



# LANGUAGE TECHNOLOGY LANDSCAPE CONFERENCE

**CHARTING THE  
GLOBAL LANDSCAPE OF  
LANGUAGE TECHNOLOGY**

**June 18, 2024**

**A EUROPEAN COMMISSION INITIATIVE**

**BY NIMDZI INSIGHTS**



LANGUAGE  
TECHNOLOGY  
LANDSCAPE  
CONFERENCE

# Speech-based LTs

## ASR, STT, MI/S2ST, Chatbots

Laszlo K. Varga, Nimdzi Insights  
Igor Szoke, Brno University of Technology  
Khalid Coukri, ELDA



LANGUAGE  
TECHNOLOGY  
LANDSCAPE  
CONFERENCE

# Housekeeping

Q&A



In-session polls



Recording



Post-event  
feedback survey





# Agenda

Time	Session
10:00-10:30	<b>Welcome &amp; Keynotes</b> Hosts: Philippe Gelin, Laszlo K. Varga Keynotes: Renate Nikolay, Renato Beninatto
10:30-11:25	<b>The Landscape of Language Technology</b> Speakers: Laszlo K. Varga
11:30-12:00	<b>Multilingualism of European Websites and the Technology Solutions Supporting It</b> Speakers: Andrejs Vasiljevs
12:00-13:00	<b>Lunch Break</b>
13:30-14:20	<b>Large Language Models and Foundational Language Technologies</b> Speakers: Laszlo K. Varga, Nadezda Jakubkova
14:25-15:15	<b>Text-based Language Technologies</b> Speakers: Laszlo K. Varga, Jourik Ciesielski
15:15-15:35	<b>Break</b>
15:35-16:25	<b>Speech-Based Language Technologies</b> Speakers: Laszlo K. Varga, Igor Szoke, Khalid Choudry
16:30-17:00	<b>Closing</b> Speakers: Philippe Gelin, Laszlo K. Varga



# Agenda

Time	Session
10:00-10:30	<b>Welcome &amp; Keynotes</b> Hosts: Philippe Gelin, Laszlo K. Varga Keynotes: Renate Nikolay, Renato Beninatto
10:30-11:25	<b>The Landscape of Language Technology</b> Speakers: Laszlo K. Varga
11:30-12:00	<b>Multilingualism of European Websites and the Technology Solutions Supporting It</b> Speakers: Andrejs Vasiljevs
13:30-14:20	<b>Large Language Models and Foundational Language Technologies</b> Speakers: Laszlo K. Varga, Nadezda Jakubkova
14:25-15:15	<b>Text-based Language Technologies</b> Speakers: Laszlo K. Varga, Jourik Ciesielski
15:15-15:35	<b>Break</b>
15:35-16:25	<b>Speech-Based Language Technologies</b> Speakers: Laszlo K. Varga, Igor Szoke, Khalid Choudry
16:30-17:00	<b>Closing</b> Speakers: Philippe Gelin, Laszlo K. Varga

# Speech-based language technologies



## Laszlo K. Varga

Lead Researcher and Analyst,

Nimdzi Insights



## Igor Szoke

Post-doc Researcher

Brno University of Technology



## Khalid Choukri

CEO

ELDA (Evaluations and Language resources Distribution Agency)

# Speech-based language technologies

## Introduction.

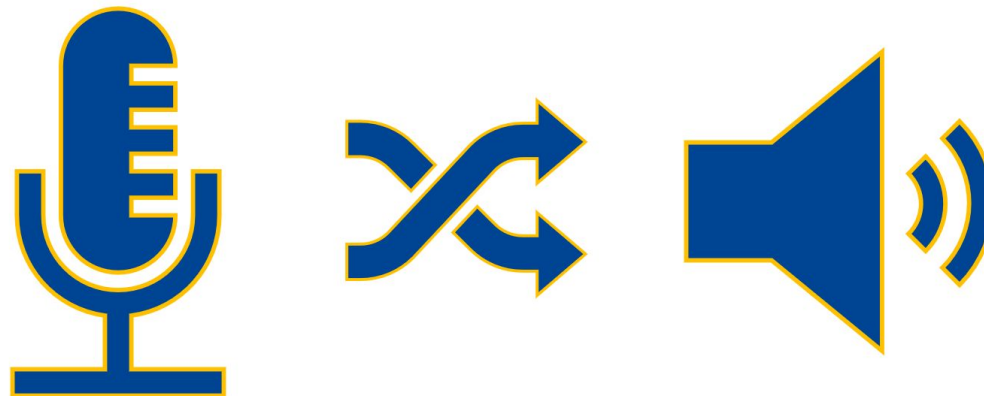
Speech LTs are a meta-category where spoken words are either input or output of the technology. With the multimodal digital life, audio (and video) communication has become ubiquitous in social media, in online meetings, web conferences, podcasts, entertainment and is expected to grow in significance.

