EUROPEAN LANGUAGE EQUALITY

D2.18

Report on the state of Language Technology in 2030

Authors	Andy Way, Georg Rehm, Jane Dunne, Maria Giagkou, José Manuel Gomez-Perez, Jan Hajič, Stefanie Hegele, Martin Kaltenböck, Teresa Lynn, Katrin Marheinecke, Natalia Resende, Inguna Skadiņa, Marcin Skowron, Tea Vojtěchová, Annika Grützner-Zahn
Dissemination level	Public
Date	30-04-2022

About this document

Project	European Language Equality (ELE)
Grant agreement no.	LC-01641480 – 101018166 ELE
Coordinator	Prof. Dr. Andy Way (DCU)
Co-coordinator	Prof. Dr. Georg Rehm (DFKI)
Start date, duration	01-01-2021, 18 months
Deliverable number	D2.18
Deliverable title	Report on the state of Language Technology in 2030
Type	Report
Number of pages	56
Status and version	Final
Dissemination level	Public
Date of delivery	Contractual: 30-04-2022 – Actual: 30-04-2022
Work package	WP2: European Language Equality – The Future Situation in 2030
Task	Task 2.3 Science – Technology – Society: Language Technology in 2030
Authors	Andy Way, Georg Rehm, Jane Dunne, Maria Giagkou, José Manuel Gomez-Perez, Jan Hajič, Stefanie Hegele, Martin Kaltenböck, Teresa Lynn, Katrin Marheinecke, Natalia Resende, Inguna Skadiņa, Marcin Skowron, Tea Vojtěchová, Annika Grützner-Zahn
Reviewers	Federico Gaspari, Gorka Azkune
EC project officers	Susan Fraser, Miklos Druskoczi
Contact	European Language Equality (ELE) ADAPT Centre, Dublin City University Glasnevin, Dublin 9, Ireland Prof. Dr. Andy Way – andy.way@adaptcentre.ie European Language Equality (ELE) DFKI GmbH Alt-Moabit 91c, 10559 Berlin, Germany Prof. Dr. Georg Rehm – georg.rehm@dfki.de http://www.european-language-equality.eu © 2022 ELE Consortium

Consortium

1	Dublin City University (Coordinator)	DCU	IE
2	Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (Co-coordinator)	DFKI	DE
3	Univerzita Karlova (Charles University)	CUNI	CZ
4	Athina-Erevnitiko Kentro Kainotomias Stis Technologies Tis Pliroforias, Ton Epikoinonion Kai Tis Gnosis	ILSP	GR
5	Universidad Del Pais Vasco/ Euskal Herriko Unibertsitatea (University of the Basque Country)	UPV/EHU	ES
6	CROSSLANG NV	CRSLNG	BE
7	European Federation of National Institutes for Language	EFNIL	LU
8	Réseau européen pour l'égalité des langues (European Language Equality Network)	ELEN	FR
9	European Civil Society Platform for Multilingualism	ECSPM	DK
10	CLARIN ERIC – Common Language Resources and Technology Infrastructure as a European Research Infrastructure Consortium	CLARIN	NL
11	Universiteit Leiden (University of Leiden)	ULEI	NL
12	Eurescom (European Institute for Research and Strategic Studies in Telecommunications GmbH)	ERSCM	DE
13	Stichting LIBER (Association of European Research Libraries)	LIBER	NL
14	Wikimedia Deutschland (Gesellschaft zur Förderung freien Wissens e. V.)	WMD	DE
15	Tilde SIA	TILDE	LV
16	Evaluations and Language Resources Distribution Agency	ELDA	FR
17	Expert System Iberia SL	EXPSYS	ES
18	HENSOLDT Analytics GmbH	HENS	AT
19	Xcelerator Machine Translations Ltd. (KantanMT)	KNTN	IE
20	PANGEANIC-B. I. Europa SLU	PAN	ES
21	Semantic Web Company GmbH	SWC	AT
22	SIRMA AI EAD (Ontotext)	ONTO	BG
23	SAP SE	SAP	DE
24	Universität Wien (University of Vienna)	UVIE	AT
25	Universiteit Antwerpen (University of Antwerp)	UANTW	BE
26	Institute for Bulgarian Language "Prof. Lyubomir Andreychin"	IBL	BG
27	Sveučilište u Zagrebu Filozofski fakultet (Univ. of Zagreb, Faculty of Hum. and Social Sciences)	FFZG	HR
28	Københavns Universitet (University of Copenhagen)	UCPH	DK
29	Tartu Ulikool (University of Tartu)	UTART	EE
30	Helsingin Yliopisto (University of Helsinki)	UHEL	FI
31	Centre National de la Recherche Scientifique	CNRS	FR
32	Nyelvtudományi Kutatóközpont (Research Institute for Linguistics)	NYTK	HU
33	Stofnun Árna Magnússonar í íslenskum fræðum SAM (Árni Magnússon Inst. for Icelandic Studies)	SAM	IS
34	Fondazione Bruno Kessler	FBK	IT
35	Latvijas Universitātes Matemātikas un Informātikas institūts (Institute of Mathematics and Computer Science, University of Latvia)	IMCS	LV
36	Lietuvių Kalbos Institutas (Institute of the Lithuanian Language)	LKI	LT
37	Luxembourg Institute of Science and Technology	LIST	LU
38	Università ta Malta (University of Malta)	UM	MT
39	Stichting Instituut voor de Nederlandse Taal (Dutch Language Institute)	INT	NL
40	Språkrådet (Language Council of Norway)	LCNOR	NO
41	Instytut Podstaw Informatyki Polskiej Akademii Nauk (Polish Academy of Sciences)	IPIPAN	PL
42	Universidade de Lisboa, Faculdade de Ciências (University of Lisbon, Faculty of Science)	FCULisbon	PT
43	Institutul de Cercetări Pentru Inteligență Artificială (Romanian Academy)	ICIA	RO
44	University of Cyprus, French and European Studies	UCY	CY
45	Jazykovedný ústav Ľudovíta Štúra Slovenskej akadémie vied (Slovak Academy of Sciences)	JULS	SK
46	Institut Jožef Stefan (Jozef Stefan Institute)	JSI	SI
47	Centro Nacional de Supercomputación (Barcelona Supercomputing Center)	BSC	ES
48	Kungliga Tekniska högskolan (Royal Institute of Technology)	KTH	SE
49	Universität Zürich (University of Zurich)	UZH	CH
50	University of Sheffield	USFD	UK
51	Universidad de Vigo (University of Vigo)	UVIGO	ES
52	Bangor University	BNGR	UK

Contents

List of Tables

List of Acronyms

AI	Artificial Intelligence
ASR	Automatic Speech Recognition
ASV	Automatic Speaker Verification
B2B	Business to Business
B2C	Business to Customer
CALL	Computer Assisted Language Learning
CAT	Computer-Assisted Translation
CAs	Conversational Agents
CEF AT	Connecting Europe Facility, Automated Translation
CH	Cultural Heritage
CRACKER	Cracking the Language Barrier (EU project, 2015–2017)
DLE	Digital Language Equality
DNN	Deep-neural-network
DPP	Data Protection and Privacy
ELE	European Language Equality *(this project)*
ELE Programme	European Language Equality Programme *(the long-term, large-scale funding programme specified by the ELE project)*
ELG	European Language Grid (EU project, 2019-2022)
ELRC	European Language Resource Coordination
EOSC	European Open Science Cloud
EP	European Parliament
FAIR	Findable, Accessible, Interoperable, Reusable
GDPR	General Data Protection Regulation
GPU	Graphic Processing Unit
HCI	Human Computer Interaction (see HMI)
HPC	High-Performance Computing
IPAs	Intelligent Personal Assistants
LID	Language Identification
LR	Language Resources/Resources
LT	Language Technology/Technologies
META-NET	EU Network of Excellence to foster META
ML	Machine Learning
MT	Machine Translation
NLP	Natural Language Processing
NLU	Natural Language Understanding
PII	Personal identifiable information
SID	Speaker Identification
SOTA	State-of-the-Art
SRIA	Strategic Research and Innovation Agenda
ST	Speech Technology
STOA	Science and Technology Options Assessment

TA	Text Analytics
TTS	Text-to-speech
WER	Word-Error-Rate

Abstract

The primary objective of the ELE project is to prepare the European Language Equality Programme, in the form of a strategic research, innovation and implementation agenda (SRIA) as well as a roadmap for achieving full digital language equality (DLE) in Europe by 2030.

This deliverable presents the current situation and state of the art in Language Technology (LT). It briefly summarises the latest breakthroughs in AI and the shift to deep learning, the importance of language models, and what implications this has for the future of LT and the equal language treatment of all languages.

The current scientific goal envisioned for 2030, laid out by the ELE consortium and the European LT community, is Deep Natural Language Understanding (NLU), which remains an open research problem with necessary breakthroughs needed. However, the benefits that NLU would bring to society are immense.

Some of the current priority research themes for NLU include Machine Translation, Speech, Text Analytics, and Data and Knowledge. A very brief overview of these research areas, along with their history, challenges, and recommendations has been provided by the ELE industry partners. As a project from the community for the community, the consortium wants to ensure that all voices are heard and taken into account for the ELE SRIA and roadmap. In addition to the expert views gathered by the consortium, further insights were gained from several online surveys and expert interviews targeting LT developers and LT users and consumers. More than 450 survey responses were collected and more than 65 expert interviews were conducted. A short 3-minute survey, targeted at European citizens, to investigate how they feel about the digital support for their languages, has already generated more than 21,000 responses at the time of writing.

1 Introduction

This deliverable summarises the necessary technological and innovation advances required to achieve the ambitious goal of DLE in Europe by 2030, and possible ways of achieving it (including technology forecasting) that will be further highlighted and investigated in the Strategic Research and Innovation Agenda (SRIA) to be published in June 2022. The SRIA, conceptualised by the ELE consortium will serve as a blueprint for achieving full DLE in Europe. The current scientific goal envisioned for 2030 is Deep Natural Language Understanding (Deep NLU) which comes with various demands and issues from society and a number of necessary breakthroughs needed.

Deep NLU remains an open research problem. Current approaches have severe limitations and are not able to serve all of Europe's languages in an adequate way. However, over the last decade, the emergence of new deep learning techniques and tools has revolutionised the approach to LT-related tasks. We are gradually moving from a methodology in which a pipeline of multiple modules was the typical way to implement LT solutions, to architectures based on complex neural networks trained with vast amounts of text data. Just to name two examples, the current state-of-the-art has enabled translation without parallel corpora and the generation of full text claimed to be almost indistinguishable from human prose.

Given the speed of the development these days, forecasting the future of LT and language-centric Artificial Intelligence (AI) is a challenge. Nevertheless, while it is undeniable that the benefits to society of these anticipated developments would be immense, they also come with great expectations and demands for the future. For instance, assistive technologies such as Text-to-Speech (TTS) help those with visual and oral impairments and learning disabilities.

The current lack of suitable data for use in training and evaluating today's state-of-the-art data-driven tools leads directly to digital language inequalities. While data availability is already a general problem, this scarcity is compounded and results in more severe limitations for lesser-spoken European languages. The European data economy relies on the availability, the interoperability and the provision of (unstructured, semi-structured and structured) data as a basis for further innovation and exponential development of technologies.

To counteract this, steps have been taken recently by the research community with respect to cultivating a culture of open data and data sharing. The EU Coordinated Plan on Artificial Intelligence states that further developments in AI require a well-functioning data ecosystem built on trust, data availability and infrastructure. In addition, the elimination of biases and the consideration of fairness and ethical aspects that are relevant to machine (and deep) learning models are important factors that need to be taken into account.

To better assess the current state of the LT landscape and to outline and define the steps necessary to achieve the ambitious goal of Deep NLU by 2030, the ELE industry partners generated, in various focus groups, four technology reports to illustrate the demands, wishes and visions of the European industry in a structured way. These deep dives have been compiled for the fields of Machine Translation (Technology deep dive Machine Translation, Bērziņš et al., 2022), Speech (Technology deep dive Speech Technologies, Backfried et al., 2022), Text Analytics (Technology deep dive Text Analytics and Natural Language Understandings, Gomez-Perez et al., 2022) and Data and Knowledge (Technology deep dive Data, Kaltenboeck et al., 2022). They offer in-depth, up-to-date analyses of their areas.

The recommendations from these expert reports serve as valuable input to pave the way for DLE in 2030. However, all initiatives of the last decade (such as META-NET, CRACKER, ELG etc.) have always been designed to also build a strong community to lobby the importance of LT in Europe. Previous projects have benefited immensely from the partners' expertise and their community reach.

This Deliverable is structured as follows. In Section 2, the state of the art (Sect. 2.1) is described as collected in the WP2 deliverables in a summarised form to make this deliverable self-contained. In the remaining two subsections of Section 2, main Gaps and shortcomings (Sect. 2.2) are described as collected from all preceding deliverables to provide the starting point for the forward looking sections; to support the visions and recommendations, Section 2.3 describes the contributions, demands and issues related to LT and their use in society at large. Section 3 presents the vision of the various stakeholders who contributed to the surveys and interviews for the LT landscape in 2030. This is followed by Section 4 that formulates the recommendations, supported by the three types of ELE surveys and their results. Section 5 summarises the report and concludes with the key points. The Appendix contains the results of the EU Citizen survey not presented in detail previously.

With all the valuable insights collected during the project by its large and well-connected consortium up to this point, a well-informed and comprehensive SRIA and roadmap will be crafted in the remainder of the ELE project to support future efforts towards achieving full DLE for all languages of Europe by 2030.

2 Language Technology: State of the Art and Current Situation

2.1 State of the Art

2.1.1 Move to Deep Learning

In recent years, the LT community has witnessed and contributed to the emergence of disruptive new deep learning techniques and tools that are revolutionising the approach to LT-related tasks. We are gradually moving from a methodology in which a pipeline of multiple modules was the typical way to implement LT solutions, to architectures based on complex neural networks trained with vast amounts of text data. For instance, the *AI Index Report 2021*[1] highlights the rapid progress in NLP, vision and robotics thanks to deep learning and deep reinforcement learning techniques. In fact, the *Artificial Intelligence: A European Perspective* report[2] establishes that the success in these areas of AI has been possible because of the confluence of four different research trends: 1) mature deep neural network technology, 2) large amounts of data (and for NLP processing large and diverse multilingual textual data), 3) increase in High Performance Computing (HPC) power in the form of Graphic Processing Units (GPUs), and 4) application of simple but effective self-learning approaches (Goodfellow et al., 2016; Devlin et al., 2019; Liu et al., 2020; Torfi et al., 2020; Wolf et al., 2020).

As a result, various IT enterprises in Europe and elsewhere have started deploying large pretrained neural language models in production. Compared to the previous state of the art, the results are so good that systems are claimed to obtain human-level performance in laboratory benchmarks when testing some difficult English language understanding tasks. For instance, DeepMind's Gopher achieved scores that suggest its comprehension skills were equivalent to that of an average high school student (Rae et al., 2021). Interestingly, large language models still perform poorly in logical and mathematical reasoning.

2.1.2 Forecasting the Future of LT

Forecasting the future of LT and language-centric AI is a challenge. Five years ago, hardly anyone would have predicted the recent breakthroughs that have resulted in systems that can translate without parallel corpora (Artetxe et al., 2019), create image captions (Hossain et al., 2019), generate full text claimed to be almost indistinguishable from human prose (Brown et al., 2020), generate theatre play script (Rosa et al., 2020), create pictures from textual descriptions (Ramesh et al., 2021, 2022) or explain jokes[3] (Chowdhery et al., 2022).[4]

2.1.3 Large Language Models

It is, however, safe to predict that even more advances will be achieved by using pretrained language models. For instance, GPT-3 (Brown et al., 2020), one of the largest dense language models, can be fine-tuned for an excellent performance on specific, narrow tasks with very few examples. GPT-3 has 175 billion parameters and was trained on 570 gigabytes of text, with a cost estimated at more than four million USD.[5] In comparison, its predecessor, GPT-2, was over 100 times smaller, at 1.5 billion parameters. This increase in scale leads to surprising behaviour: GPT-3 is able to perform tasks it was not explicitly trained on with zero to few training examples (referred to as zero-shot and few-shot learning, respectively). This behaviour was mostly absent in the much smaller GPT-2 model. Furthermore, for some tasks (but not all), GPT-3 outperforms state-of-the-art models explicitly trained to solve those tasks with far more training examples.

It is impressive that a single model can achieve a state-of-the-art or close to a state-of-the-art performance in limited training data regimes. Most models developed until now have been designed for a single task, and can thus be evaluated effectively by a single metric. Moreover, OpenAI has trained language models that are much better at following user intentions than GPT-3. The InstructGPT[6] models are trained with humans in the loop. The

1 https://aiindex.stanford.edu/report/
2 https://ec.europa.eu/jrc/en/publication/artificial-intelligence-european-perspective
3 https://ai.googleblog.com/2022/04/pathways-language-model-palm-scaling-to.html
4 https://openai.com/blog/dall-e/
5 https://lambdalabs.com/blog/demystifying-gpt-3/
6 https://openai.com/blog/instruction-following/#{}guide

team claims to have made them more truthful and less toxic by using techniques developed through alignment research.

Making larger models is not the only way for improving its performance. For instance, Megatron-Turing NLG,[7] built by Nvidia and Microsoft, held the title of the largest dense neural network at 530B parameters – already 3x larger than GPT-3 – until very recently (Google's PaLM[8] holds the title now at 540B). But remarkably, some smaller models that came after MT-NLG reached higher performance levels. Smaller models, like Gopher (280B), or Chinchilla[9] (70B) – barely a fraction of its size – are way better than MT-NLG across tasks. It seems that current large language models are "significantly undertrained".

Combining large language models with symbolic approaches (knowledge bases, knowledge graphs), which are often used in large enterprises because they can be easily edited by human experts, is a non-trivial challenge. Techniques for controlling and steering such outputs to better align with human values are nascent but promising.

Such language models have an unusually large number of uses, from chatbots to summarisation, from computer code generation to search or translation. Future users are likely to discover more applications, and use existing technologies positively (such as knowledge acquisition from electronic health records) and negatively (such as generating deep fakes), making it difficult to identify and forecast their impact on society. As argued by Bender et al. (2021), it is important to understand the limitations of large pretrained language models, which they call "stochastic parrots" and put their success in context.

