcheva, David Jones, Ian Roberts, Jan Hajic, Jana Hamrlová, Lukáš Kačena, Khalid Choukri, Victoria Arranz, Andrejs Vasiļjevs, Orians Anvari, Andis Lagzdiņš, Jūlija Meļņika, Gerhard Backfried, Erinç Dikici, Miroslav Janosik, Katja Prinz, Christoph Prinz, Severin Stampler, Dorothea Thomas-Aniola, José Manuel Gómez Pérez, Andres Garcia Silva, Christian Berrío, Ulrich Germann, Steve Renals, and Ondrej Klejch (2020a). "European Language Grid: An Overview". In: *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)*. Ed. by Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Christopher Cieri, Khalid Choukri, Thierry Declerck, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis. Marseille, France: ELRA, pp. 3359–3373. https://w ww.aclweb.org/anthology/2020.lrec-1.413/.

Rehm, Georg and Stefanie Hegele (2018). "Language Technology for Multilingual Europe: An Analysis of a Large-Scale Survey regarding Challenges, Demands, Gaps and Needs". In: *Proceedings of the 11th Language Resources and Evaluation Conference (LREC 2018)*. Ed. by Nicoletta Calzolari, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, Stelios Piperidis, and Takenobu Tokunaga. Miyazaki, Japan: ELRA, pp. 3282–3289. https://aclanthology.org/L18-1519.pdf.

Rehm, Georg, Katrin Marheinecke, Stefanie Hegele, Stelios Piperidis, Kalina Bontcheva, Jan Hajic, Khalid Choukri, Andrejs Vasiļjevs, Gerhard Backfried, Christoph Prinz, José Manuel Gómez Pérez, Luc Meertens, Paul Lukowicz, Josef van Genabith, Andrea Lösch, Philipp Slusallek, Morten Irgens, Patrick Gatellier, Joachim Köhler, Laure Le Bars, Dimitra Anastasiou, Albina Auksoriūtė, Núria Bel, António Branco, Gerhard Budin, Walter Daelemans, Koenraad De Smedt, Radovan Garabík, Maria Gavriilidou, Dagmar Gromann, Svetla Koeva, Simon Krek, Cvetana Krstev, Krister Lindén, Bernardo Magnini, Jan Odijk, Maciej Ogrodniczuk, Eiríkur Rögnvaldsson, Mike Rosner, Bolette Pedersen, Inguna Skadina, Marko Tadić, Dan Tufiș, Tamás Váradi, Kadri Vider, Andy Way, and François Yvon (2020b). "The European Language Technology Landscape in 2020: Language-Centric and Human-Centric AI for Cross-Cultural Communication in Multilingual Europe". In: *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)*. Ed. by Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Christopher Cieri, Khalid Choukri, Thierry Declerck, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis. Marseille, France: ELRA, pp. 3315–3325. https://www.aclweb.org/anthology/2020.lrec-1.407/.

Rehm, Georg, Stelios Piperidis, Kalina Bontcheva, Jan Hajic, Victoria Arranz, Andrejs Vasiļjevs, Gerhard Backfried, José Manuel Gómez Pérez, Ulrich Germann, Rémi Calizzano, Nils Feldhus, Stefanie Hegele, Florian Kintzel, Katrin Marheinecke, Julian Moreno-Schneider, Dimitris Galanis, Penny Labropoulou, Miltos Deligiannis, Katerina Gkirtzou, Athanasia Kolovou, Dimitris Gkoumas, Leon Voukoutis, Ian Roberts, Jana Hamrlová, Dusan Varis, Lukáš Kačena, Khalid Choukri, Valérie Mapelli, Mickaël Rigault, Jūlija Meļņika, Miro Janosik, Katja Prinz, Andres

Garcia-Silva, Cristian Berrio, Ondrej Klejch, and Steve Renals (2021). "European Language Grid: A Joint Platform for the European Language Technology Community". In: *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations (EACL 2021)*. Kyiv, Ukraine: ACL, pp. 221–230. https://www.aclweb.org/anthology/2021.eacl-demos.26.pdf.