This session also includes insights from the ASR data market research.

## Categories explained.

In this section, the following LT categories were grouped into speech-based LTs for convenience:

- Speech-to-text (STT) or automated speech recognition (ASR)
- Text-to-speech (TTS)
- Machine interpreting (MI) and speech-to-speech translation (S2ST)
- Chatbots and voicebots, voice assistants



# Automated speech recognition (ASR) 1

Main LT category	Speech-based technologies (ASR, STT)
Market size estimate (2023)	EUR 3-5 billion (EUR 20-30 billion including professional services)
Growth potential	High
Investment interest	Moderate
Market character	Three levels: 1. Core tech is well-established, big-tech dominated, commoditized 2. Custom enterprise solution providers, professional-services heavy 3. Base of composite LTs and services (MT, MI/S2ST, Chatbots, Speech NLP)
AI / ML adoption / disruption level	Fundamental
Technology maturity level	Evolving

## Introduction. Demand.

ASR and transcription is widely used by businesses and other organisations worldwide. ASR is also a core component to downstream tasks. Usage scenarios include online meetings, interviews such as in healthcare, legal, or HR settings, dictation, customer service and help desk calls. Demand for ASR technology has been increasing and will continue to do so:

- The explosion of audio-visual content created and distributed over the internet.
- Technological advancements that have improved quality
- Increased adoption in electronic devices
- Accessibility and DEI

## Market size and character.

The speech-based technology market is estimated around EUR 3-5 billion from the revenues of main actors, factoring in long-tail suppliers and big-tech companies' dominating presence. However, the market size including services is estimated to be 5-6 times that of technology alone.



# Automated speech recognition (ASR) 2

## ASR demand.

- by languages offered
- by use-case / specialisation
- generic models VERSUS specific and niche applications.