Indeed, today we find ourselves in the midst of a significant paradigm shift in LT and language-centric AI. This revolution has brought noteworthy advances to the field along with the promise of substantial breakthroughs in the coming years.

2.1.4 Equal Language Treatment

We believe that this unprecedented time of significant technological transition represents a unique opportunity to redress the balance, and that now is the moment to create a level playing field for all European languages in the digital realm. There are ample reasons for optimism. Although there is more work that can and must be done, Europe's leading language resources, repositories, platforms, libraries, models and benchmarks have the potential to make significant progress. Recent research in the field has considered the implementation of cross-lingual transfer learning and multilingual language models for low-resource languages, an example of how the state of the art in LT could benefit from better digital support for low-resource languages.

2.2 Main Gaps and Shortcomings

Here we discuss the main gaps and shortcomings facing LTs across 4 main axes – Data, Technologies, Benchmarking and Expertise. Within these contexts, there is much overlap of shared issues across all four areas of focus in the technology deep dives.

[7] https://developer.nvidia.com/blog/using-deepspeed-and-megatron-to-train-megatron-turing-nlg-530b-the-worlds-largest-and-most-powerful-generative-language-model/

[8] https://ai.googleblog.com/2022/04/pathways-language-model-palm-scaling-to.html

[9] https://towardsdatascience.com/a-new-ai-trend-chinchilla-70b-greatly-outperforms-gpt-3-175b-and-gopher-280b-408b9b4510

2.2.1 Data

Data Availability

The availability of suitable data for use in both training and evaluating today's state-of-the-art data-driven tools is crucial. However, the current lack of parity in such resources for different languages translates directly to digital language inequalities.

The type of data required for TA tools can vary according to the task at hand. For example, when building large transformer-based language models, current systems can be built upon raw (unlabelled) text (e. g. Wikipedia, books, etc.). However, more sophisticated tasks such as named entity recognition, syntactic parsing, sentiment analysis, etc., require training and test data to be labelled. **Labelling data can be a time-intensive task** that often requires skilled domain expertise, which is a costly overhead for both the research and industry communities. The lack of in-house expertise to create labelled datasets has increased the demand for third-party data providers. In addition, online platforms such as Amazon's Mechanical Turk are also popular for crowd-sourcing campaigns for (trivial, non-expert) labelling tasks. These online platforms, however, are not useful when dealing with complex labelling tasks or for regional or lesser-spoken languages.

With respect to MT, as translation data management and file standards improve, parallel data is becoming more and more available. However, there is much **untapped potential across public sectors**. In fact, Berzins et al. (2019) report on the difficulties experienced across a number of EU member states in accessing public sector language data – due to the lack of awareness or implementation of the EU Open Data Directive. They also found that lack of specialised user training and negative dispositions towards Computer-Assisted Translation (CAT) tools (along with their high costs) were blocking factors for translators to embrace CAT tools, hindering the creation of appropriate translation files (e. g. TMX) and language data sharing.

Obtaining training data for speaker recognition and language identification presents a different set of challenges. In the case of speaker identification (SID) and language identification (LID), the situation is more favourable since the only annotation needed is the identity of the speaker or language. However, it is crucial that the training data for SID and LID contain **many recordings** of the same speaker or of the same language. On the other hand, it is preferable to have as **many (different) speakers** as possible in ASR training data. While there has been much progress in collecting data from videos online, progress on telephony data is still limited by privacy concerns and lack of data.

The **diversity** of contexts and speakers represented by popular ASR benchmarks for read speech[10] and spontaneous speech[11] is limited. Recent works attempt to address this problem by introducing benchmarks that mimic real-world settings, with the goal of detecting model biases and flaws (Riviere et al., 2021). Contemporary models often reveal significant performance differences by accent, and much greater differences depending on the socio-economic background of the speakers; which also highlights the need to develop better and more robust conversational language models.

As we have seen, data availability is already a general problem, but when it comes to **lesser-spoken European languages with less digital content**, this scarcity is compounded and results in significant implications. For a few languages with high commercial interest, an abundance of training data is available. However, for many (the majority of) European languages, this is not the case and only corpora which are minuscule in comparison to English are available, often exclusively in general-purpose domains. This of course has a knock-on effect on performance quality of the relevant technologies and in the prospects of developing novel LTs for these languages.

[10] Librispeech (Panayotov et al., 2015; Garnerin et al., 2021)
[11] See Tüske et al. (2021)

That said, most of today's TA solutions are language-specific. If language-agnostic tools are not a realistic goal, more innovation and investment are required in making this language adaptation process easier and less of a roadblock for LT providers, their customers, governments and the wider linguistically diverse public. This broadening of linguistic coverage cannot rely solely on being market-driven – which is the main reason why relatively lesser spoken languages are being left behind. Likewise in terms of ST, the unbalanced availability and quality of resources (e. g. data-sets, annotations, models) strongly impact the performance of ST for different groups of languages. In extreme cases, selected functionalities and/or support for minor languages may not be available at all. In addition to the support of a language per se, language varieties, dialects or accents (e. g. non native) may not be supported or only supported on very limited levels.

Studies involving evaluations of English-only language data and applications are no longer sufficient, as there is a need to extend well-established evaluation protocols and benchmarking exercises to as many languages as possible. In this regard, the so-called Bender rule (Bender, 2011) originally called upon researchers to "state the name of the language that is being studied, even if it's English. Acknowledging that we are working on a particular language foregrounds the possibility that the techniques may in fact be language-specific, which in turn may open up interesting and promising opportunities to investigate the portability of some techniques to other languages. Conversely, neglecting to state that the particular data used were in, say, English, gives [a] false veneer of language-independence to the work."

Domain coverage is also an important consideration. LT is a type of horizontal technology, which cuts across nearly every domain, from health and pharmacy, publishing, broadcasting and the legal domain, to construction, finance and insurance and the public administration – to just name a few core areas. While general language data may be useful for developing a language model, domain-specific language data (e. g., medical, legal, user-generated content, etc.) is needed to ensure sufficient coverage of certain terminology and phrasing for certain applications.

In addition, databases with more expressive and spontaneous recordings are required to build TTS systems suitable for more **emotion-demanding** applications like audiobook reading, movie dubbing and human-computer interaction that aims to be similar to interactions between humans. Moreover, the vast majority of datasets correspond to adult voices and there is a lack of data to generate child and elderly voices, which involves ethical issues with regard to eliciting and using spoken data provided by potentially vulnerable speakers. Adding emotions and affections into the applications and tools for human-machine interaction, recognising intent and taking into account a broad variety of contexts holds the potential to turn these interactions into truly human-like experiences. The components related to **emotional understanding and empathy**, while relevant to all Intelligent Personal Assistants (IPAs) and Conversational Agents (CAs), are especially relevant for systems functioning in social domains, such as healthcare, education, and customer service.

Digitisation for at least one kind of content still has a long way to go: **educational material**. Much educational material in several languages is still widely published on paper. In many countries, there is a small but focused industry with a long tradition of creating books and related material conforming to country and region-specific requirements of educational content. While digitized content is available in some cases, it is essentially replacing analog recordings, e. g. for foreign language teaching, or recorded lectures as experienced during the pandemic. Current digital technology, LT and AI as much more to offer, yet this is still to be picked up by the key players in the educational industry, and vice versa – LT is not well informed about the needs of the educational sector. This has then negative consequences on data availability for LT in education, and as a consequence, the possibly groundbreaking applications in this area cannot be developed.

Data Accessibility

Steps have been taken recently in the research community with respect to cultivating a culture of **open data and data sharing**. Many top-tier publications require the release of datasets (where possible) in order to facilitate reproducibility of studies. Additionally most shared tasks (benchmark or evaluation campaigns) require a release of their specifically designed datasets for use by the wider research community (Escartín et al., 2021). These practices are only helpful, however, when related to datasets that are not restricted by copyright, licensing agreements or privacy regulations.

Enterprise data, e. g., tends to be locked in regulatory and corporate silos. As enterprise data is by nature confidential and companies need to respect data protection regulations, the barriers for making data available are high. Research and solutions for language technologies that address problems of business and social relevance is therefore underdeveloped.

When it comes to MT training data, translation memories and terminology data are often licensed for **non-commercial use only**. When commercial licences do exist, their prices are often prohibitively high for many users and developers. This acts as a major barrier to SMEs developing MT applications, especially when there is a limited amount of data available in the language pairs and/or domains of interest.

With regards to copyrighted content, **copyright laws** pose a barrier in Europe. While copyright law is subject to fair-use exceptions in countries such as the US, European law is far less flexible. Many European laws severely restrict the use of parts of copyrighted works for purposes such as data mining. In the context of speech technologies, it is found that in terms of copyright, rules in Europe are more restrictive than in other economic regions and countries such as the United States. For example, difficulties faced in accessing closed captions from TV broadcasts or subtitles from a copyrighted film to train and evaluate ST models.

The EU Coordinated Plan on Artificial Intelligence[12] correctly states that "Further developments in AI require a well-functioning data ecosystem built on trust, data availability and infrastructure." But it underestimates the effect that one of its cornerstones has had on data collection in the language AI field – The General Data Protection Regulation (GDPR). Since unconstrained, unstructured text can by its very nature often include personal data, data protection and privacy (DPP) policies can put limits on the type of data that can be made available for the development of all LT technologies. As such the GDPR may have an adverse effect on a large part of the European LT industry. Additionally, the principles of DPP and legal provisions such as GDPR stipulate that data should only be used for a-priori defined narrow purposes and that these purposes must be made transparent to the data subject upfront. This proves problematic, especially when dealing with induced models or datasets from online sources that have been reused without the consent of website owners or individual contributors, that would be highly impractical to trace in most situations. Moreover, non-European AI companies have been able to continue to operate without GDPR restrictions, which has gained them a considerable competitive advantage over EU companies.

As the main issue related to GDPR restricted data concerns Personal identifiable information (PII), steps have been taken recently towards developing tools that can anonymise language data in an attempt to overcome these barriers.[13] However, the task of anonymisation is difficult and does not always work with sufficient precision and reliability. Any text anonymisation in practice has to accept a potential residual risk of DPP non-compliance. Special usage rights have been called for to help advance NLP, particularly in domains where PII is prevalent in datasets (similar to the exemptions granted in the field of medical research under very specific circumstances and subject to approval of the relevant authorities).

---

[12] Coordinated Plan on Artificial Intelligence, COM(2018) 795 final https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018DC0795

[13] For example, the CEF-funded MAPA project set out to develop a toolkit for effective and reliable anonymisation of texts in the medical and legal fields for all official EU languages: https://mapa-project.eu

Fairness and Ethical Considerations

The development, application and adoption of LTs are also connected to a range of issues relating to fairness, biases and ethical aspects that need to be accounted for.

Unfortunately, machine (and deep) learning models are notoriously **sensitive to bias and noise** within datasets. The dominant data-driven approach to speech and language processing and the quest for accuracy have yielded both opaque tools that are hard to interpret, and biased tools that perpetuate social stereotypes that exist within datasets on a gender, racial and ethnic basis (e. g., Vanmassenhove et al., 2019; Sheng et al., 2021). These dataset biases replicate regrettable patterns of socio-economic domination and exclusion that are conveyed through language, since these biases are present in the training data and are then amplified by models which tend to choose more frequent patterns and discard rare ones. Furthermore, they can generate unpredictable and factually inaccurate text or even recreate private information.[14] One way to achieve this is the examination of training data, identifying biased parts or gaps, and enriching the data by providing alternatives, or by replacing them altogether. Modifying models could reduce biases, too, for example by introducing weights for probabilities of words related to bias.

Voice assistants frequently utilise **female voices**. Some of them offer the possibility of using male voices, but the default voice is usually female. This fact has been extensively criticised as it can contribute to the outdated view of women as the gender that must help and take care of others. Moreover, nowadays the generation of gender-neutral voices is gaining importance, as many people do not identify themselves with the classic binary genders.

Similar to gender-related biases, **race-related biases** may also be present in many kinds of LT models. Due to the fact that models depend on the amount and composition of training data, ethically-concerning aspects of language and language use that is present in these data may also be present in the resulting models. Systems capable of self-learning may adapt into directions completely unplanned and undesired by the developers or be gamed (attacked) by users into doing so.[15] Due to these inherent conditions, systems may subsequently perform at different levels of accuracy for particular sections of the population. Furthermore, disabilities related to language production may not be accounted for and **exclude sections of the population** from using ST systems at all. Various ethnic groups may however be under-represented in the training data and thus less accurately recognised. Biased tools therefore have a direct impact in society as a whole and can have a negative impact on marginalised populations (Sheng et al., 2021).

2.2.2 Technologies

Technology Capabilities

The **paradigm shift** to neural MT systems, neural language models in TA[16] and end-to-end ST systems means that current state-of-the-art LT research and development is based on access to **huge, and previously unthinkable, amounts of data and processing power**. Access to hardware, experts, and involvement in research have also shifted in such a way that elite universities and large firms have an advantage due to their ease of access to such resources (Ahmed and Wahed, 2020). Thus, it is no surprise that the companies with the largest pools of data and the most extensive infrastructure are now the leading actors in their respective fields, leaving only niche markets and domains to smaller, but highly specialised players.

[14] https://ai.googleblog.com/2020/12/privacy-considerations-in-large.html

[15] After Microsoft's release of its chatbot Tay in 2016, the chatbot began to post racist, sexually-charged, inflammatory and offensive tweets prompting Microsoft to shut down the service again within 16 hours of its launch (https://en.wikipedia.org/wiki/Tay_(bot))

[16] Also known as pre-trained language models (Han et al., 2021)

According to the ELE report on existing strategic documents and projects in LT/AI, there is a lack of necessary resources (experts, High Performance Computing (HPC) capabilities, etc.) in Europe, compared to large U.S. and Chinese IT corporations (e. g., Google, OpenAI, Facebook, Baidu, etc.) that lead the development of new LT systems. The report also highlights an uneven distribution of resources, including scientists, experts, computing facilities, and IT companies, across countries, regions and languages (Aldabe et al., 2021b).

While most research focuses on a single user's interactions, speech technologies embodied in virtual assistants are becoming increasingly popular in social spaces. This highlights a gap in our understanding of the opportunities and constraints unique to **multiple user scenarios**. These include detecting if users are addressing the system or other participants. For example, speaker diarisation (see Park et al., 2022, for a review of recent advances in speaker diarisation with deep learning methods), understanding aspects of social dynamics, and finding interaction barriers are some of the factors that restrict the usefulness of voice interfaces in group settings.

In ASR, the focus on rather constrained conditions has left gaps in more **diverse settings** such as: distant speech recognition instead of single microphones; noisy environments; accented speech, non-native speech, dialectal speech and sociolinguistic factors affecting speech; spontaneous, unplanned speech; emotional speech (including speech during stressful or dangerous situations) and connected aspects concerning sentiments expressed (empathy); the integration of speech technologies into collaborative environments, multiple, simultaneous speakers engaged in discussions; as well as the integration of technologies addressing paralinguistic aspects. All of these issues warrant future attention and research.

Even more important is the lack of consideration for those users with **disabilities**, another community often marginalised through advances in technology. For example, while state-of-the-art ASR systems achieve great accuracy on typical speech, they perform poorly on disordered speech and other atypical speech patterns. While on-device personalisation of ASR recently showed promising, preliminary results in a home automation domain for users with disordered speech (Tomanek et al., 2021), more research is required to further increase the ASR performance for these groups of users and provide support for open conversations with longer phrases. Text-based interactive tools or applications, such as computer-assisted language learning apps, also need to consider those students with learning disabilities such as dyslexia or visual impairments. TTS (e. g. for screen readers) is not employed widely enough with these users in mind.

Interoperability ensures the seamless interplay of different (natural) LT systems with respect to interfaces and data. It is often connected with the requirement of related standards in the field. Interoperability allows easy data integration of heterogeneous data from different sources, which is a crucial task for adequate LT systems that ingest and make use of data from relevant sources.

There has been a significant move towards open-source tooling and ease-of-sharing for LTs (e. g. Github[17] and Hugging Face[18]). As a result, many NLU system components are available for a 'plug-and-play' interaction with complex pipelines during software development. This has facilitated interoperability in academic or open-source research areas. However, at an enterprise level, in the absence of standards, interoperability can prove to be more challenging with respect to proprietary software or data formats. Accordingly, technical solutions need to be built with investment protection and interoperability in mind. Otherwise, risks such as vendor lock-in are likely to surface.

Additionally, official standards are important ingredients for protecting investments since they facilitate interoperability and reuse. A special dimension related to standards concerns **conformance**. "Conformance is the fulfillment of specified requirements by a product, pro-

[17] https://github.com
[18] https://huggingface.co

cess, or service".[19] In the context of regulated industries, certification – the assignment of a label, based on transparent testing, and compliance with conformance criteria – may need to be considered.

Multimodal Tools

Different modalities can be combined to provide complementary information that helps to convey information more comprehensively and effectively (Palanque and Paternò, 2012). This convergence across modalities requires synergies from AI research fields that until now have been conducted individually, such as NLP, ASR and computer vision. For example, TA is not only the process of analysing a source text sentence by sentence. Rather, key pieces of contextual information (i. e. pragmatics) such as the author, intended audience, societal factors and the purpose of communication – the interactive and communicative context – need to also be considered. As such, there is much scope for improving contextualised and personalised analytics. Conversely, an intimate **interaction of ASR, SID and TTS with downstream NLP and NLU technologies** is required to allow the correct interpretation of input so that recognition, meaning and output can be produced in a natural and correct manner.

In addition, the ever-growing demand for translation of audiovisual content over multiple delivery platforms has sparked interest in the development of **MT-centric TTS and STT applications**. The New European Media Strategic Research Agenda states that in the future AI will be used to **translate speech to subtitles, text to Sign Language and Sign Language to text** (New European Media Initiative, 2020). Likewise, ST predominantly addresses the modality of using voice for human computer interaction (HCI). This encompasses linguistic as well as paralinguistic elements and may extend to sign language. The inclusion of gestures, facial expression, emotions or haptics as well as the generation of multimodal outputs reflecting these elements could result in a much richer and more natural user experience and lead to wider adoption and acceptance of ST.