Rehm, Georg and Hans Uszkoreit, eds. (2012). *META-NET White Paper Series: Europe's Languages in the Digital Age*. 32 volumes on 31 European languages. Heidelberg etc.: Springer.

Rehm, Georg and Hans Uszkoreit, eds. (2013). *The META-NET Strategic Research Agenda for Multilingual Europe 2020*. Heidelberg etc.: Springer. http://www.meta-net.eu/vision/reports/meta-net-sra-version_1.0.pdf.

Rehm, Georg, Hans Uszkoreit, Sophia Ananiadou, Núria Bel, Audronė Bielevičienė, Lars Borin, António Branco, Gerhard Budin, Nicoletta Calzolari, Walter Daelemans, Radovan Garabík, Marko Grobelnik, Carmen García-Mateo, Josef van Genabith, Jan Hajič, Inma Hernáez, John Judge, Svetla Koeva, Simon Krek, Cvetana Krstev, Krister Lindén, Bernardo Magnini, Joseph Mariani, John McNaught, Maite Melero, Monica Monachini, Asunción Moreno, Jan Odjik, Maciej Ogrodniczuk, Piotr Pęzik, Stelios Piperidis, Adam Przepiórkowski, Eiríkur Rögnvaldsson, Mike Rosner, Bolette Sandford Pedersen, Inguna Skadiņa, Koenraad De Smedt, Marko Tadić, Paul Thompson, Dan Tufiş, Tamás Váradi, Andrejs Vasiļjevs, Kadri Vider, and Jolanta Zabarskaite (2016). "The Strategic Impact of META-NET on the Regional, National and International Level". In: *Language Resources and Evaluation* 50.2, pp. 351–374. DOI: 10.1007/s10579-015-9333-4. http://link.springer.com/article/10.1007/s10579-015-9333-4.

Rehm, Georg, Hans Uszkoreit, Ido Dagan, Vartkes Goetcherian, Mehmet Ugur Dogan, Coskun Mermer, Tamás Váradi, Sabine Kirchmeier-Andersen, Gerhard Stickel, Meirion Prys Jones, Stefan Oeter, and Sigve Gramstad (2014). "An Update and Extension of the META-NET Study "Europe's Languages in the Digital Age"". In: *Proceedings of the Workshop on Collaboration and Computing for Under-Resourced Languages in the Linked Open Data Era (CCURL 2014)*. Ed. by Laurette Pretorius, Claudia Soria, and Paola Baroni. Reykjavik, Iceland, pp. 30–37. http://georg-re.hm/pdf/CCURL-2014-META-NET.pdf.

Soria, Claudia, Nicoletta Calzolari, Monica Monachini, Valeria Quochi, Núria Bel, Khalid Choukri, Joseph Mariani, Jan Odijk, and Stelios Piperidis (2014). "The language resource Strategic Agenda: the FLaReNet synthesis of community recommendations". In: *Language Resources and Evaluation* 48, pp. 753–775. https://doi.org/10.1007/s10579-014-9279-y.

STOA (2018). *Language equality in the digital age – Towards a Human Language Project*. STOA study (PE 598.621), IP/G/STOA/FWC/2013-001/Lot4/C2. https://data.europa.eu/doi/10.2861/136527.

Zhang, Daniel, Saurabh Mishra, Erik Brynjolfsson, John Etchemendy, Deep Ganguli, Barbara Grosz, Terah Lyons, James Manyika, Juan Carlos Niebles, Michael Sellitto, Yoav Shoham, Jack Clark, and Raymond Perrault (2021). "The AI Index 2021 Annual Report". In: DOI: 10.48550/ARXIV.2103.06312. https://arxiv.org/abs/2103.06312.