## Market segmentation.

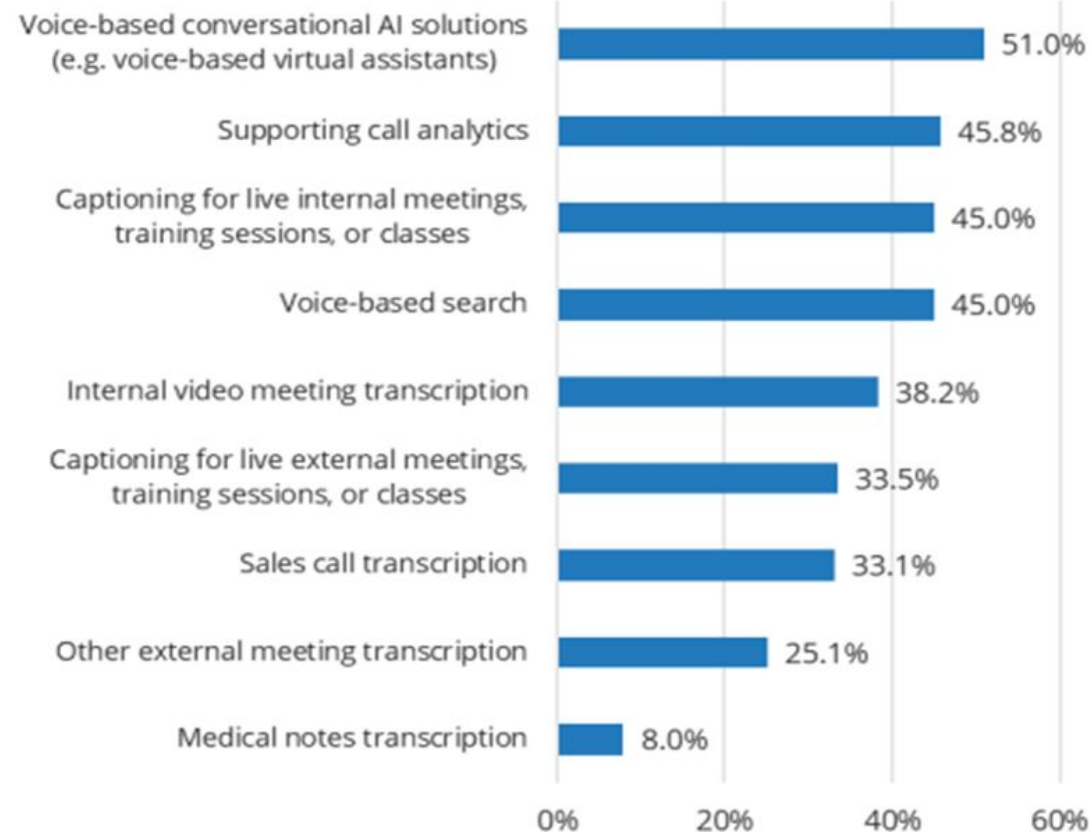
### By business sectors

- High tech & telecom & government authorities
- Banking, financial services & insurance
- Retail
- Automotive & transportation
- Healthcare
- Other (education, legal, ...)

### By organizations' profiles

- Large enterprises
- Small and medium-sized enterprises (SMEs)
- Public sector / government
- Consumer / individual user

## For which of the following use cases is your organization using Speech to Text?



*Deployment of ASR solution per use case from IDC 2022 report*

# Automated speech recognition (ASR) 3

## Main actors.

Big-tech dominated, commoditized. Key big tech players include Google, Microsoft, Amazon, Nvidia, Baidu, Tencent, Alibaba. Their solutions are typically used in API calls by various applications within or outside their cloud infrastructures. Pricing of such ASR solutions is closely matching across providers.

Challengers: OpenAI with Whisper, AppTek, Assembly.ai, DeepGram, Omilia, Rev, Phonexia, Verbit, and others. Even ASR developers use big-tech's base models to complement language coverage.

## Quality and technology outlook.

WER-based benchmarking, but need individual evaluation. Wildly varying between languages, depending on pre-training. Generally good quality base models (incl open source), low-resource languages challenged.

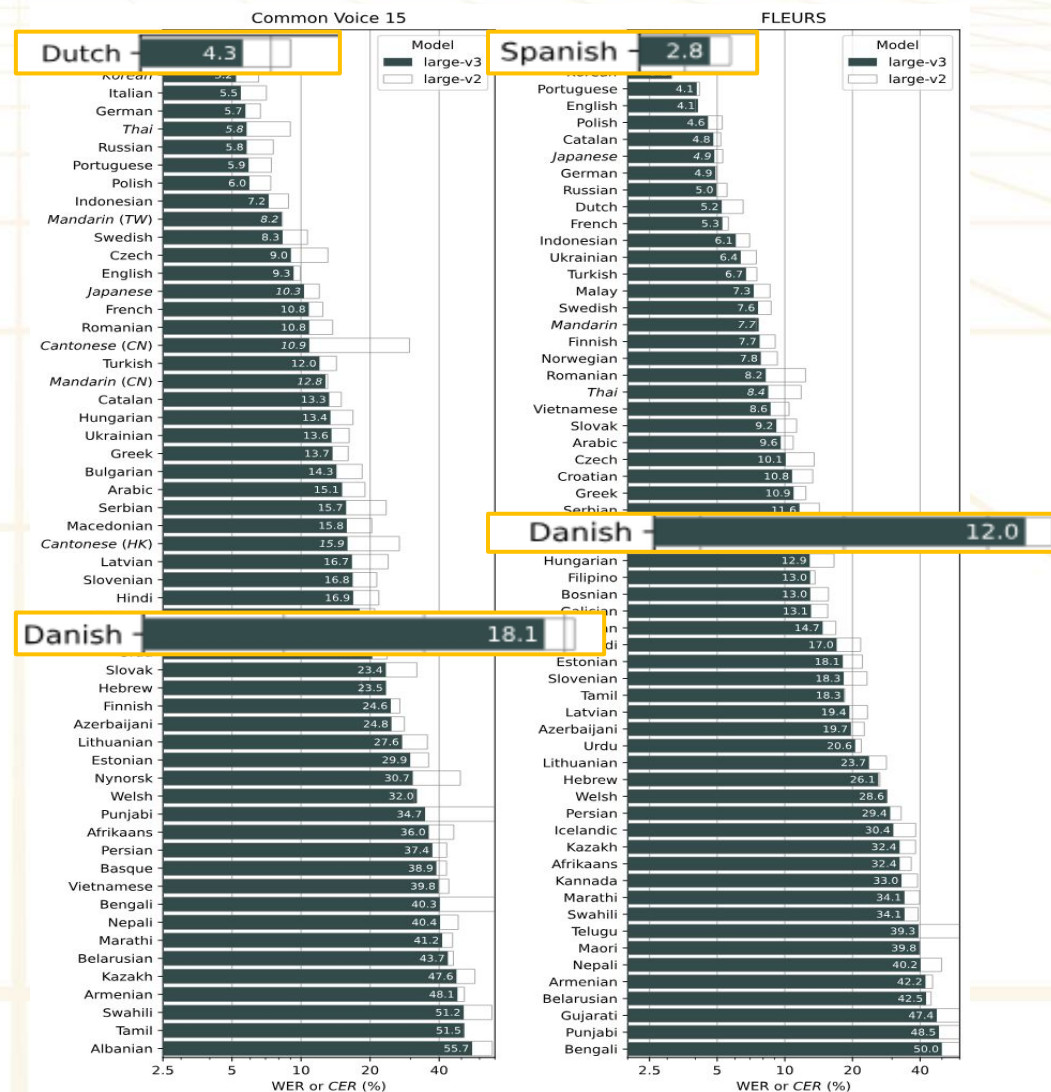
Advanced direct speech NLP features such as speaker and intent recognition, direct emotional and sentiment analysis is a frontier. Multilingual ASR and speech translation models are also being researched.