Explainable AI

Interpretability is a major concern in modern AI and LT research. Data-driven approaches such as Machine Learning (and to a greater degree, Deep Learning) have been criticised for their **'blackbox' nature**. That is to say, when language data is converted to numeric or opaque vector representations in order to enable modelling or pattern inducing, it becomes difficult to (i) assess why a model is under-performing and (ii) overtly specify processing expectations of a system. The EU Coordinated Plan on Artificial Intelligence[20] recognises this problem and advocates the need for trustworthy AI, mainly from the perspective of the end-user. The **lack of transparency** makes it difficult to build trust between users and system predictions, having negative consequences in overall technology adoption.

In cases where decisions are made based on AI model prediction, it is important that businesses can assess these models' level of accuracy, fairness and transparency. As such, a priority for many businesses and organisations is to **build trust and confidence** in these AI models. As a result, there has been a notable increase in attention given to and demand for **Explainable AI**.

[19] https://www.w3.org/TR/qaframe-spec/#specifying-conformance

[20] Coordinated Plan on Artificial Intelligence, COM(2018) 795 final, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018DC0795

Responsible AI

Training neural MT, TA and ST engines is resource-intensive and has a heavy **carbon footprint**. One area where EU laws are perhaps too relaxed is in relation to carbon emissions in the field of AI research and development. Researchers have warned of the marginal performance gains associated with expensive compute time and non-trivial carbon emissions. An MIT study (Strubell et al., 2019b) found that training a large AI model to handle human language can lead to emissions of nearly 300,000 kilograms of carbon dioxide equivalent, about five times the emissions of the average car in the US, including its manufacture. In line with this study, Swedish researchers have forecast that data centres could account for 10% of total electricity use by 2025.[21]

Through the European Green Deal[22] and the Horizon Europe Work Programme,[23] the European Commission has committed to making "Europe the world's first climate-neutral continent by 2050". To achieve this, the economy must be transformed with the aim of climate neutrality. More efficient AI infrastructure can help in reducing the amounts of energy that are required for data storage and algorithm training.[24]

The increase in the complexity and combination of technologies and models requires a careful balance with regard to **privacy and trust**. The standard today is to store audio (voices) and text in the cloud and label them manually. Concerns have arisen regarding trust, privacy, intrusion, eaves-dropping, or the hidden collection and use of data. These concerns have been recognised by many actors but are only addressed to a limited degree. This general approach raises critical privacy concerns and it has led to market and data concentration in the hands of a few, big corporations. Dramatic improvements in speech synthesis (Székely et al., 2019), voice cloning (Vestman et al., 2020) and speaker recognition (Snyder et al., 2018) pose severe privacy and security threats to the users. Further work and investigation into these topics will be necessary commercially, academically, as well as for policy-making.

In the long run, the question will be whether any possible **breaches, leaks or scandals** involving LT will erode trust to a level that users will **no longer volunteer to provide their data for training purposes** (e. g. in ST, deep fakes may pose a particular risk). Of course, the distrust will be weighed against the commodity of using certain devices and platforms whose terms of use may simply require the user to do so.

Privacy and security also emerge as matters of utmost interest in the MT industry. Text submitted for translation may include sensitive product or customer information and clients are often reluctant to hand these details over to third-party technology provides, make them available to external post-editors and even to the MT systems, which can learn from edits made to the raw output. The partial understanding of how MT works and the unclear legal rights, obligations and consequences of misuse have clients seeking solutions backed with specific privacy and security functionalities.

2.2.3 Benchmarking

Benchmarking is the practice of establishing an evaluation reference point against which the performance of a system can be measured. Benchmarking campaigns, evaluation campaigns and shared tasks have the common objective of establishing standard datasets on

[21] Strubell et al. (2019b) recommend that time spent retraining should be reported for NLP learning models and that researchers should prioritise developing efficient models and hardware. The EU has the opportunity to be a pioneer in training and developing green LT by following and enforcing these recommendations.

[22] https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_en

[23] https://ec.europa.eu/info/research-and-innovation/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe_en

[24] See for example https://ec.europa.eu/research-and-innovation/en/horizon-magazine/ai-can-help-us-fight-climate-change-it-has-energy-problem-too

which systems can be evaluated, establishing appropriate evaluation metrics and providing 'leaderboard' reports on best-performing systems so as to identify state-of-the-art (SOTA) performance. Current benchmarking presents issues across all areas of speech and language technologies.

In academia, benchmarking is mainly used as a way to advance research (leaderboard-driven), while for industry it is a way to determine the technical or market readiness of a product. Moreover, savvy customers in this space will often set minimum accuracy scores in terms of the quality of the systems they require. With respect to TA, while metrics and benchmarks exist for various sub-fields, it is often difficult for users or buyers to determine how well their own content is or could be processed. Similarly, certain tasks are notoriously difficult to establish benchmarks for, such as information retrieval. In terms of the nature of datasets used in benchmarking, businesses require realistic data. Some evaluation datasets are also often criticised in academic shared tasks, where they are sometimes referred to as "toy" examples that are not applicable to real-world problems.

In particular, there is still a lack of agreement within the MT community on a single metric which can be used universally to assess the quality of MT engines prior to deployment. The community still relies to a large extent on one of the first automatic metrics, Bilingual Evaluation Understudy (Papineni et al., 2002), and there is a noticeable reluctance to abandon this measure despite a large body of research pointing out its drawbacks (Mathur et al., 2020; Kocmi et al., 2021). Future systems should be evaluated by new automatic metrics which represent better approximations of human judgments and also ideally abandon the dependence on single human reference translations, which is a serious limitation.

The single most frequently mentioned hindering factor for the broad adoption of speech technology is accuracy. The perceived accuracy and its exact meaning have changed dramatically – from individual words being misrecognised to intentions not correctly interpreted in complex situations, with accuracy reaching well beyond the actual accuracy of ASR only, regarding it in a more comprehensive and embedded manner. Whereas Word-Error-Rate (WER) as an evaluation measure has had its merits to measure progress in ASR (and still does so), more comprehensive approaches to measuring the impact of ASR performance on downstream tasks and actual deployments may require novel approaches. WER alone clearly does not provide the full picture when it comes to the perceived performance and usability of complete systems comprising several kinds of speech and language technologies.

Similarly, the availability of proofing tools also influences a society or community's connectedness. While speech technology is becoming more prevalent in Business to Business (B2B) and Business to Customer (B2C) interactions, much of our personal interactions with each other still rely on language technologies that facilitate written communication (e.g., emails, online social networks, instant messengers, chat rooms, etc). As this continues to be the trend, we can see clearly how, through the lack of basic technological support, a language community could not continue forging or strengthening these connections through their own language. Such scenarios inevitably leads to disconnect and possible divide.

2.2.4 Expertise Gap

A significant gap, concerning all areas of speech and language processing, is the scarcity of trained personnel and expertise, as well as the risk of losing emerging talent to innovative power-players outside of Europe (with possibilities and salaries which can generally not be matched by European players). Indeed, with respect to multimodal approaches, there is a demand for those with blended expertise. As is the case for the field of computational linguistics, such interdisciplinary fields of research require a broad amount of knowledge and expertise. As such, traditional silos of learning (e.g. third level institutions, training programmes) will need to adapt and expand. Therefore, respective educational programs in

LTs form the foundation for future European success in these areas and may hinder it if not appropriately established and strengthened.

Today, most work in the ML-driven LT ecosystem requires expert-level skills in the realm of tools related to data management, data science and NLP processing. This creates bottlenecks since it does not allow domain experts (e. g. experts in finance) to become actively involved without rather extensive tool training, and without the need for understanding the underlying technology. The 'design' of this ecosystem also causes overhead and delays since work between tool experts (e. g. data scientists) and domain experts needs to be coordinated. As such, only 1 in 10 enterprises feel they have a competent approach to mining data, which ultimately hampers AI efforts. A shortage of AI skills and risk managers' lack of familiarity with the technology increase the risk.

2.3 Contributions to Society, Demands and Issues

With the democratisation of AI and the development of accurate and smart solutions that communicate with users in natural language, AI technologies already impact business activities, society and individual users' lives. From an economic perspective, Gartner (November, 2021) forecasts the worldwide artificial intelligence (AI) software revenue to total $62.5 billion in 2022, an increase of 21.3% from 2021.[25] With respect to areas of high impact, the top five use-case categories for AI software spending, according to Gartner, in 2022 will be knowledge management, virtual assistants, autonomous vehicles, digital workplace and crowdsourced data.

Intelligent, AI-based, virtual assistants are already in demand in the digital market and use of them in the workplace is growing. Gartner (August, 2020) predicts that by 2025, 50% of knowledge workers will use a virtual assistant on a daily basis, up from 2% in 2019. For public sector and businesses this means an opportunity to use an intelligent virtual assistant technology to take care of more repetitive and auxiliary business processes. By 2030, Gartner predicts that the decision support/augmentation will be the largest type of AI accounting for 44% of business value, while agents representing 24%.[26] These predictions of course only hold for countries with lesser-spoken languages if the technology is there to support them. If not, it is evident how an economic divide will emerge, as countries with sufficient language technologies will gain advantage.

The following summarises the ways in which language technologies already play a central role in our daily lives and how they can have a positive impact on governments, businesses and consumers. Also highlighted are the negative impacts that the lack of such technology will eventually have on societies and economies that may be digitally left behind, as well as the unintended side effects of AI-driven LT.

Government Affairs and Public Services

Today, many government organisations already apply LT solutions to help them deliver efficient public services and improve governance. According to the Gartner Digital Transformation Divergence Across Government Sectors survey[27] **chatbots** are leading the way in government AI technology adoption – 26% of government respondents reported that they have already deployed them, while 59% are planning to have deployed them within the next

---

[25] https://www.gartner.com/en/newsroom/press-releases/2021-11-22-gartner-forecasts-worldwide-artificial-intelligence-software-market-to-reach-62-billion-in-2022
[26] https://www.gartner.com/en/newsroom/press-releases/2019-08-05-gartner-says-ai-augmentation-will-create-2point9-trillion-of-business-value-in-2021
[27] https://www.gartner.com/en/newsroom/press-releases/2021-10-05-gartner-says-government-organizations-are-increasing-

three years. In the case of machine learning-supported data mining – only 16% have currently deployed it with a further 69% planning to do so within the next three years.

In the case of government organisations, one of the key challenges faced is obtaining relevant information from huge volumes of unstructured text. In these cases, LT can be used to: help to **solve routine tasks** (e. g., with help of virtual assistants many common citizen information-related questions could be answered without human intervention), improve public services (e. g., through analysis of public feedback or engagement), assist process analysis (e. g., identifying potential risks, investigating or enhancing policy analysis) or even address critical government issues.

Public administrations across Europe have large translation demands, as demonstrated by the extensive translation data collection of the ELRC (Berzins et al., 2019) over the past several years. MT is therefore imperative as a support tool in such professional translation settings for ensuring such translation demands are met.

Likewise, the COVID-19 pandemic showed a clear need for a multilingual and crosslingual information sharing. In such crisis time, communities that do not adequately understand or speak the major or official national languages are easily excluded from latest information updates (e. g. availability of vaccines or specific medication). This lack of information can lead to grounds for misinformation, toxic content and bias to grow.

From the perspective of **national security** and integrity, LT is often employed to flag or identify possible risks that can be detected in written format. National concerns such as threats to national security, money-laundering and people-trafficking are often intercepted through advanced technology in this space. When relevant documents or audio/visual recordings are in a technologically unsupported language however, such instances of national interest remain undetected.

Similarly, new advances have been made in **event detection**, based on what is being reported in real time in social media by citizens and eye-witnesses (e. g., natural disasters, accidents), supporting information gathering for first responders, governments and newsrooms. Of course, this analysis on large amounts of data is only possible for the content in languages that are supported sufficiently through LT. Where a language is not supported, any relevant content written in that language is therefore disregarded and rendered unusable.

Courts and criminal justice systems are now benefiting from multimodal approaches to content retrieval combining speech processing and NLU to assist in the discovery of evidence amongst large amounts of unstructured audio and video content. Inequalities are likely to arise in the legal system however, as processing times will improve only for those whose languages are suitably supported through these technologies.

Sentiment analysis of online political commentary (e. g. news articles, social media, etc.) is often used by governments and **political parties** to gauge their popularity based on the electorate's opinions online (i. e., what is being said about them). In addition and true to predictions[28] that the future of government service ratings would lie in the hands of sentiment mining, the UK is one such example of a government who has embraced the power of topic modelling and sentiment analysis to analyse the feedback provided by citizens in their GOV.UK website.[29] Similarly, online data mining is often used as a technique for **predicting election outcomes**. However, in a multilingual society, only the opinions or comments of those in the technologically supported languages will be represented. In other words, the voices of many will be left unheard, unrepresented and unaccounted for.

[28] https://datasmart.ash.harvard.edu/news/article/from-comment-cards-to-sentiment-mining-301
[29] https://dataingovernment.blog.gov.uk/2016/11/09/understanding-more-from-user-feedback/

Business and Consumer Benefits

From an EU **Digital Single Market** perspective, the importance of being able to reach wider markets and consumer bases through the use of machine translation should not be underestimated. Nor should the importance of effective multilingual online dispute resolution.

Additionally, all European economies have seen a shift towards **eCommerce** in the past several years. This shift has benefited both businesses (wider market reach) and consumers (convenience and more choice). TA plays an important role in supporting both parties. From a commercial perspective, businesses no longer need to conduct market research polls to gauge customer satisfaction. Instead they can use sentiment analysis to assess online reviews, mentions in social media and customer feedback forms. Personalised advertisement also helps to find the right potential customer base.

From a customer's perspective, more efficient **customer service** (through chatbots, virtual assistants or automatically generated FAQ sections) makes buyer-seller interactions more seamless. Multilingual systems widens these benefits even further. Effective online search through product websites is also supported through TA and MT.

It is clear therefore, that for economies and societies to grow and evolve at the same pace, they need equal access to such advancements in LT.

A further economical aspect concerns the impact of LT on automation of tasks and as a consequence on the job market as a whole. As technologies such as chatbots are being adopted in pursuit of efficiency, they also perform an increasing number of tasks previously reserved for humans. LT and AI thus blur the boundary between humans and technology, leading to shifts in jobs and entire industries. Clearly, a message of cooperation and support rather than of rivalry and replacement needs to be communicated and acted upon.

Education

School-based learning and education is changing rapidly in terms of technological support. In many learning environments, there is a shift away from the traditional pencil and copybook approach towards **technology supported learning**. This shift is supporting learning and growth, and ultimately improving quality of lives and leading to better societies.

For instance, Computer Assisted Language Learning (CALL) tools are increasingly employing LT to create intelligent learning support systems. For example, personalised or **adaptive learning** is a technology that allows the identification of a student's progress and gaps in their knowledge, while adapting the curriculum, learning pace or learning goals to suit the learner. Such adaptive learning has proven invaluable for subjects such as language learning, maths and science (Chen et al., 2021).

In bilingual countries there is often a more dominant language that influences the language medium through which education is offered across society. In these cases, **language immersion schools** are also offered to those who, instead, want their children to receive an education through their mother tongue. While these lesser-spoken-language medium schools are key to ensuring continued use of the language across generations, the availability or lack of language technology to support learning could eventually create a divide in the levels of education on offer to citizens within regions or states, contributing further to inequalities.

Likewise, a major challenge for assessing large groups of students is the ability to track their learning progress. **Learning progress analytics** is being made possible through TA and NLU, in settings such as automatic scoring as applied to English content in the US.[30] Very little research has progressed to market for these types of applications for other languages.

[30] https://www.ets.org

Disability and Special Needs Support

Text-to-speech (TTS) is considered assistive technology and as such, it may contribute to better integrating into society people with **visual impairments** and **learning disabilities such as dyslexia**. By developing robust systems capable of reading any text from any source, including books, websites and social media, these people would be able to enjoy the same advantages as any person without a disability. It facilitates equal access to education for people with visual and learning disabilities as well as for foreigners who may struggle with the language. In addition, it can contribute to the integration of immigrants into society by making it easier for them to learn the local language, as TTS allows one to listen to words and sentences while reading them. TTS can also help people with literacy issues and pre-literate children learn to speak and access contents presented in written form.

Another contribution of TTS to society relates to orally impaired people, where technology is able to provide a voice for those who have lost their own. **Synthetic voices** can be personalised so they suit the characteristics desired by each user, by applying speaker adaptation techniques. It is even possible to generate synthetic voices that can reproduce the sound of the voice the person had before they lost it is possible, provided recordings are available. This way individuals can speak with synthetic voices that match their personality and character instead of using the standard voices provided by default by companies.

Ethical Implications for Society

As discussed in Section 2.2, while advances in LTs are considerably improving our lives, some technologies also carry unintended hidden dark sides that can negatively impact societies.

As technologies are entering the homes and offices of users on a broad scale, an enhanced level of attention to privacy concerns, ethics and policy is essential. Additionally, the main applications of Automatic Speaker Verification (ASV) are the areas of access control, surveillance, forensics or voice assistants (e. g. to authorise access to resources such as a bank account or building, or detecting and identifying a wanted criminal in a collection of audio recordings). Trust is therefore viewed as the main currency and key to the adoption and acceptance of technologies. Scandals, data breaches and opaque behaviour on the part of ST providers may have detrimental effects.

Current DNN based TTS systems have reached a quality level and a degree of similarity with the voice of real people that could be used to generate **deepfake voices**, which could be used as a tool for illegal activities such as committing fraud or discrediting people. New regulations and the development of ad-hoc legislation is critical to mitigating this pernicious effect of the TTS technology. New tools to detect and prevent speech deepfakes must be produced, and anti-spoofing techniques that discriminate synthesised from natural speech must be developed in close collaboration with teams working in TTS.

LT and subsequent automation and multiplication of services could be beneficial for underrepresented minorities from an inclusion perspective. Parts of the population may not have access to smart devices or not be media-literate. Language conveyed by means other than audio (e. g. sign languages) may be at a disadvantage and technically require different processing channels (visual processing). For speech output, powerful TTS technology ready for use in many languages (any language) and equipped with efficient interfaces is imperative to achieve an inclusive society where everybody has equal access to information, education and communication.

A further area of concern is the extent of **unlawful surveillance by governments, state agencies or (large) corporations**, infringing citizens' rights, liberties, adversely affecting public discourse, democratic values and influencing the political powers (Stahl, 2016). The concerns about the extent of privacy invasion, accountability of intelligence and security services, the (non-)conformity of mass surveillance activities with fundamental rights (Garrido, 2021), their effects on the social fabric of nations can only be considered and analysed jointly with the rapidly extending technological capacities, and the pervasiveness of devices able to capture, process and transmit relevant data. The growing extent of mass surveillance and especially its unlawful application may lead to erosion of public trust in governments and state agencies.

Career and Growth Opportunities

The world of job-seeking and career moves has changed significantly over the past several years. Today, in the English-speaking world at least, professional networks and job databases such as LinkedIn have changed the way in which recruiters find potential candidates and job seekers find **potential career options**. In turn, these opportunities empower and strengthen a workforce and societies. TA and NLU are fundamental in this process and much of the language technology powering these kind of systems is AI-driven. In many ways, they also benefit from the power of knowledge graphs and relationship linking to enable the right recruiters find the right candidates by matching users' CVs to job descriptions. This provides an advantage to both businesses and individuals.

Upskilling and re-education are also high in demand nowadays, with learning platforms providing tailored learning based on users' interests, previous experiences and so on. These personalised systems are also enabled through TA technologies, matching the right courses with the right users. Such learning platforms are therefore enabling growth and opportunity that will improve not only the lives of individuals but also leading to wider impact at a society level as a result of a strengthened and more skilled workforce.

In the absence of wide language support in this sphere, it is evident that only specific language communities (including businesses and citizens alike) are set to gain advantage through a more skilled workforce.

Digital Interaction and Connected Societies

Language support and proofing tools (e. g., spell-checker, grammar-checker, auto-correct, predictive text) facilitate efficient and seamless creation of digital text content. Today, it is unusual to find (for English at least) a platform or application that does not provide such language support (e. g., customer review forms, micro-blogging platforms such as Twitter, blogs, messenger tools, etc.). As such, they are often viewed as fundamental requirements for any text-based content creation technology. However, very often such support does not extend to other languages. Consider the simple examples where a user attempts to write content in their own language but their words are instead "auto-corrected" to a word in another *supported* language or underlined in red as a typo or invalid word. This is a frequent occurrence and challenge for speakers of minority languages. In such cases, one of two outcomes occur: (1) over time, users will default to writing in another supported language (if they can speak one) or (2) they will stop using the technology. In the case of (1), this is a clear step towards **language shift** and eventual language decline, particularly amongst younger generations. In the case of (2), this creates a **divide in levels of accessibility and usability** across language communities.

Health

According to the Health Europa,[31] virtual cognitive assistants could drastically reduce the administrative burden and lead to **improved patient experience and health outcomes**. Al-

[31] https://www.healtheuropa.eu/patient-experience-virtual-cognitive-assistants/91679/

ready in the medical industry we can see investment in cognitive agents like virtual medical billing assistants, virtual radiology assistants, virtual plan of care assistants, virtual medical testing assistants, etc.[32]

According to *Research And Markets*,[33] the virtual medical assistant market is expected to grow from $1.1 billion in 2021 to $6.0 billion by 2026. The smart speakers segment of the **healthcare virtual assistants** market should grow from $813.1 million in 2021 to $4.4 billion by 2026, while chatbot segment – from $317.3 million in 2021 to $1.6 billion by 2025.

At the height of the COVID-19 pandemic, the role of virtual assistants increased in the medical domain, since virtual assistants were able to provide the public with convenient and fast access to trustworthy information such as the latest regional, national and international illness statistics, relevant contact information including information hotline numbers, information about the virus, border crossing, the nearest analysis delivery points, how to act in various situations etc.

Integrated with virtual assistants, TTS systems are able to provide support to the elderly, assisting them with reminders of appointments and medication needs, providing them access to online information and improving both their ability to live by themselves and strengthen their autonomy. Studies have already shown that this technology can also benefit any individual living alone by allowing them to have conversations and being a kind of social companion, helping to reduce loneliness (e. g., Zsiga et al., 2018; Cooper et al., 2020). Similarly, ST applications in health and **elderly care technologies** enable interventions to be triggered by the detection of certain emotional states in users' voices. Furthermore, ST can prove helpful for ageing populations with degrading eyesight.

Multilingual and cross-lingual text analytic tools for medical domain can also help in **knowledge transfer**, fact finding and fast solution finding when rare and less common information is necessary. This is particularly relevant if solution needs to be provided in urgent situations, where immediate response is crucial.

A growing area of research and development in the health domain is the emergence of **medical transcription tools** that will support doctor-patient interactions. Research has shown that these interactions lack in terms of the attention the doctor can spend engaging with the patient face-to-face, due to the overhead of note-taking. Medical transcription or scribe tools, using a combination of speech and NLU technologies, are being introduced to improve this interaction and also make note-taking more consistent and structured. The quality of the data then captured through these tools will further lead to **improvements in healthcare**. Societies and language communities that do not have technologies to support their local language will not benefit from these advances in the health sector.

All of the above raises the following questions:

- Will the commercially important languages continue to stay ahead of the majority of languages in the long run?
- What impact will this have on speakers of such smaller (lesser spoken) languages?
- Will a lack of commercial interest in such "small languages", also translate to a lack of improvements and innovation in these communities and societies?
- How much will the imbalance between language support cause language shift where speakers choose to use English (or another major language) as this might provide a better experience instead?
- Will the digital footprint of minor languages be reduced to a minimum and eventually be marginalised?

[32] A Review of Cognitive Assistants for Healthcare is recently published by Preum et al. (2021)
[33] ResearchAndMarkets.com

- Will these marginalised linguistic communities lose out on the advances (through LT) in their education, health, economies, public sectors and general societal improvements?

3 The LT Technological Landscape in 2030

3.1 The Vision

The ELE Programme, in the form of the SRIA as well as a roadmap, conceptualised by the ELE consortium, will serve as a blueprint for achieving full DLE in Europe. The current scientific goal envisioned for 2030 is Deep Natural Language Understanding. Human languages are incredibly complex. We do not yet have algorithms or machines that are able to accurately and seamlessly integrate modalities, situational and linguistic context, general knowledge, reasoning, emotion, irony, sarcasm, humour, metaphors, culture, explain themselves at request, or that are able to do all of this as required on the fly and at scale reliably across domains for the many languages of Europe and beyond. All of these bear on and are the hallmarks of truly deep language processing. Deep understanding is understood in the sense that the resulting application using LT is able to explain itself: why did it make the decision it made given the linguistic context, the situational context (across modalities), linguistic knowledge, world knowledge etc.

Since 2010, the topic has been receiving more and more attention, recently also increasingly on a political level. In 2017, the study "Language Equality in the Digital Age – Towards a Human Language Project", commissioned by the European Parliament's Science and Technology Options Assessment Committee (STOA), concluded that the topics of LT and multilingualism are not adequately considered in current EU policies (STOA, 2017).

Over the coming years, AI is expected to transform not only every industry but society as a whole. The scientific and technological roots of LT are deeply embedded in AI and Computational Linguistics, especially with regard to the development of knowledge-based systems for language understanding. LT and NLP are, by now, considered important driving forces. An increasing number of researchers perceive full language understanding to be the next barrier and one of the ultimate goals of the next generation of innovative AI technologies (STOA, 2017). The European Parliament adopted, on 11 September 2018, with a landslide majority of 592 votes in favour, a resolution on "language equality in the digital age" that also includes the suggestion to intensify research and funding to achieve Deep Natural Language Understanding (European Parliament, 2018).

Both the STOA Report and the EP Resolution emphasise that there is an enormous need for a large-scale, multidisciplinary LT development and deployment programme that benefits European society, industry and politics. The opportunities of developing technologies for cross-cultural communication in Europe, and beyond, are almost endless.

Various internal and external consultations and surveys conducted by the ELE consortium have confirmed that there is still a huge gap between the LT support for English and the other European languages, with dramatic differences in several cases. Even though there is an increased interest in bridging this gap and in expanding technological support to more languages, limited funding, uneven demand and various obstacles with regard to available resources make it a very challenging endeavour. Basic research is still urgently needed. The fragmentation of the LT industry remains a serious hindrance. At the same time, the last decade has seen progress on a larger scale than could have been imagined 10 years ago. Many experts highlight European excellence, also on a global level and consider leadership in LT and language-centric AI to be possible if the necessary conditions are created by political decision-makers (Way et al., 2022).

While the goal of Deep NLU by 2030 is ambitious, it can be reached by setting up a shared programme between the EU, the Member States, local/regional authorities and other stakeholders, including industry. It must necessarily include a balanced mix of basic research, applied research, technology development, resource development, innovation and commercialisation; education and talent retention must be taken into account, too, to ensure long-term sustainability. The programme should run for at least ten years, so that the political and societal goal as well as the scientific goal can be adequately addressed. Public procurement and a policy change towards "LT enabled multilingualism" are crucial related aspects.

3.2 Priority Research Themes: Towards Deep Natural Language Understanding

The ELE industry partners generated, in various focus groups, four technology reports to illustrate the demands, wishes and visions of the European industry in a structured way. These deep dives were compiled for the fields of Machine Translation, Text Processing and Text Analytics, Spoken Language Research and Applications as well as Data and Knowledge Resources. They offer in-depth and up-to-date analyses of their respective areas.

3.2.1 Multilingualism and Machine Translation

Machine Translation (MT) is one of the most traditional LT applications, which has been researched for more than 70 years now. It has been analysed, criticised and praised from different perspectives and in different contexts.

Today translation technologies are widely used the by the general public, public sector and government agencies, SMEs, LSPs and many other industries where generating and consuming high-quality multilingual content is indispensable. The use of translation technology will definitely continue growing, covering new application areas (e. g., Internet of Things, smart homes etc.), markets, supporting Europe's Digital Single Market and language equality.

With the help of neural networks, MT has recently improved significantly in its quality, consistency and productivity for an ever increasing range of language combinations and domains. However, in many cases the focus of new technologies is still on big, fully-resourced languages, in particular English, thus limiting diversity and reinforcing already-existing disparities. At the same time the neural network techniques have opened the path to developing a universal translation engine aiming to translate between any language pair with help of a single model. The application of neural networks to MT allows also to forego the independence constraints and move towards context-aware methodologies in MT. A novel approach attracting the attention of many researchers is unsupervised MT, where monolingual data suffices to build a working system. While much work remains to be done in this area, it emerges as one of the key pillars to drive language equality.

An important aspect for language equality that deserves special attention is the availability of data necessary for MT training and methods allowing to overcome data scarcity for less- and low-resourced languages and domains.

Needed breakthroughs include explainability, contextualisation, data collection and EU policies, focusing on carbon-neutral and trustworthy AI.

Training neural MT engines is resource-intensive, requires massive infrastructure and has a heavy carbon footprint. By developing efficient models and hardware, the EU has the opportunity to be a pioneer in training and developing green LTs (Bērziņš et al., 2022).

Many current LTs process sentences in isolation, typically ignoring the previous and subsequent parts of the text. However, a text is more than a random collection of juxtaposed sentences. Today's LTs also have limited capabilities related to meaning and intent. They also hardly consider colloquial language and often cannot resolve references or draw inferences. Next-generation LTs should feature contextualised, adaptive, multi-modal, knowledge-rich, genuine semantic understanding, including pragmatic interpretation.

In terms of core technology, evaluation methodologies, metrics and data for training and evaluation, MT needs NLP that goes beyond traditional capabilities such as detection of terms / keywords / labels, entities, relations, and sentiments. These capabilities – amongst others referred to as Deep NLU – will, in the context of MT, solve shortcomings that clearly identify MT output as being generated by a machine. There are long lists of those, but as examples, the following can be named:

- awareness of context and ability to consider annotations/metadata;
- output faithful to the intended communication purpose;
- take translation purpose/specifications/requirements into account;
- explain text rather than translating it, reflecting cultural diversity between the source and target languages and users;
- show empathy with the reader/listener when necessary and appropriate.

3.2.2 Text Processing and Text Analytics

Text Processing and Analytics tools aim to process unstructured text and to extract knowledge or meaningful information and insights from text sources supporting strategic decisions in different contexts. Tools have been in the market for several years and have proved useful to extract meaningful information and insights from documents, web pages and social media feeds etc. Text analysis processes are designed to gain knowledge and support strategic decision-making that leverages the information contained in the text. Typically, such a process starts by extracting relevant data from text that later is used in analytics engines to derive additional insights. Nowadays text analysts have a wide range of accurate features available to them to help recognise and explore patterns, while interacting with large document collections.

The success of deep learning has caused a noticeable shift from knowledge-based and human-engineered methods to data-driven architectures in text processing. The text analytics industry has embraced this technology and hybrid tools are incipiently emerging nowadays.

While the progress made in the last years is undeniably impressive, we are still far from having perfect text analytics and natural language understanding tools that provide appropriate coverage to all European languages, particularly to minority and regional languages (Gomez-Perez et al., 2022).

3.2.3 Spoken Language Research and Applications

Speech – as the most spontaneous and natural manner for humans to interact with each other and ideally also with computers – has always attracted enormous interest in academia and the industry. Speech Technologies (ST) have consequently been the focus of a multitude of research and commercial activities over the past decades. From humble beginnings in the 1950s, they have come a long way to the current state-of-the-art, deep-neural-network (DNN) based approaches.

Especially over the past couple of decades, ST have evolved dramatically and become omnipresent in many areas of human-machine interaction. Embedded into the wider fields of AI and NLP, the expansion and scope of ST and their applications have accelerated further and gained considerable momentum. In recent years, these trends were paired with the ongoing, profound paradigm shift related to the rise of various data-driven models.

Current technologies often require the presence of large amounts of data to train systems and create corresponding models. Despite the lack of massive volumes of training material (e. g., transcribed speech in case of ASR or annotated audio for TTS), recent advances in ML and ST have begun to enable the creation of models also for less common languages. These approaches however are generally more complex, expensive and less suitable for wide adoption. While recently presented results indicate that novel approaches could indeed be applied to address some of the challenges related to the creation of models for low-resourced languages, the scope of their application and inherent limitations are still the subject of ongoing research (Backfried et al., 2022).

3.2.4 Data and Knowledge Resources

LT requires a range of specific language data resources that can be used to develop working monolingual, multilingual and cross-lingual applications.

While the acquisition, filtering, cleaning, annotation and preservation of language resources might seem a necessary, but methodologically known task, it is in fact the opposite. With the growing number of areas where LTs are used and applied, the need for specific data in specific domains and for specific purposes is also growing.

This is true for all types of language resources: monolingual corpora, bilingual/multilingual corpora (including parallel and/or comparable), monolingual/multilingual lexical and terminological resources. In addition, the growing number of applications generates the need to annotate data for very specific tasks, at least in reasonable quantities, even if the existence of large language models might help here.

Research is thus needed to find faster, cheaper, more reliable and if possible massively multilingual methods and procedures that will generate the necessary datasets in a short time and in good quality. This of course goes hand in hand with fundamental research on language models and in general on Deep Learning, since progress there can change the need for data in volume, annotation and other aspects.

In addition, the will be more need for LRs combined with image, video, gesture, facial expression and possibly other types of modalities.

4 The Path to Digital Language Equality in 2030: Recommendations

4.1 Overview

The main political and societal goal of the ELE project is the development of a plan to achieve DLE in Europe by 2030. The "working" scientific vision is "Deep Natural Language Understanding by 2030". Its implementation is being discussed and consolidated with the ELE community and will underpin the SRIA and Roadmap produced by the project. This important process has been based on and inspired by, in particular, the analysis of the current state of technology support for Europe's languages and the identification of gaps and issues with regard to LT.

Independent of a specific technology area, two points are particularly critical to achieve DLE in Europe by 2030:

- Neural language models and related techniques are key to sustain progress in LTs. Therefore, being able to build neural language models for other languages with the same quality as English is key for language equality;

- Multilingual data is the key element to train such models in a variety of languages. We should not take for granted that large amounts of publicly available corpora of good quality can be readily obtained for all European languages, rather the contrary. The effort to ensure that all languages have large amounts of publicly available corpora of good quality, taking into account fairness issues, should be at the center of any future efforts towards DLE.

4.2 Survey Results and Recommendations

In order to achieve DLE, there are many aspects that need to be considered. Depending on the context of individuals or groups, different aspects are particularly relevant along the way. Notwithstanding the large consortium of this project, without the inclusion of external stakeholders, important aspects for the journey might not be sufficiently considered. To ensure that our 2030 vision includes as many perspectives as possible, we conducted and analysed three large-scale surveys for the target groups of LT developers, LT users, and LT consumers. In these, participants were asked about their opinions, perspectives, and their needs regarding LTs and DLE. In the following sections, the participants' recommendations for achieving DLE in 2030 are presented.

4.2.1 LT Developers Survey

The European LT developers community is composed of industry and research. Besides this distinction, the development of LTs crosses different disciplines, such as Computational Linguistics, Computer Science and Artificial Intelligence, resulting in a diverse group of stakeholders. From this heterogeneous group, 321 respondents from 223 different organisations participated in the survey. Academic institutions are represented with 73%, while private companies constitute 22% (the remaining 5% belong to the group "Other"). Moreover, the organisations represented 32 different countries, covering all EU member states and other European countries. Further information about the study was published in Deliverable 2.17 (Way et al., 2022).

Regarding the predictions and visions for the future, the participants named several times the availability of resources. All European languages should be supported by a critical mass of resources in different domains for free or at a reasonable cost by 2030, as these are needed for the development of LTs. LT developers want to work intensively in the next years on the automation of data collection, annotation and curation and on the problem of data bias. Therefore, we expect the situation regarding language data to be significantly improved by 2030. Additionally, the participants envisioned a development in the next years solving the step from language processing to language understanding to enable seamless human-like interaction for all Europeans in their own language.