Company	Country of origin	Languages supported	Estimate ASR revenue (2023)	Investment / funding (till March, 2024)
Alibaba	CN	11, non-EU24	(not core)	(not core)
Amazon AWS	US	around 100, non-EU24 (noIrish)	(not core)	(not core)
AppTek	US	46, non-EU24 (no Irish)	EUR 20 million	undisclosed
Assembly.ai	US	20 for Best, 102 for Nano, non-EU24	EUR 20 million	EUR 50 million
Baidu	CN	4, non-EU24	(not core)	(not core)
DeepGram	US	33, non-EU24	EUR 30 million	EUR 100 million
Google	US	100+, all EU24	(not core)	(not core)
IBM	US	14 + flavours, non-EU24	(not core)	(not core)
iFlyTek	CN	15, non-EU24	(not core)	(not core)
Microsoft	US	100+, all EU24	n/a - non-core	(not core)
Nuance	US	86, non-EU24	(not core)	(part of Microsoft)
OpenAI	US	57, non-EU24	undisclosed	EUR 10 billion
Rev	US	60+, non-EU24	EUR 50 million	EUR 30 million
Sonix	US	49, non-EU24	EUR 10 million	EUR 1 million
Speechmatics	UK	30, non-EU24	EUR 20 million	EUR 80 million
Verbit	US	5, non-EU24	EUR 80 million (incl. services)	EUR 500 million
Phonexia	CZ	60+, non-EU24	EUR 10 million	undisclosed

# Automated speech recognition (ASR) 4

On language coverage  
and language quality

Not all languages are equal.



Source: OpenAI Whisper gitHub page. Highlight: Danish.