Important instruments helping to achieve DLE in 2030 were considered to be long-term programmes enabling the needed groundbreaking research in the direction of language understanding, and investment in already existing research infrastructures supporting LTs. Recommendations regarding the technological level stressed the investment in the development of new methodologies for transfer/adaption of resources/technologies to other domains or languages as an effective measure to boost less supported languages. Given the many gaps that need to be filled, most of the participants would appreciate an increase of qualified LT personnel and incentives for talent retention. The funding instruments of the last years helped to establish Europe in the LT field. Further investments in the next years are needed in all domains, especially in the basic research and not only in the applied aspects of LT. Some participants also would like to provide incentives to language communities that strive to preserve their language. Research collaboration with the industry should be further supported, with ideally less bureaucracy to ease the inclusion of small companies. In order for an increased visibility of the local industry and a better collaboration between the communities in the different countries, national centres of excellence in LT were considered to be critically important. Regulatory documents such as guidelines or recommendations, e. g. the FAIR principles, are an important instrument for driving research and development in the right direction. These should be increasingly implemented and expanded. The creation of such a document could have positive effects in some areas, as content accessibility regulations for multimedia creation. Awareness raising in the community of LT researchers and developers was considered another important point towards DLE. Besides this, increased incentivising for journals and conferences dedicated to less supported languages is considered necessary. Finally, social and linguistic diversity are strongly connected. Therefore, actions towards social diversity, like large-scale policies against racism and discrimination, will have an impact on the development of LTs and LRs, as the need for multilingual resources and tools will also rise.

The most important aspect for the future steps in Europe is that the resources and tools will strictly adhere to key European values such as privacy, transferability, fairness, diversity, openness, transparency and accountability, public wealth, individual rights and collective purposes (Way et al., 2022).

4.2.2 Users Survey

The LT users and consumers consist of professionals and communities that use LT on a regular basis. Various stakeholders from this group were surveyed in order to collect data for an analysis of the level of technological support for the EU official languages and EU lesser-used languages. This survey received a total of 246 responses from professionals working in a diverse range of sectors and activities. Most of the respondents work in the Education and Research sector with 130 responses (53%) out of 246, that is, most respondents were researchers, university professors, assistant professors, lecturers or held other academic positions. The survey was also filled out by representatives of NGOs, large enterprises, SMEs, government departments and independent contractors and consultants in diverse economic sectors. The 15 (6%) respondents who selected the option "other" represented non-governmental bodies, non-profit organisations, public sector organisations, social organisations and independent government departments.

Respondents were based mainly in European countries, although some participants indicated that they were based outside Europe such as United States of America and Republic of Congo. In Europe, the most represented countries were Croatia (33 responses), Spain (23 responses), the UK (23 responses), Ireland (17 responses) and Germany (16 responses). Detailed figures can be found in Deliverable D2.17.

The survey showed that 74% of the respondents work with English, which is followed by a well-balanced group of languages composed by German, French and Spanish. In relation to other European languages, respondents mentioned Basque, Catalan, Macedonian, Luxembourgish, Moldovan, Welsh and Galician. 50 respondents (20%) indicated that they plan to work with additional languages, most often English, German, Spanish and French. Thus, the survey shows that in a multilingual and multicultural Europe, most minority, regional, lesser-used languages are disregarded either for not being commercially interesting or simply for lack of institutional investment and engagement. Detailed figures can be found in Deliverable D2.17.

Regarding the evaluation of the current situation, the survey showed that English is the best supported language, followed by German, French and Spanish. In relation to the most used tools, the survey results revealed that the most used LT tools in EU official languages are translation tools, followed by proofing tools, search engines, and language learning tools. Search engines are less likely to be used in minority, regional, lesser-used languages due to poor performance.

The survey also showed that respondents perceive gaps in the tools they use. The most common gaps perceived are in relation to the amount and variety of applications available. Within this group of responses, this gap was more frequently perceived by respondents working with LTs in Estonian (100% of respondents), Maltese (86% of respondents), Latvian (83% of respondents), Bulgarian (72% of respondents), Czech (67% of respondents), Slovak (58% of respondents), Irish (56% of respondents) and Romanian (50% of respondents). In contrast, for English, this gap is only perceived by 4% of respondents, German 10%, French 10%, Spanish 11% and Italian 14%. Gaps in the quality of available applications were more frequently perceived by respondents using LT tools in Icelandic, Maltese, Croatian and Bulgarian, but less perceived by respondents using LT tools in Italian and English. Gaps in the variety of linguistic phenomena covered by the tools were perceived by 50% of respondents using them in Icelandic, 43% in Maltese and 39% in Irish, but this gap was only perceived by 1.9% of respondents for English.

The responses to the open-ended questions show that the LT users and consumers wish to increase the variety of tools and resources available for minority, regional, lesser-used languages. Respondents indicated several things they would like to see in a tool that would make LT more useful in their work. For instance, respondents wish for higher-quality tools for certain languages such as “better parsing of Danish than currently available” or the availability of tools that do not yet exist for some languages but exist for other languages such as “speech recognition for Welsh”, “speech recognition for Catalan, better grammar checking for Catalan”, “free spell check for Irish”, “more reliable speech recognition, information extraction, summarisation, semantic parsing and semantic search for Greek”, “A good Georgian-English Translator” and ”better MT for Croatian language”. A further problem related to this is the documentation for the language technology only being available in English for many of these existing language tools. The lack of open-source language tools and language resources (language learning materials, school books, open-source dictionaries, translations resources, stop words, stemmers, written documents, audio data or spell checkers) – which is especially true for minority, regional, lesser-used languages – has also been mentioned by the respondents as a serious hindrance for reaching more digital equality for languages in Europe. Another gap identified was the insufficient long-term funding for projects and institutions (e. g., libraries) working with regional and minority language.

Some visions that the respondents formulated concerned multilingual translation tools (translating into multiple languages at once) or real-time collaborative translation tools that allow speakers of different languages to work together on one text. Furthermore, a linked open data environment for lexicographic data could allow for stronger links and translations from one minority, regional, lesser-used languages to another.

The most important finding of this survey is the respondents’ concern regarding the differences in technological support between European languages, specifically the poor technological support of minority, regional and lesser-used languages. As we could see from the findings described, there is a huge gap in support between the European Languages which are reflected in terms of differences in performance of tools across languages as well as in terms of lack of availability of tools for certain low-resource languages. Thus, the results show that, in order to achieve full DLE as a crucial step to maintain linguistic diversity, the survey shows the necessity for action and an implementation agenda with the objective of fostering and supporting a multilingual and linguistically inclusive Europe that brings solutions to all European citizens.

4.2.3 EU Citizen Survey

In addition to the survey targeting representatives of LT developers and LT consumers, the ELE consortium also organised the EU Citizen Survey with the aim of taking into account the opinions, individual needs, wishes and demands of Europe's citizens. The preliminary results were collected in 28 European countries with the help of a service provider in 28 languages. Table 1 shows all countries where the European Citizens Survey was disseminated, the sample size per country and the sample size per language.

Although the at the time of writing the survey has not yet closed, the preliminary results with 18,963 responses collected via a service provider (see Table 2 in Appendix 5.6), overall, the survey shows that the most frequently used languages appear to be English, German and French. These languages appear to be frequently used even in non-English, non-French and non-German speaking countries. Results also show that the top three most used LT tools are MT, search tools and proofing tools and those are the most well evaluated tools in terms of performance – this result is consistent across all languages included in this sample. Automatic subtitling tools are also among the most well evaluated tools in four languages, namely, Hungarian, Serbian, Lithuanian and Dutch. The results also show that these three most used tools are also the top rated tools in a 5-point scale.

The survey also asked respondents to indicate why certain tools are used with certain languages but not with other languages. The most selected response for the question "what holds you back from using these tools?" was the lack of available tools. This option was more frequently selected by respondents that use LTs in Valencian/Catalan, Czech, Bulgarian, Slovenian, Polish to name a few languages. Regarding future demands, the survey shows that personal assistant tools are the most desired tool for the future in many European languages (e. g. Bulgarian, Croatian, Czech, Hungarian and Lithuanian). These results suggest that personal assistant tools such as Alexa or Siri are not available in certain languages or, if available, they are not yet frequently used in many European languages. Thus, the survey shows that there is currently a high demand for the use of these tools in the (near) future.

In the last section of the survey, respondents were requested to select the top 3 advantages of improving apps and tools for all languages. The preliminary results show that the top 3 advantages in respondents' opinions are "to increase peoples' exposure to these languages" with 8828 responses (close to 50% of the total sample); the second advantage in respondents' opinions is "to improve communication between speakers of different languages" with 8,538 responses (48%); and, finally, "increase the number of speakers of languages, including minority and regional languages" with 7,112 responses (40%).

The survey is, at the time of writing, still available on the ELE website.[34] It can be answered in 35 European languages. In addition to the 18,963 responses collected by our service provider, the dissemination promoted by the ELE consortium collected so far 2,423 responses, thus totaling 21,386 responses up to 29/04/2022. See Table 3 in Appendix 5.6.

4.3 Technology Areas

The technology areas covered by the surveys and Deep Dive reports are Machine Translation, Speech Technology, Text Analytics and Natural Language Understanding. The cross-sectoral topic, namely Data and Knowledge Bases, is covered in Section 4.4.2.

Needed breakthroughs for **machine translation** are related to system development (including interoperability, explainability, contextualisation, hardware needs and opportunities and opportunities offered by quantum computing), data collection and EU policies, focusing on carbon-neutral and trustworthy AI. Special emphasis is placed on deep learning architectures (e. g., Transformer models), research on NMT model efficiency, use of broader con-

[34] https://european-language-equality.eu/language-surveys/

Countries	Total Sample Size	Language(s)	Sample per Language
Austria	900	German	900
Belgium	900	French	350
		German	50
		Flemish Dutch	500
Bulgaria	750	Bulgarian	750
Croatia	600	Croatian	600
Czech Republic	900	Czech	900
Denmark	600	Danish	600
Estonia	150	Estonian	150
Finland	300	Finnish	250
		Swedish	50
France	900	French	900
Germany	900	German	900
Greece	900	Greek	900
Hungary	900	Hungarian	900
Ireland	550	English	450
		Irish	100
Italy	900	Italian	900
Latvia	200	Latvian	200
Lithuania	300	Lithuanian	300
Netherlands	900	Dutch	900
Norway	600	Norwegian	600
Poland	900	Polish	900
Portugal	900	Portuguese	900
Romania	900	Romanian	900
Serbia	100	Serbian	100
Slovakia	550	Slovak	550
Spain	900	Spanish	750
		Catalan	50
		Galician	50
		Basque	50
Sweden	900	Swedish	900
Switzerland	400	French	150
		German	200
		Italian	50
United Kingdom	1000	Welsh	100
		English	900
Slovenia	250	Slovenian	250

Table 1: Sample size per country and language

texts (e. g., documents instead of isolated sentences) and multiple source inputs (e. g., source sentences in multiple languages), use of linguistic knowledge (e. g., morphology, syntax, semantics) and external knowledge (e. g., domain-specific terminology, domain information, etc.), multi-lingual and multi-domain NMT, use of pre-trained models (e. g., BERT, mBART, etc.), multi-task learning, automatic post-editing, and other methods that allow achieving state-of-the-art translation quality for NMT systems.

When looking forward to 2030, we expect the movement towards Deep Natural Language Understanding smoothly and seamlessly enabling efficient and real-time translation to support human-to-human or human-to-machine communication. We expect a major breakthrough towards efficient, omnipresent, high quality **real-time translation between any European language pair** and in any domain, regardless of the modality (written, spoken, sign language) of the input.

While text-to-text translation is widely used today, speech, sign language and multi-modal MT is still relatively in its early stages. There is a growing need for the translation of audiovisual content and development of MT-centric text-to-speech and speech-to-text applications that can support the meaningful integration of the written and spoken word and images. Speech translation and voice interaction with devices are the key techniques to break the language barrier for human communication. In order to achieve human-like language processing capabilities, machines should be able to jointly process **multimodal data**, and not just text, images, or speech in isolation. There is also a need for accessible content in the form of subtitles and audio descriptions.

Future systems should be evaluated by **new automatic metrics** which represent better approximations of human judgments and also ideally abandon the dependence on human reference translations. Moreover, evaluation should not be carried out on isolated sentences/segments. Increased attention should be paid to the human judgments used for tailoring the automatic metrics, as well as to manual evaluation in general.

Going towards the ambitious goals to be achieved by 2030, different aspects regarding the Text Processing and Analytics tools deserve further investigation. Firstly, multilingual text processing and analytics needs to be strengthened. Currently, research on unsupervised and zero-shot learning (Radford et al., 2019; Brown et al., 2020; Gao et al., 2021) as well as on multilingual language models (Conneau and Lample, 2019), language-agnostic models (Aghajanyan et al., 2019) and neural MT (Johnson et al., 2017) enhances the processing and support of regional and minority languages. With further investment in this direction, we expect the language coverage to be improved by 2030.

Another crucial element that needs to be adapted to the new research and its results by 2030 is benchmarking. The currently used benchmarking systems hardly give room for newer, better results, because the current results are already classified as very good. When adapting the benchmarking, points such as data validity and specificity, reliable annotation, statistical significance, complexity and cost and disincentives for biased models should be taken into account (Bowman and Dahl, 2021). These aspects would push further research more in the direction of DLE than benchmarking efforts, valuable though they have been, have done so far. Another important aspect of benchmarking is the consideration whether the data is realistic regarding its setting and its composition.

Concerning **Speech Technology (ST)**, several recommendations and development trends can be identified:

Speech technologies integration: An intimate relation of ASR, SID and TTS with downstream NLP and NLU technologies is needed to allow the correct interpretation of the input so that recognition, meaning and output can be produced in a natural and correct manner. This future oriented and recommended approach is based on the combination of technologies, enabling interactions in multimodal ways (including visuals) and the efficient combination of inter-linked models will be able to guarantee the best experience possible. In turn, the successful combination will result in an enhanced easiness and naturalness of use, hiding individual components and allowing to perceive systems as assistants using natural language much in the way that human assistants would.

Support for less-resourced languages: To be able to provide first-rate ST in any language, additional high-quality datasets are essential. Ideally, they should be open and available without usage rights limitations for all the languages and include recordings with a variety of conditions and representative settings. These include a variety of speakers, language varieties, dialects, sociolects, data including spontaneous speech, varied prosodic patterns, diverse sentence lengths and a wide range of emotions. Creating this wide set may not be feasible in general, but could be achieved at least for several major European languages. New techniques for transfer learning and model adaptation from systems trained for one resource-rich language to systems able to function in languages with more reduced quantities of available data should be developed. These techniques would allow the development of cutting-edge ST systems also for less-resourced languages. Also, new recommended (see D2.14 for more details) architectures allow using resources from several languages in such a way that commonalities among languages are learned in a more robust way by cross-lingual knowledge-sharing or methods for the creation of multilingual or language-agnostic models which can be applied to a number of different languages are of utmost importance.

Multimodal models: Recently introduced neural net architectures, e. g., Perceiver IO (Jaegle et al., 2021), support encoding and decoding schemes of various modalities. They can directly work with BERT-style masked language modelling using bytes instead of tokenised inputs. Another advantage of this type of architecture is that the computation and memory requirements of the self-attention mechanism don't depend on the size of the inputs and outputs, as the bulk of computing happens in a common Transformer-amenable latent space. In the near future, this type of architecture will be commonly used in a range of applications where multimodal content needs to be jointly analysed. Further, the future line of work relates to the training of a single, shared neural net encoder on several modalities at the same time, and only using modality-specific pre- and post-processors. In the longer-term perspective, such multimodal, *plug and play* architectures and models, will provide strong baselines in many areas, potentially also supporting less technical users with visual design tools, tractable hyper-parameter search, automated architecture, popularising the access to high performance, multimodal analysis and inferences. It is recommended that the research on multimodal models be continued strengthened.

Addressing the existing technological gaps: In the area of ASR, continues efforts towards better understanding and modelling human speech perception might result in sophisticated speech recognition addressing several of the technical limitations and gaps identified in current approaches. Improved handling of audio conditions currently perceived as difficult (e. g., multiple simultaneous speakers in noisy environments speaking spontaneously and highly emotionally in a mix of languages) will be possible by such advances. At the same time, a wider deployment and further popularisation of ST will also require solutions that offer high robustness, low latency, efficient customisation and the ability to provide possible equal support for a diverse set of speakers. It is recommended that addressing these technological challenges should further drive the R&D activities in the ST fields.

User and application contexts: A trend towards the integration of richer context is to be expected, regardless of the sub-field of voice processing. The research in this area should be further strengthened, providing additional highly valuable cues for modelling non-laboratory human-AI interactions.

Development pace: The pace of development in voice-based technologies is driven by general advances in ML and associated hardware as well as domain-specific advances in the modelling of speech perception and production. The former can be expected to accelerate even more due to general interest in ML and AI from a wide portfolio of domains. Advances in transfer learning, reinforcement learning, fine-tuning, the use of pre-trained models and components as well as the arrival of platforms such a Hugging Face have created additional momentum. GPU support and extension of GPU capabilities can likewise be expected to continue at a fast pace, which might also have effects on the availability of hardware resources. The latter topics have been receiving increased attention as voice and language technologies entered the mainstream. Voice, being the most natural way to interact with systems can surely be assumed to attract even more commercial and academic interest in the future.

Training and evaluation: Simultaneously, there will be further improvements introduced in the process of creation and distribution of ever-growing, ever more coherent (labelling quality), and diverse datasets. These will also include the creation of and increase in a number of large, multilingual, multi-domain and multimodal datasets, that will become de facto standard sets for the training and evaluation of the ST methods and systems that include ST components. In the next years, we will also witness an increase in labelling efficiency, a wider adaptation of continuous learning, self-adaptation and self-modification paradigms. While the number of languages available in the datasets will continue to grow, the quality and amount of data available for the most common, currently rich-resourced and the less common, currently low-resourced languages are unlikely to converge in a shorter term. This development in the creation of more complex and multifaceted datasets calls for a more comprehensive evaluation and quality criteria; a shift that would change a focus from an individual speech technology to an end-user assessment of a complete experience when conducting a specific task in a given, non-laboratory environment and in a given operational and personalised contexts. Whereas current learning paradigms focus predominately on training models on massive amounts of data in one go, human learning takes place in complex steps over time, refining itself constantly along the way. New paradigms incorporating complex sequence learning may not only provide further insight into human language acquisition but likewise lead to even more powerful ST (NLP, NLU) models.

Customisation: Technologies may have reached an advanced level of maturity for many languages and domains. However, numerous further niches remain which require expertise and adaptation of base models to cover the last mile to the customer. In all areas of ST, the opportunity to capitalise on efforts and tasks which fall into this category exists and should be taken up by all parties involved in R&D of ST, including the local champions.

Ambient Intelligence: The confluence of individual technologies to form an entity that is larger than the sum of the individual technologies is a recurrent theme within this document. This is especially important when combining human-like modalities for input and output with knowledge representation and reasoning, potentially in an augmented or virtual environment. Viewing ST as a means for intelligent interaction, integrating nuanced and fine-grained context and input from multiple modalities can be expected to lead to more human-like systems where the perception of individual components will blur into an overall experience for end-users. Such combinations may be a step towards a broader kind of AI as opposed to the narrow, highly-specialised versions in use today. This line of work should be further explored and supported.

Supermodels: Recent years have witnessed a fierce race between renowned institutions and research labs on who can build the largest model for NLP. It has become customary that only actors with enormous resources at their disposal can participate in this race. Whereas the huge foundation models suffer from the same shortcomings as their predecessors in terms of bias, the integration of toxic language, the lack of explainability, etc., performance on many tasks is still improving with the number of parameters and no end of this race is currently in sight. As is the case for search technologies, the US and Chinese giants are leading these activities. European efforts like the German OpenGPT-X project[35] aim to mitigate this imbalance. In the recently published work, Bommasani et al. 2021 (Bommasani et al., 2021), provides a thorough account of the opportunities and risks of such foundation models, ranging from their capabilities, technical principles, applications and societal impacts. The

[35] https://www.iais.fraunhofer.de/de/presse/news/news-210701.html

research and development of supermodels should further focus the attention of ST and NLP communities, including the studies on their multifaceted and profound impacts.

Towards Deep Natural Language Understanding: The contribution of ST towards achieving Deep NLU is in the improvement and extension of the individual technologies (both from accuracy as well as a language-/domain-coverage perspective), from their integration into E2E systems allowing for joint operation and optimisation, including different kinds of knowledge sources and from their flexible and dynamic configuration depending on the state and context of an application or user. This recommended approach, combining several modalities for input and output may likewise prove beneficial for achieving Deep NLU.

In many cases, the real power of NLU will become apparent when it features as part of a complex system functioning as a human-like counterpart in communication – exhibiting contextual and historical awareness and elements of general intelligence. However, it may also be then, that NLU is overshadowed by the cognitive downstream processing and eventually perceived as a mere commodity. The element of admiration and awe on part of the user will then concern the complete system performance, with NLU itself disappearing in importance as a small part of a much larger and complex integrated intelligent system.

From the perspective of **Text Analytics and Natural Language Understanding**, support beyond widely spoken languages, including minority and under-resourced languages, is constant work in progress. The increasing adoption of approaches based on self-supervised, zero-shot, and few-shot learning opens new possibilities to increase the coverage of minority and under-resourced languages (e. g., (Conneau et al., 2020)). At the core of this trend, neural language models have shown promising results also in zero and few-shot settings across a wide range of tasks (Radford et al., 2019; Brown et al., 2020; Gao et al., 2021). This may have potentially interesting applications to eliminate or at least reduce the need of additional labeled data for fine-tuning over downstream tasks, which is a scarce resource for many languages. In addition, we expect that the language coverage of text analytics tools will be enhanced thanks to a **mixture of research breakthroughs** on multilingual language models (Conneau and Lample, 2019), language agnostic models (Aghajanyan et al., 2019), and others that fall more on the realm of neural machine translation (Johnson et al., 2017). It is thus recommended to continue research in these directions, paving the way for truly multilingual language technologies in the area of text analytics.

In a similar vein, the field of **neurosymbolic approaches to NLP and NLU** is also expected to contribute to alleviate the dependency on training data, as anticipated in e. g. Hitzler et al. (2019) and Gómez-Pérez et al. (2020). The integration of existing knowledge bases within pre-trained language models, as shown by approaches like KnowBert Peters et al. (2019) and K-Adapter Wang et al. (2021), will enhance such models, making them aware of the entities contained in a knowledge base and the relations between them as well as enabling a faster, cheaper and more scalable adaptation to vertical domains and organisations. Also, recommendable is the development of a greater methodological clarity in terms of what type of approaches to use, either neural, knowledge and rule based or a mix, depending on parameters like **data availability or interpretability requirements**.

We reiterate the importance of creating **new benchmark datasets** that take into account not only model accuracy but other types of metrics aimed at measuring the reliability with which they are annotated, their size, and the ways they handle social bias, including potential discrimination by language. In the area of digital language equality, to the best of our knowledge this is still fairly unexplored territory that will need to be progressively charted.

Also encouraged is the progressive development of **large multimodal models** that address not only text in isolation but also combined with other modalities. Models like CLIP (Radford et al., 2021) show that scaling a simple pre-training task is sufficient to achieve competitive zero-shot performance on a great variety of image classification datasets by leveraging information from text. The approach uses an abundantly available source of supervision based on pairs of text and images found across the internet, resulting in a gigantic language-vision dataset. Unfortunately, CLIP is available in English, Italian and Korean only, showing how language inequality also impacts on language-vision tasks. Investment in multilingual resources will also be necessary to make this type of technology available across all European languages as well as underrepresented languages in general.

Finally, we advocate for a next generation of language processing tools that care about the needs and expectations of end users, making them part of the design and learning process. **Human feedback** will serve as a guide for model training, telling the machine what users want and what they do not want (Christiano et al., 2017). Reinforcement learning from human feedback (Stiennon et al., 2020; Li et al., 2016) and interactivity with domain experts and general users (see Shapira et al., 2021; Hirsch et al., 2021) are key areas for further advances beyond the usual supervised paradigm.

4.4 Infrastructural Support

4.4.1 HPC and Hardware Infrastructure

Today large hardware infrastructures are required to accommodate for the required computation power and storage of Deep Neural Networks. While in North America and Asia public and private resources can be allocated to only a limited number of languages, to effectively honour the well-entrenched commitment to promote multilingualism in Europe resources must be distributed across a large number of official and unofficial EU languages, so that the respective language communities are treated fairly. As a result, the scale at which European research can be conducted is limited in comparison. There is also an uneven distribution of resources across countries, regions and languages (Aldabe et al., 2021a). Considering the massive infrastructure that is required to train very large state-of-the-art LT systems, Europe starts with a systemic handicap. Europe's strong foundation in research and innovation can compensate for the disadvantage European organisations have with respect to infrastructure, provided that a concerted effort is undertaken in researching the development of new hardware platforms and respective AI training paradigms.

In general, the hardware on which LT runs must be scaled down. Several approaches to replace GPU-based computing, or at least to make it more power-efficient, are already under investigation. By ensuring that the capabilities of the hardware are aligned with the needs of ML training and inference models, smaller models would be easier to integrate and use on any device and also be greener by requiring fewer resources, since training neural models is resource-intensive and has a heavy carbon footprint Strubell et al. (2019a). The EU has the opportunity to be a pioneer in developing such LT models by focusing also on efficiency both in terms of hardware and software. This would not only have positive environmental consequences, but it will also level the paying field for smaller and not well-resourced institutions and companies.

4.4.2 Data and Knowledge Infrastructure

In addition to hardware infrastructures, we also see a clear need for a comprehensive and interconnected data infrastructure that needs to be put in place to achieve the specified objectives.

To fill the identified gaps in data, language resources, and Knowledge Graphs we recommend and suggest a future path for Europe towards comprehensive and interlinked data infrastructures. These infrastructures have to provide interoperability out-of-the-box by following harmonised and well-proven standards, regarding (i) data (semantic data) interoperability as well as (ii) services and (iii) innovative metadata and data management tools that are available along all steps of the data life cycle.

Metadata, data, data-driven services and data-driven tools to be easily docked into these data infrastructures, without todays' huge efforts in data cleaning and data integration, or service- and tool integration. This future technology vision of integrated and interoperable data infrastructures shall follows the idea of a Semantic Data Fabric including rich semantics, and thereby context and meaning as well as dynamic metadata and augmented metadata and data management. By this approach a federated network and infrastructure of interlinked data spaces for language technology can be realised. Existing data spaces as well as newly developed ones should be integrated, where appropriate and possible.

In such a federated ecosystem relevant data regarding a domain and/or language can easily be identified, loaded, and evaluated for specific use cases. Data driven services are provided and can be used along end users requirements.

Integrated crowdsourcing and/or citizen science mechanisms allow human-machine interaction to foster data acquisition, cleaning and enrichment (e. g., annotation, classification, quality validation and repair, domain specific model creation, et al.). Raw data can be loaded into available tools to train algorithms or create memories and/or (language) models for specific use cases, but also existing algorithms, models or vocabularies are available and can be easily loaded and re-used to avoid unnecessary energy consumption / computing power to foster the idea of energy efficient data management.

In addition high importance needs to be put on privacy protection (related to personal identifiable information, PII and beyond), the avoidance of bias (for example on gender et al.), and on data sovereignty.

The approach of such data infrastructures require working and sustainable business models that allow data trading, -sharing and collaboration. And require supporting policies, as well as sustainable data governance models around data creation, data provision and data sharing. Well targeted publicly funded/supported programmes and activities in the area of data literacy are required from early education onward, to ensure that sufficient human resources in the field are available in the future.

In addition an action plan for the collection and the development of data and language resources that are relevant for language technology, as well as for Knowledge Graphs is needed to ensure the availability of sufficient data in the EU languages, as well as in dialects and important non-EU languages. The recommendation for this is to look into three areas, as: (i) Language Equality Action Plan by means of targeted national and European funding along a matrix of relevant resources and languages, combined with (ii) more measures in the fields of crowdsourcing and citizen science, and (iii) the development of functioning data related business models.

Beside technology, interoperability or data related attributes there must be a strong focus established on applying all these mechanisms and methodologies to the widest range of languages possible, at least to EU languages but also local and regional dialects of these languages, as well as to non-EU languages that are wide-spread across Europe. Without such data and language resources in place a digital language equality cannot be reached.

The availability of high quality data, language resources and knowledge graphs in at least EU 24 languages, but furthermore in as many languages as possible, that are easily accessible with fair conditions and costs in a clearly specified legal environment providing transparent rules and regulations can support clear benefits and competitive advantage for the stakeholders. For the European research community to foster innovations in the field, for the industry to successfully compete in a global market, and thereby for the European citizens and its society, that is constantly growing in regard to its diversity and a wide and increasing variety of languages. Data, language resources, and Knowledge Graphs are thereby a crucial factor on our way to digital European Language Equality.

5 Summary and Conclusions

5.1 Paradigm Shift: A Challenge and Opportunity to Achieve Digital Language Equality

The main purpose of this report has been to summarise the envisioned state of LT in 2030 with a particular focus on European languages, as gathered from various sources, especially those identified during the ELE project. The views expressed here are supported by careful analysis of the quantitative and qualitative input from the various surveys and interviews conducted, and by the analysis of the state of the art and the perceived gaps in several priority technology areas.

This report has assembled and discussed in a comprehensive fashion the contents from many previous deliverables, with primary input from D2.17 (Way et al., 2022), which in turn was based on Deliverables produced within WP2. The structure has been arranged in such a way to systematically cover the most important aspects of the SRIA, which will further elaborate on the views, foresight and priorities described herein. To summarise the main findings presented in this deliverable, we list below the most important points that in our view deserve to be considered particularly carefully in driving forward the effort to achieve DLE for all languages of Europe by 2030.

We are currently in the midst of a very significant paradigm shift, namely towards building the underlying LT as well as LT applications by using Deep Learning by Artificial Neural Networks in the end-to-end fashion, complemented with unsupervised techniques which require less costly data acquisition and preparation (but in larger quantities). This revolution in LT/AI has been possible because of the confluence of four different research trends: 1) mature deep neural network technology, 2) large amounts of data (and for NLP processing large and diverse multilingual textual data), 3) increase in High Performance Computing (HPC) power including GPUs, and 4) application of simple but effective self-learning (unsupervised) approaches.

Forecasting the future of LT and language-centric AI even for the end of the current decade is of course a challenge, due to the fast-pace developments that are taking place. Five years ago, hardly anyone would have predicted the recent breakthroughs. It is expected that more advances will be achieved by using pre-trained language and translation models as an example of a trend towards reusing and sometimes re-purposing trained models, with only modest costs for fine-tuning to various domains and applications. However, there are still many unsolved issues, presenting both a challenge and an opportunity.

At this time of technological transition, with the view of achieving DLE, we have an unprecedented and unique opportunity before us to address equal and fair support for all European languages, and eventually better digital provision for low-resourced languages, bringing them on par with the currently supported “large” languages in both quality and related application coverage. There are many aspects and challenges involved in achieving this ambitious but worthy goal, which this report analysed in detail.

5.2 Gaps in LT

Based on the analysis of the state of the art in the whole field with respect to the impact of LT on society, there are numerous gaps in several vertical as well as horizontal areas, which we review in the following concluding remarks.

Data

The uneven availability of suitable high-quality data for use in both training and evaluating today's state-of-the-art data-driven tools is a result of, and in turn regrettably reinforces, digital language inequalities. Obtaining clean and curated training data is a huge challenge, not only for several languages, but also in multiple vital domains. Labelling data can be a time-intensive task that often requires skilled domain expertise. Domain-specific language data (e. g., medical, legal, user-generated content, etc.) is needed to ensure sufficient coverage of certain terminology. In particular, the digitisation of educational material still has a long way to go. Much educational material in several languages is still largely published on paper. For a few languages with high commercial interest, an abundance of training data is available. However, for many (in fact, the majority of) European languages, this is not the case, and there is a need to instigate change to reverse well-established patterns.

With regard to accessibility, important steps have recently been taken in the research community with respect to cultivating a culture of open data and data sharing. Many top-tier publications require the release of datasets (where possible) in order to facilitate the reproducibility of studies. Shared tasks such as those involving benchmarking or evaluation campaigns require the release of their specifically designed datasets. However, enterprise data, for example, tends to be locked in regulatory and corporate silos. Particularly stringent copyright laws may pose a further barrier to research and development efforts in Europe, more so than in other competing areas of the world. The development, application and adoption of LTs are also connected to a range of issues relating to fairness, biases and ethical aspects that need to be accounted for. Similar to gender-related biases, race-related and ethnically-based biases and stereotypes may regrettably be present also in many LT models, and there is a need to prevent the serious harm that they are likely to cause. Biased tools are bound to have a direct negative impact on society as a whole and can cause substantial damage to already disadvantaged marginalised populations. Biases are a significant drawback which is yet to be addressed in full, especially considering their high potential to cause damage and embarrassment, which may undermine the credibility and appeal of LTs and related applications among both policy-makers, investors and the citizens at large.

Technology

Access to hardware, experts, and involvement in research have also shifted in such a way that elite universities and large firms have an advantage due to their ease of access to the required high-end facilities and expensive resources, which are often also very demanding to maintain and power. The lack of necessary resources (expert personnel, HPC capabilities, etc.) in Europe, compared to large U.S. and Chinese IT corporations is of special concern and needs to be alleviated by concerted efforts leveraging synergies between public bodies and private organisations. In addition, the lack of consideration for the specific needs and support required for users with a range of physical, sensory, cognitive and learning disabilities leads to other communities being regrettably marginalised despite advances in technology – this problem is compounded by the increasing amount of aging population all over Europe.

Interpretability is a major concern in modern AI and LT research. As such, a priority for many businesses and organisations is to build trust and confidence in these AI models. In particular, a notable increase in attention has been recently observed with regard to explainable AI. In addition, there are challenges in making responsible AI a reality: training neural MT, TA and ST engines is resource-intensive and has a heavy carbon footprint, which is another major concern that needs to be urgently and specifically addressed, to ensure environmentally-friendly and sustainable development in the future.

5.2.1 Benchmarking

Current benchmarking presents issues across all areas of speech and language technologies. In particular, there is still a lack of agreement within the MT community, as with the increasing quality of MT the widely used automatic metrics start to diverge from the true needs of assessing MT quality and suitability for various purposes. The situation is similar for other LTs, and for novel uses and applications, benchmarks are not even established – both in terms of methodology and the necessary datasets.

5.2.2 Expertise

Another significant gap that concerns all areas of speech and language processing is the scarcity of trained personnel and expertise, with the serious risk of losing emerging talent to innovative power-players outside of Europe, many of which can offer salaries and general working conditions that cannot typically be matched by academic or industry employers in Europe. In many cases, as the surveys and interviews have shown, the problem is not that education in Europe is inferior, but it is a question of retention, even though a higher number of trained staff (at all levels, including advanced users and maintainers of LT) would be very helpful both to industry and academia.

5.3 Impact on Society

LTs already impact extensively business activities, society and individual users' lives on multiple levels. The apparently highest impact so far has been in the area of businesses and consumers, where – apart from creating novel applications and whole areas of business – there have been benefits to the Digital Single Market by providing technology to reach wider markets and consumer bases, for example through the use of MT. In general, the possibility for companies to offer their services in all European (and some other) languages opens up desirable market opportunities for them.

The same benefits apply to Government Affairs and Public Services. Given today's increased mobility of citizens, migratory flows and progressive integration of international and cross-national decision-making processes, many government organisations already apply LT solutions to help them deliver efficient public services and improve governance.

Another broad area of impact is education. School-based learning and education are rapidly changing in terms of technological support, as we could see during the pandemic. LTs can bring new and more effective opportunities in the education process – not only specifically in the teaching of foreign languages, but to teaching in general, with a more inclusive and positive learning experience for a broader range of students (including, as mentioned above, those with disabilities).

Language support and proofing tools (e. g., spell-checker, grammar-checker, auto-correct, predictive text, as well as MT) facilitate efficient and seamless creation of digital text content and its communication within and across language borders.

A still under-explored area from the point of view of LT is healthcare, in all its various forms. TTS and ASR are at the core of a range of assistive technologies, helping to better integrate into society people with visual impairments and learning disabilities such as dyslexia or other cognitive as well as physical disabilities; incidentally, many of the relevant assistive devises and tools can also benefit Europe's increasingly aging population (e. g. for visual and hearing aids). Some speech technologies are used in diagnostics and cure of diseases related to speech understanding and production. Virtual cognitive assistants could drastically reduce the administrative burden and lead to improved patient experience and health outcomes. In general, speech and language technologies integrated in the emerging diagnostic tools, telemedicine, and social services can improve the well-being of everyone, and a fast increase in their use is expected in the near future; for this to happen in a fair, balanced and inclusive way, substantial progress is needed for most European languages.