# Automated speech recognition (ASR) 5

## Market and technology evolution

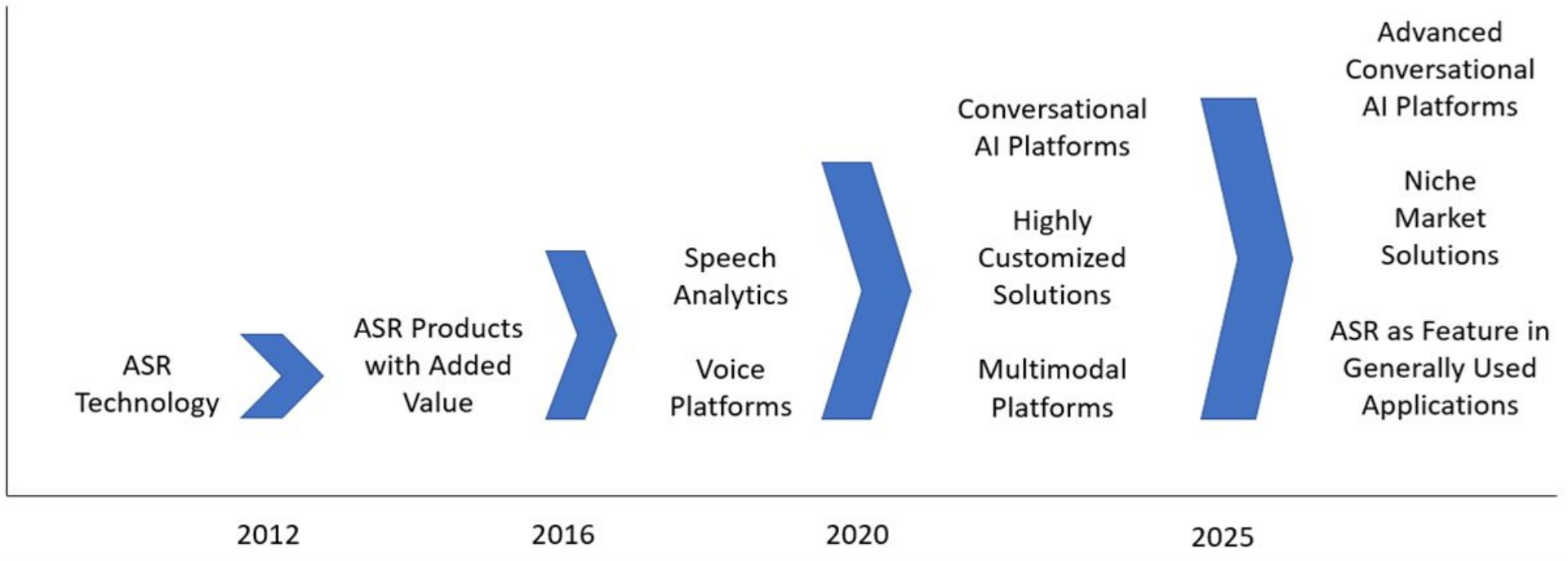


Illustration of the ASR market transformation

# Automated speech recognition (ASR) 6

There is a large market for specific, niche solutions...



Whisper Large



Whisper Large

# Automated speech recognition (ASR) 6

There is a large market for specific, niche solutions...