5.4 LT Landscape in 2030

The central scientific goal envisioned for 2030 is Deep NLU for all European languages, including minority ones, as well as in the languages of Europe's partners. This goal describes and encompasses the numerous individual and/or goals in different dimensions:

- In research, development and innovation goals in the technology areas, supported by data and knowledge bases
- In covering and bridging the gaps and shortcomings identified in technology, use, regulation, and education
- In the legal and regulatory area, to facilitate the development of more effective and relevant LTs, in accordance with European values and contributing to strengthening them
- In the use (and the promotion of use) of LT and AI in government and public services
- Last but not least, in coordinating funding schemes among the EC and national or regional funding bodies to create an ecosystem of support that can lead to much-needed scientific breakthroughs, innovation and growth.

These goals are not completely new, at least in part; they have been endorsed by the European Parliament on several occasions, such as in the STOA Report "Language equality in the digital age – Towards a Human Language Project"[36] and especially the landmark EP Resolution "Language equality in the digital age".[37] The findings and results of the ELE project so far, as summarised in this report, have foregrounded the importance of these goals, and also identified more gaps, taking into account the recent developments in the computational and statistical foundations and advances in machine learning. Another major contribution has been received from analysing the views of the respondents and experts regarding the technology situation in LT and AI in 2030 and its role in society, resulting in the above list.

Given that the focus of the ELE project is DLE through technology, the forecasting focused on LT (and LT within and combined with additional AI technologies). Expert teams described their vision in the four Deep Dives. The vision has been summarised in this report in Section 3.2. Here we outline the main points of their vision for key LTs in 2030.

The priority research themes for NLU are MT, text analytics, speech and horizontally, data resources.

In MT, one of the most traditional LT applications that can be used directly by all types of users including ordinary citizens, the main features that are expected to be available by 2030 are awareness of context, including the environment ("metadata"), awareness of communication purpose as well as other translation requirements, ability to explain the translation decisions (through full NLU or other means), awareness of cultural diversity and, if appropriate, "transfer" and the presence of empathy with the users and their needs, if and as appropriate. These features will be all be available for both written and spoken translation systems while minimising the computing and space footprint, contributing also to the preservation of the environment.

In text processing and analytics, the main goal (aligned very closely to the overall NLU goal) is to extract knowledge, in all possible forms, from unstructured text. Research will

[36] https://www.europarl.europa.eu/stoa/en/document/EPRS_STU(2017)598621, March 2017
[37] https://www.europarl.europa.eu/doceo/document/TA-8-2018-0332_EN.html, Sep. 2018

progress on how to combine current human-generated and human-understandable knowledge (in the form of knowledge graphs, databases and other representations of world knowledge) with deep learning approaches and models. Basic text analysis tools will be capable of grounding, to connect the text with the aforementioned structured data. Today's tools will be integrated to provide a wealth of such information, for any type of text and domain, in all EU and other languages (minority, regional, possibly Europe's partners'), and leveraged by combining them with data from structured sources (databases). The size of data and text available for such analysis will grow, as will the speed of processing, allowing to analyse streamed data (including multimodal data).

In speech technologies, which are still evolving dramatically while bringing higher quality in more and more challenging environments, huge leaps are envisaged. However, these leaps are, in the view of the experts, also dependent on the progress in the other technology areas, namely MT and text analytics, to combine the audio signal analysis and generation with the possibilities offered by these two other areas. By 2030, speech technology will be integrated seamlessly not only in MT applications, but also with text analytics tools, allowing to process streamed and multilingual data in real time. Advances in Deep Learning will allow to increase the quality also for low-resource languages, and advances in audio signal processing will allow to increase audio quality also in increasingly more noisy environments. At the same time, audio signal processing technologies will allow for high quality, almost instant language identification, speaker identification and diarisation, speaker adaptation and discourse-related feature extraction in a form helping the overarching Deep NLU goal and novel applications. Advances will include sign language analysis and generation, teaming up with AI methods for image analysis and generation. For TTS, more variety and even more naturalness will be achieved to overcome adoption barriers.

In the area data and knowledge resources, a horizontal topic covering the three technology areas described above, the main goal is still to increase data availability in size (for large language models) and application suitability. However, with advances in deep learning which will decrease the need for large datasets in every language and every domain and application area, it will become more important to have these data available in an accessible way, at a zero or low and reasonable cost, shared among all stakeholders. The experts envision a network of repositories that preserve and distribute data of all kinds and types, under legally clear conditions and in an efficient way. In addition, ways will be found to create annotated data in a cheaper, more reliable and faster way if needed for supervised machine learning methods. Finally, speech and text data will be linked to other types of data, mainly structured and multimodal, in order to support the Deep NLU goal and novel research and applications in all areas of language-centric AI.

5.5 The Path to DLE in Europe by 2030: Key Recommendations

The surveys and expert interviews discussed here targeted LT developers, users and – equally importantly – the EU citizens. The surveys investigated language coverage, evaluated the current situation of LT in Europe and encouraged participants to share their predictions and visions for the future. More than 450 survey responses were collected and dozens of expert interviews were conducted. In addition, the EU citizen survey has been created so that citizens can provide their opinions regarding digital support for their languages. A very new approach to gather insight was the large-scale Citizen Survey which has already collected more than 21,000 responses. The answers broadly show that raising awareness for the LT potential in Europe on a political and institutional level is more important now than ever before. The European LT community is in a position where change is needed in order to compete with innovative systems and tools built elsewhere.

The analysis of the responses from all the surveys and interviews, as well as from the technology areas assessed by experts in the Deep Dives, resulted in recommendations described in detail (together with the supporting evidence) in Section 4. Here, we summarise the key recommendations that are likely to have major impact on driving forward the agenda of DLE for all European languages by 2030.

5.5.1 LT Developers

The key recommendations as extracted from the surveys and interviews with LT developers both from academia and from industry (Section 4.2.1) reflect the identified gaps, and take into account the visions where LT is going to be in 2030, ensuring, at the same time, DLE:

- Increase effort for collecting data across technologies, domains, and use cases
- Provide the data following the FAIR principles to ensure the broadest possible uptake
- Support basic research on LT/AI, especially in the following directions: full NLU, more efficient ML algorithms, algorithms and models avoiding bias, and tackling specifically (very) low-resourced languages
- Increase infrastructural support, both in terms of compute and services as well as data
- Support creating a network of closely collaborating national centres of excellence in LT/AI
- For public support, decrease bureaucracy especially for SMEs
- In academia, work on promoting FAIR data creation, annotation, preservation, and curation as a worthy and appreciated contribution to science
- Support talent retention in Europe
- Devise and/or adapt supportive regulation that favours the use of LT for accessibility and other similar purposes

5.5.2 LT Users

The recommendations as extracted from the surveys and interviews with LT users (Section 4.2.2) have some common ground with those listed above. However, the users have been looking at LT from a different angle, bringing new insights, gaps and shortcomings as seen from the users' perspective, thus producing an additional set of recommendations:

- Increase the variety of tools available for minority languages
- Increase the availability and quality of tools for all languages other than the biggest ones (English, German, Spanish, French, Italian)
- Develop real-time, collaborative, spoken MT for all language pairs
- Increase availability of human-readable and computer-processable lexical resources linked to available tools, especially for low-resource languages
- Introduce measures at all levels to keep language diversity, from education to public services to businesses, employing available LT

5.5.3 Technology Areas

The experts involved in preparing the Deep Dives (Bērziņš et al., 2022; Backfried et al., 2022; Gomez-Perez et al., 2022; Kaltenboeck et al., 2022) provided a very detailed analysis of the state of the art in MT, speech technologies and text analytics, including a data and knowledge infrastructural view. They identified a range of required breakthroughs, which are reflected in their recommendations.

In particular, breakthroughs needed for MT are related to system development (including interoperability, explainability, contextualisation, hardware needs and opportunities and opportunities offered in the future by quantum computing), data collection and EU policies, focusing on carbon-neutral and trustworthy AI-powered tools. Text processing and analytics tools need to focus on multilingual text processing and analytics needs to be strengthened. Another crucial element is benchmarking. For speech technology several development directions have been identified that contribute to DLE as well as to general progress: speech technologies integration, support for less-resourced languages, multimodal models, addressing the existing technological gaps, user and application contexts, development pace, training and evaluation etc. to name some of the issues identified as highest priorities.

From these gaps and identified breakthroughs needed for further progress as well as the experts' visions for the future with regard also to DLE in 2030, the following recommendations have been formulated:

- For MT research, support integration of speech for real-time, multi-agent and multi-language "instant" spoken MT among all EU languages
- Also especially for MT (but not only), support the creation of fundamentally new benchmarks and automated metrics that take into account DLE
- In speech, support a seamless integration of speech (ASR, TTS, SID) and downstream NLU/NLP in order to have intelligent systems, such as digital and virtual assistants, for all languages
- Support research in the direction of combining speech (and NLU/NLP) with other modalities, such as image and vision
- For ASR, support research on the digital audio signal and possibilities to address current limitations, such as noise in the environment
- As a common recommendation with the text analytics experts, support research in NLU which integrates speech, NLP and contextual information as well as additional modes of perception
- Support basic research in neurosymbolic approaches to NLP/NLU, including grounding and the use of human-understandable databases and sources
- Support the role of humans ("human in the loop") in LT/AI systems and applications
- Given the success of large language models for various applications, support the collection or acquisition or large datasets and train these general-purpose large language models for all EU languages, possibly mixed with other modalities

Separately, the "horizontal" Data and Knowledge Deep Dive resulted in additional recommendations, which also take into account certain eInfrastructural issues, such as compute and data storage.

In general, infrastructural support also needs to be significantly improved and extended. Today, large hardware infrastructures are required to accommodate for the necessary computation power and storage of Deep Neural Networks. Beside hardware infrastructures we see a clear need for a comprehensive and interconnected data infrastructure in place to achieve the specified objectives. To fill the identified gaps in data, language resources, and Knowledge Graphs we recommend and suggest a future path for Europe towards comprehensive and interlinked data infrastructures, considering the ELG the first foundational step in such a direction, heralding several promising and much-needed developments.

The key recommendations regarding both hardware facilities (such as data centres and HPCs) as well as the data and knowledge infrastructure (Section 4.4.1 and Section 4.4.2) can be summarised as follows:

- Increase the capacity of HPCs across Europe to cater for the needs of ML (e. g., include GPUs and provide simpler access to them), including staging large data for processing
- At the same time, support work on algorithms and general approaches that minimise the need for data and/or power supply for ML training
- Support interlinking, interoperability and sharing of metadata and FAIR[38] data in a transparent and open manner, in cooperation with major initiatives such as EOSC[39] as well as national projects and national funding in general
- Support the creation of a clear legal framework that allows data sharing and reuse, including for business development; this includes specific supportive regulations targeting the most widespread uses of LT, while preserving privacy

5.6 Towards Digital Language Equality

This report has provided the forward-looking vision and recommendations towards DLE in 2030, including a detailed analysis of the state of the art of LT and the related gaps and shortcomings. The results of WP1, with detailed analyses of the surveys and interviews conducted, as presented in a number of previous reports (WP1 Deliverables and Deliverables D2.2 to D2.16 and their main points summarised in D2.17) served as the main source of supporting evidence. In addition, this report has also been based on a large body of relevant literature, recent reports and studies related to LT and AI technology, societal impact and issues, business cases and general use of LT, with a special focus on Europe. In turn, this report will serve as one the main sources to develop the SRIA and roadmap as the main results of the ELE project.

References

Armen Aghajanyan, Xia Song, and Saurabh Tiwary. Towards language agnostic universal representations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4033–4041, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1395. URL https://aclanthology.org/P19-1395.

Nur Ahmed and Muntasir Wahed. The de-democratization of ai: Deep learning and the compute divide in artificial intelligence research. *arXiv preprint arXiv:2010.15581*, 2020. URL https://arxiv.org/abs/2010.15581.

Itziar Aldabe, Georg Rehm, German Rigau, , and Andy Way. Deliverable D3.1 Report on existing strategic documents and projects in LT/AI, 2021a. URL https://european-language-equality.eu/wp-content/uploads/2021/12/ELE___Deliverable_D3_1__revised_.pdf. Project deliverable; EU project European Language Equality (ELE); Grant Agreement no. LC-01641480 – 101018166 ELE.

[38] Findable, Accessible, Interoperable, Reusable
[39] European Open Science Cloud

Itziar Aldabe, Georg Rehm, and Andy Way. Report on existing strategic documents and projects in lt/ai, 2021b. URL https://european-language-equality.eu/wp-content/uploads/2021/05/ELE___Deliverable_D3_1.pdf.

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. An effective approach to unsupervised machine translation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 194–203, Florence, Italy, 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1019. URL https://aclanthology.org/P19-1019.

Gerhard Backfried, Marcin Skowron, Eva Navas, Aivars Bērziņš, Joachim Van den Bogaert, Franciska de Jong, Andrea DeMarco, Inma Hernaez, Marek Kováč, Peter Polák, Johan Rohdin, Michael Rosner, Jon Sanchez, Ibon Saratxaga, and Petr Schwarz. Deliverable D2.14 Technology Deep Dive – Speech Technologies, 2022. URL https://european-language-equality.eu/wp-content/uploads/2022/03/ELE___Deliverable_D2_14__Speech__Technologies.pdf. Project deliverable; EU project European Language Equality (ELE); Grant Agreement no. LC-01641480 – 101018166 ELE.

Emily M Bender. On achieving and evaluating language-independence in NLP. *Linguistic Issues in Language Technology*, 6, 2011.

Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 610–623, 2021.

Aivars Berzins, Khalid Choukri, Maria Giagkou, Andrea Lösch, Helene Mazo, Stelios Piperidis, Mickaël Rigault, Eileen Schnur, Lilli Small, Josef van Genabith, Andrejs Vasiljevs, Andero Adamson, Dimitra Anastasiou, Natassa Avraamides-Haratsi, Núria Bel, Zoltán Bódi, António Branco, Gerhard Budin, Virginijus Dadurkevicius, Stijn de Smeytere, Hrístina Dobreva, Rickard Domeij, Jane Dunne, Kristine Eide, Claudia Foti, Maria Gavriilidou, Thibault Grouas, Normund Gruzitis, Jan Hajic, Barbara Heinisch, Verónique Hoste, Arne Jönsson, Fryni Kakoyianni-Doa, Sabine Kirchmeier, Svetla Koeva, Lucia Konturová, Jürgen Kotzian, Simon Krek, Gauti Kristmannsson, Kaisamari Kuhmonen, Krister Lindén, Teresa Lynn, Armands Magone, Maite Melero, Laura Mihailescu, Simonetta Montemagni, Micheál Õ Conaire, Jan Odijk, Maciej Ogrodniczuk, Pavel Pecina, Jon Arild Olsen, Bolette Sandford Pedersen, David Perez, Andras Repar, Ayla Rigouts Terryn, Eirikur Rögnvaldsson, Mike Rosner, Nancy Routzouni, Claudia Soria, Alexandra Soska, Donatienne Spiteri, Marko Tadic, Carole Tiberius, Dan Tufis, Andrius Utka, Paolo Vale, Piet van den Berg, Tamás Váradi, Kadri Vare, Andreas Witt, Francois Yvon, Janis Ziedins, and Miroslav Zumrik. *Sustainable Language Data Sharing to Support Language Equality in Multilingual Europe - Why Language Data Matters: ELRC White Paper*. ELRC Consortium, 2 edition, 2019. ISBN 978-3-943853-05-6.

Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2021.

Samuel Bowman and George Dahl. What will it take to fix benchmarking in natural language understanding? In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4843–4855, 2021.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf.

Aivars Bērziņš, Mārcis Pinnis, Inguna Skadiņa, Andrejs Vasiļjevs, Nora Aranberri, Joachim Van den Bogaert, Sally O'Connor, Mercedes García–Martínez, Iakes Goenaga, Jan Hajič, Manuel Herranz, Christian Lieske, Martin Popel, Maja Popović, Sheila Castilho, Federico Gaspari, Rudolf Rosa, Riccardo Superbo, and Andy Way. Deliverable D2.13 Technology Deep Dive – Machine Translation, 2022. URL https://european-language-equality.eu/wp-content/uploads/2022/03/ELE___Deliverable_D2_13__Machine_Translation_.pdf. Project deliverable; EU project European Language Equality (ELE); Grant Agreement no. LC-01641480 – 101018166 ELE.

Xieling Chen, Di Zou, Haoran Xie, and Gary Cheng. Twenty years of personalized language learning: Topic modeling and knowledge mapping. *Educational Technology & Society*, 24(1):205–222, 2021. ISSN 11763647, 14364522. URL https://www.jstor.org/stable/26977868.

Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, Parker Schuh, Kensen Shi, Sasha Tsvyashchenko, Joshua Maynez, Abhishek Baindoor Rao, Parker Barnes, Yi Tay, Noam M. Shazeer, Vinodkumar Prabhakaran, Emily Reif, Nan Du, Benton C. Hutchinson, Reiner Pope, James Bradbury, Jacob Austin, Michael Isard, Guy Gur-Ari, Pengcheng Yin, Toju Duke, Anselm Levskaya, Sanjay Ghemawat, Sunipa Dev, Henryk Michalewski, Xavier García, Vedant Misra, Kevin Robinson, Liam Fedus, Denny Zhou, Daphne Ippolito, David Luan, Hyeontaek Lim, Barret Zoph, Alexander Spiridonov, Ryan Sepassi, David Dohan, Shivani Agrawal, Mark Omernick, Andrew M. Dai, Thanumalayan Sankaranarayana Pillai, Marie Pellat, Aitor Lewkowycz, Erica Oliveira Moreira, Rewon Child, Oleksandr Polozov, Katherine Lee, Zongwei Zhou, Xuezhi Wang, Brennan Saeta, Mark Diaz, Orhan Firat, Michele Catasta, Jason Wei, Kathleen S. Meier-Hellstern, Douglas Eck, Jeff Dean, Slav Petrov, and Noah Fiedel. Palm: Scaling language modeling with pathways. *ArXiv*, abs/2204.02311, 2022.

Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper/2017/file/d5e2c0adad503c91f91df240d0cd4e49-Paper.pdf.

Alexis Conneau and Guillaume Lample. Cross-lingual language model pretraining. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/file/c04c19c2c2474dbf5f7ac4372c5b9af1-Paper.pdf.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.747. URL https://aclanthology.org/2020.acl-main.747.

Sara Cooper, Alessandro Di Fava, Carlos Vivas, Luca Marchionni, and Francesco Ferro. Ari: the social assistive robot and companion. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 745–751, 2020. doi: 10.1109/RO-MAN47096.2020.9223470.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1423. URL https://aclanthology.org/N19-1423.

Carla Parra Escartín, Teresa Lynn, J. Moorkens, and Jane Dunne. Towards transparency in nlp shared tasks. *ArXiv*, abs/2105.05020, 2021.

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

Tianyu Gao, Adam Fisch, and Danqi Chen. Making pre-trained language models better few-shot learners. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3816–3830, Online, 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.acl-long.295. URL https://aclanthology.org/2021.acl-long.295.

Mahault Garnerin, Solange Rossato, and Laurent Besacier. Investigating the impact of gender representation in asr training data: a case study on librispeech. In *3rd Workshop on Gender Bias in Natural Language Processing*, pages 86–92. Association for Computational Linguistics, 2021.

Miguelángel Verde Garrido. Why a militantly democratic lack of trust in state surveillance can enable better and more democratic security. In *Trust and Transparency in an Age of Surveillance*, pages 221–240. Routledge, 2021.

José Manuél Gómez-Pérez, Ronald Denaux, and Andrés García-Silva. *A Practical Guide to Hybrid Natural Language Processing - Combining Neural Models and Knowledge Graphs for NLP*. Springer, 2020. ISBN 978-3-030-44829-5. doi: 10.1007/978-3-030-44830-1. URL https://doi.org/10.1007/978-3-030-44830-1.

Jose Manuel Gomez-Perez, Andres Garcia-Silva, Cristian Berrio, German Rigau, Aitor Soroa, Christian Lieske, Johannes Hoffart, Felix Sasaki, Daniel Dahlmeier, Inguna Skadiņa, Aivars Bērziņš, Andrejs Vasiļjevs, and Teresa Lynn. Deliverable D2.15 Technology Deep Dive – Text Analytics, Text and Data Mining, NLU, 2022. URL https://european-language-equality.eu/wp-content/uploads/2022/03/ELE___Deliverable_D2_15__Text_Analytics_.pdf. Project deliverable; EU project European Language Equality (ELE); Grant Agreement no. LC-01641480 – 101018166 ELE.

Ian J. Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, Cambridge, MA, USA, 2016. http://www.deeplearningbook.org.

Xu Han, Zhengyan Zhang, Ning Ding, Yuxian Gu, Xiao Liu, Yuqi Huo, Jiezhong Qiu, Liang Zhang, Wentao Han, Minlie Huang, et al. Pre-trained models: Past, present and future. *AI Open*, 2021.

Eran Hirsch, Alon Eirew, Ori Shapira, Avi Caciularu, Arie Cattan, Ori Ernst, Ramakanth Pasunuru, Hadar Ronen, Mohit Bansal, and Ido Dagan. iFacetSum: Coreference-based interactive faceted summarization for multi-document exploration. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 283–297, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.emnlp-demo.33. URL https://aclanthology.org/2021.emnlp-demo.33.

Pascal Hitzler, Federico Bianchi, Monireh Ebrahimi, and Md. Kamruzzaman Sarker. Neural-symbolic integration and the semantic web a position paper. In *Semantic Web, IOS Press*, 2019.

MD Zakir Hossain, Ferdous Sohel, Mohd Fairuz Shiratuddin, and Hamid Laga. A comprehensive survey of deep learning for image captioning. *ACM Computing Surveys (CsUR)*, 51(6):1–36, 2019.

Andrew Jaegle, Sebastian Borgeaud, Jean-Baptiste Alayrac, Carl Doersch, Catalin Ionescu, David Ding, Skanda Koppula, Daniel Zoran, Andrew Brock, Evan Shelhamer, et al. Perceiver io: A general architecture for structured inputs & outputs. *arXiv preprint arXiv:2107.14795*, 2021.

Melvin Johnson, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes, and Jeffrey Dean. Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation. *Transactions of the Association for Computational Linguistics*, 5:339–351, 10 2017. ISSN 2307-387X. doi: 10.1162/tacl_a_00065. URL https://doi.org/10.1162/tacl_a_00065.

Martin Kaltenboeck, Artem Revenko, Khalid Choukri, Svetla Boytcheva, Christian Lieske, Teresa Lynn, German Rigau, Maria Heuschkel, Aritz Farwell, Gareth Jones, Itziar Aldabe, Ainara Estarrona, Katrin Marheinecke, Stelios Piperidis, Victoria Arranz, Vincent Vandeghinste, and Claudia Borg. Deliverable D2.16 Technology Deep Dive – Data, Language Resources, Knowledge Graphs, 2022. URL https://european-language-equality.eu/wp-content/uploads/2022/03/ELE___Deliverable_D2_16_

_Data_and_Knowledge_.pdf. Project deliverable; EU project European Language Equality (ELE); Grant Agreement no. LC-01641480 – 101018166 ELE.

Tom Kocmi, Christian Federmann, Roman Grundkiewicz, Marcin Junczys-Dowmunt, Hitokazu Matsushita, and Arul Menezes. To Ship or Not to Ship: An Extensive Evaluation of Automatic Metrics for Machine Translation. In *Proceedings of the 6th Conference on Machine Translation (WMT 2021)*, 2021. URL https://arxiv.org/abs/2107.10821. 17pp.

Jiwei Li, Will Monroe, Alan Ritter, Dan Jurafsky, Michel Galley, and Jianfeng Gao. Deep reinforcement learning for dialogue generation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1192–1202, Austin, Texas, 2016. Association for Computational Linguistics. doi: 10.18653/v1/D16-1127. URL https://aclanthology.org/D16-1127.

Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. Multilingual denoising pre-training for neural machine translation. *Transactions of the Association for Computational Linguistics*, 8:726–742, 2020. doi: 10.1162/tacl_a_00343. URL https://aclanthology.org/2020.tacl-1.47.

Nitika Mathur, Timothy Baldwin, and Trevor Cohn. Tangled up in BLEU: Reevaluating the evaluation of automatic machine translation evaluation metrics. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4984–4997, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.448. URL https://aclanthology.org/2020.acl-main.448.

Philippe Palanque and Fabio Paternò. *Formal methods in Human-computer interaction*. Springer Science & Business Media, 2012.

Vassil Panayotov, Guoguo Chen, Daniel Povey, and Sanjeev Khudanpur. Librispeech: an asr corpus based on public domain audio books. In *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 5206–5210. IEEE, 2015.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of ACL 2002*, pages 311–318, Philadelphia, Pennsylvania, 2002.

Tae Jin Park, Naoyuki Kanda, Dimitrios Dimitriadis, Kyu J Han, Shinji Watanabe, and Shrikanth Narayanan. A review of speaker diarization: Recent advances with deep learning. *Computer Speech & Language*, 72:101317, 2022.

Matthew E. Peters, Mark Neumann, Robert Logan, Roy Schwartz, Vidur Joshi, Sameer Singh, and Noah A. Smith. Knowledge enhanced contextual word representations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 43–54, Hong Kong, China, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-1005. URL https://aclanthology.org/D19-1005.

Sarah Masud Preum, Sirajum Munir, Meiyi Ma, Mohammad Samin Yasar, David J. Stone, Ronald Williams, Homa Alemzadeh, and John A. Stankovic. A review of cognitive assistants for healthcare: Trends, prospects, and future directions. *ACM Comput. Surv.*, 53(6), feb 2021. ISSN 0360-0300. doi: 10.1145/3419368. URL https://doi.org/10.1145/3419368.

Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. Technical report, OpenAI, 2019.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision, 2021.

Jack W. Rae, Sebastian Borgeaud, Trevor Cai, Katie Millican, Jordan Hoffmann, Francis Song, John Aslanides, Sarah Henderson, Roman Ring, Susannah Young, Eliza Rutherford, Tom Hennigan, Jacob Menick, Albin Cassirer, Richard Powell, George van den Driessche, Lisa Anne Hendricks, Maribeth Rauh, Po-Sen Huang, Amelia Glaese, Johannes Welbl, Sumanth Dathathri, Saffron Huang, Jonathan Uesato, John F. J. Mellor, Irina Higgins, Antonia Creswell, Nathan McAleese, Amy Wu, Erich Elsen, Siddhant M. Jayakumar, Elena Buchatskaya, David Budden, Esme Sutherland, Karen Simonyan, Michela Paganini, L. Sifre, Lena Martens, Xiang Lorraine Li, Adhiguna Kuncoro, Aida Nematzadeh, Elena Gribovskaya, Domenic Donato, Angeliki Lazaridou, Arthur Mensch, Jean-Baptiste Lespiau, Maria Tsimpoukelli, N. K. Grigorev, Doug Fritz, Thibault Sottiaux, Mantas Pajarskas, Tobias Pohlen, Zhitao Gong, Daniel Toyama, Cyprien de Masson d'Autume, Yujia Li, Tayfun Terzi, Vladimir Mikulik, Igor Babuschkin, Aidan Clark, Diego de Las Casas, Aurelia Guy, Chris Jones, James Bradbury, Matthew G. Johnson, Blake A. Hechtman, Laura Weidinger, Iason Gabriel, William S. Isaac, Edward Lockhart, Simon Osindero, Laura Rimell, Chris Dyer, Oriol Vinyals, Kareem W. Ayoub, Jeff Stanway, L. L. Bennett, Demis Hassabis, Koray Kavukcuoglu, and Geoffrey Irving. Scaling language models: Methods, analysis & insights from training gopher. *ArXiv*, abs/2112.11446, 2021.

Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. Zero-shot text-to-image generation. *arXiv preprint arXiv:2102.12092*, 2021. URL https://arxiv.org/abs/2102.12092.

Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *ArXiv*, abs/2204.06125, 2022.

Morgane Riviere, Jade Copet, and Gabriel Synnaeve. Asr4real: An extended benchmark for speech models. *arXiv preprint arXiv:2110.08583*, 2021.

Rudolf Rosa, Ondřej Dušek, Tom Kocmi, David Mareček, Tomáš Musil, Patrícia Schmidtová, Dominik Jurko, Ondřej Bojar, Daniel Hrbek, David Košťák, Martina Kinská, Josef Doležal, and Klára Vosecká. Theaitre: Artificial intelligence to write a theatre play. In *Proceedings of AI4Narratives2020 workshop at IJCAI2020*, 2020.

Ori Shapira, Ramakanth Pasunuru, Hadar Ronen, Mohit Bansal, Yael Amsterdamer, and Ido Dagan. Extending multi-document summarization evaluation to the interactive setting. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 657–677, Online, June 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.naacl-main.54. URL https://aclanthology.org/2021.naacl-main.54.

Emily Sheng, Kai-Wei Chang, Premkumar Natarajan, and Nanyun Peng. Societal biases in language generation: Progress and challenges. In *Proceedings of the Conference of the 59th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2021.

David Snyder, Daniel Garcia-Romero, Gregory Sell, Daniel Povey, and Sanjeev Khudanpur. X-vectors: Robust dnn embeddings for speaker recognition. In *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 5329–5333. IEEE, 2018.

Titus Stahl. Indiscriminate mass surveillance and the public sphere. *Ethics and Information Technology*, 18(1):33–39, 2016.

Nisan Stiennon, Long Ouyang, Jeff Wu, Daniel M. Ziegler, Ryan Lowe, Chelsea Voss, Alec Radford, Dario Amodei, and Paul Christiano. Learning to summarize from human feedback, 2020.

STOA. Language equality in the digital age – Towards a Human Language Project. STOA study (PE 598.621), IP/G/STOA/FWC/2013-001/Lot4/C2, March 2017. Carried out by Iclaves SL (Spain) at the request of the Science and Technology Options Assessment (STOA) Panel, managed by the Scientific Foresight Unit (STOA), within the Directorate-General for Parliamentary Research Services (DG EPRS) of the European Parliament, March 2017. http://www.europarl.europa.eu/stoa/.

Emma Strubell, Ananya Ganesh, and Andrew McCallum. Energy and policy considerations for deep learning in nlp. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3645–3650, Florence, Italy, 2019a.

Emma Strubell, Ananya Ganesh, and Andrew McCallum. Energy and policy considerations for deep learning in nlp. *arXiv preprint arXiv:1906.02243*, 2019b.

Éva Székely, Gustav Eje Henter, Jonas Beskow, and Joakim Gustafson. Spontaneous conversational speech synthesis from found data. In *INTERSPEECH*, pages 4435–4439, 2019.

Katrin Tomanek, Françoise Beaufays, Julie Cattiau, Angad Chandorkar, and Khe Chai Sim. On-device personalization of automatic speech recognition models for disordered speech. *arXiv preprint arXiv:2106.10259*, 2021.

Amirsina Torfi, Rouzbeh A Shirvani, Yaser Keneshloo, Nader Tavvaf, and Edward A Fox. Natural language processing advancements by deep learning: A survey. *arXiv preprint arXiv:2003.01200*, 2020. URL https://arxiv.org/abs/2003.01200.

Zoltán Tüske, George Saon, and Brian Kingsbury. On the limit of english conversational speech recognition. *CoRR*, abs/2105.00982, 2021. URL https://arxiv.org/abs/2105.00982.

Eva Vanmassenhove, Dimitar Shterionov, and Andy Way. Lost in translation: Loss and decay of linguistic richness in machine translation. In *Proceedings of Machine Translation Summit XVII: Research Track*, pages 222–232, Dublin, Ireland, August 2019. European Association for Machine Translation. URL https://aclanthology.org/W19-6622.

Ville Vestman, Tomi Kinnunen, Rosa González Hautamäki, and Md Sahidullah. Voice mimicry attacks assisted by automatic speaker verification. *Computer Speech & Language*, 59:36–54, 2020.

Ruize Wang, Duyu Tang, Nan Duan, Zhongyu Wei, Xuanjing Huang, Jianshu Ji, Guihong Cao, Daxin Jiang, and Ming Zhou. K-Adapter: Infusing Knowledge into Pre-Trained Models with Adapters. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1405–1418, Online, 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.findings-acl.121. URL https://aclanthology.org/2021.findings-acl.121.

Andy Way, Georg Rehm, Jane Dunne, Jan Hajič, Teresa Lynn, Maria Giagkou, Natalia Resende, Tereza Vojtěchová, Stelios Piperidis, Andrejs Vasiljevs, Aivars Berzins, Gerhard Backfried, Marcin Skowron, Jose Manuel Gomez-Perez, Andres Garcia-Silva, Martin Kaltenböck, and Artem Revenko. Deliverable D2.17 Report on all external consultations and surveys, 2022. URL https://european-language-equality.eu/wp-content/uploads/2022/04/ELE___Deliverable_D2_17__Report_on_External_Consultations_-2.pdf. Project deliverable; EU project European Language Equality (ELE); Grant Agreement no. LC-01641480 – 101018166 ELE.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online, 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-demos.6. URL https://aclanthology.org/2020.emnlp-demos.6.

Katalin Zsiga, András Tóth, Tamás Pilissy, Orsolya Péter, Zoltán Dénes, and Gábor Fazekas. Evaluation of a companion robot based on field tests with single older adults in their homes. *Assistive Technology*, 30(5):259–266, 2018. URL https://doi.org/10.1080/10400435.2017.1322158.

Appendix – EU Citizen Survey Additional Material

Countries	Sample size	Total responses	Completed	Ratio	%
Austria	900	965	900	1	100.0
Belgium – Dutch	500	538	500	1	100.0
Belgium – German	50	57	50	1	100.0
Belgium – French	350	380	350	1	100.0
Bulgaria	750	797	750	1	100.0
Croatia	600	643	600	1	100.0
Czech	900	977	920	1.02	102.2
Denmark	600	677	600	1	100.0
Estonia	150	170	150	1	100.0
Finland – Swedish	50	55	50	1	100.0
Finland – Finnish	250	271	250	1	100.0
France	900	1003	901	1	100.1
Germany	900	1008	901	1	100.1
Greece	900	972	900	1	100.0
Hungary	900	934	900	1	100.0
Ireland	450	532	473	1.05	105.1
Ireland – Irish	100	173	117	1.17	117.0
Italy	900	958	900	1	100.0
Latvia	200	209	200	1	100.0
Lithuania	300	321	300	1	100.0
Netherlands	900	959	900	1	100.0
Norway	600	638	600	1	100.0
Poland	900	1026	900	1	100.0
Portugal	900	995	900	1	100.0
Romania	900	950	900	1	100.0
Serbia	100	95	100	1	100.0
Slovakia	550	604	550	1	100.0
Spain	750	820	748	0.99	99.7
Spain – Galician	50	82	50	1	100.0
Spain – Basque	50	72	50	1	100.0
Spain – Catalan	100	126	100	1	100.0
Sweden	900	747	900	1	100.0
Switzerland – Italian	50	64	50	1	100.0
Switzerland – German	200	389	200	1	100.0
Switzerland – French	150	171	150	1	100.0
UK	842	973	842	1	100.0
UK – Wales	58	100	58	1	100.0
Slovenia	250	399	253	1.012	101.2
Totals	18900	20850	18963		

Table 2: Number of responses through our service provider per country and language

Languages	Totals
Basque	147
Bosnian	157
Bulgarian	47
Catalan	79
Croatian	19
Czech	43
Danish	55
Dutch	35
English	228
Estonian	58
Finnish	49
French	48
Galician	172
German	121
Greek	48
Hungarian	47
Icelandic	134
Irish	126
Italian	81
Latvian	14
Lithuanian	74
Luxembourgish	4
Macedonian	61
Maltese	79
Norwegian	19
Polish	13
Portuguese	19
Romanian	13
Serbian	12
Slovakian	29
Slovenian	59
Spanish	32
Swedish	35
Turkish	42
Welsh	224
Total	2423

Table 3: Number of responses through ELE dissemination channels (as of 29 April 2022)