Whisper Large

ATCO: RY4Z, altitude 4000ft, QNH 1008  
Pilot: **ATC talking to the radio**

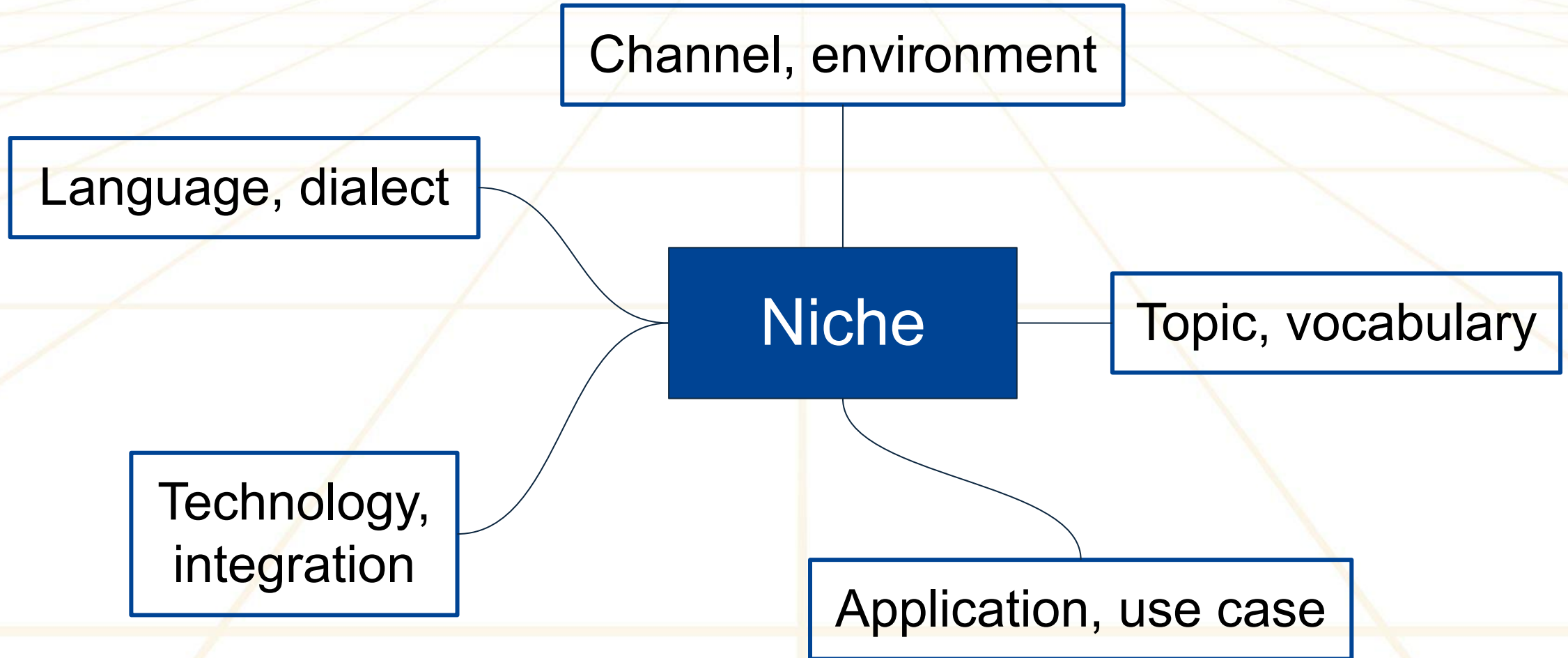


Whisper Large

ATCO: **ATC talking to the radio**



# Automated speech recognition (ASR) 7



# Automated speech recognition (ASR) 8

## Technology insights

### Data is the oil of the AI era.

- Who controls the data?

### ASR will be tightly connected to LLMs.

- Better understanding of the speech.
- Enables implicit high level tasks like search, summary, extractions, etc.

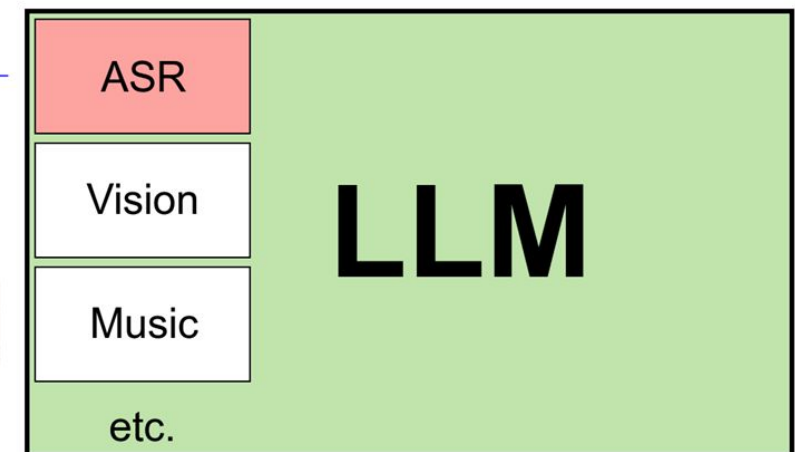
### Winner takes all.

- **Single universal world wide model**
  - Accurate
  - Robust
  - Freely available
- You **fine-tune** it on your “niche”.

NOW



FUTURE





# ASR data & market 1 - history

	Volume of data collected	Source of data
<b>Before 2000</b>	10-50 hours	mostly in-house
<b>In 2000s</b>	100-1000 hours	large co-funded projects from data vendors / producers / Darpa
<b>Since 2015</b>	10,000+ hours	large co-funded projects “open” sources; Internet / web crawling (IPR?)
<b>Nowadays</b>	100,000+ hours	From big institutions, such as EUPARL, but also broadcasters, Youtube

## Recent ASR data examples.

- OpenAI Whisper 680K hours on 2022,
- META like 400K (Voxpopuli of META only about 2K transcribed)
- Multilingual LibriSpeech Dataset (about 44.5K hours of English and a total of about 6K hours for other languages.)
- Mostly major languages

## Experts:

Khalid Choukri (ELDA, FR)  
Radim KUDLA (BUT, CZ)  
Inguna Skadiņa (Tilde, LV)  
Igor Szoke (BUT, CZ)

# ASR data & market 2

## Examples of EC co-funded Language resources.

### SpeechDat Family



2014 - 202X

> ELRC (European Language Resources Coordinations)

Over **3586** Language Resources

--- Textual resources mostly bilingual (2600)

# ASR data & market 3

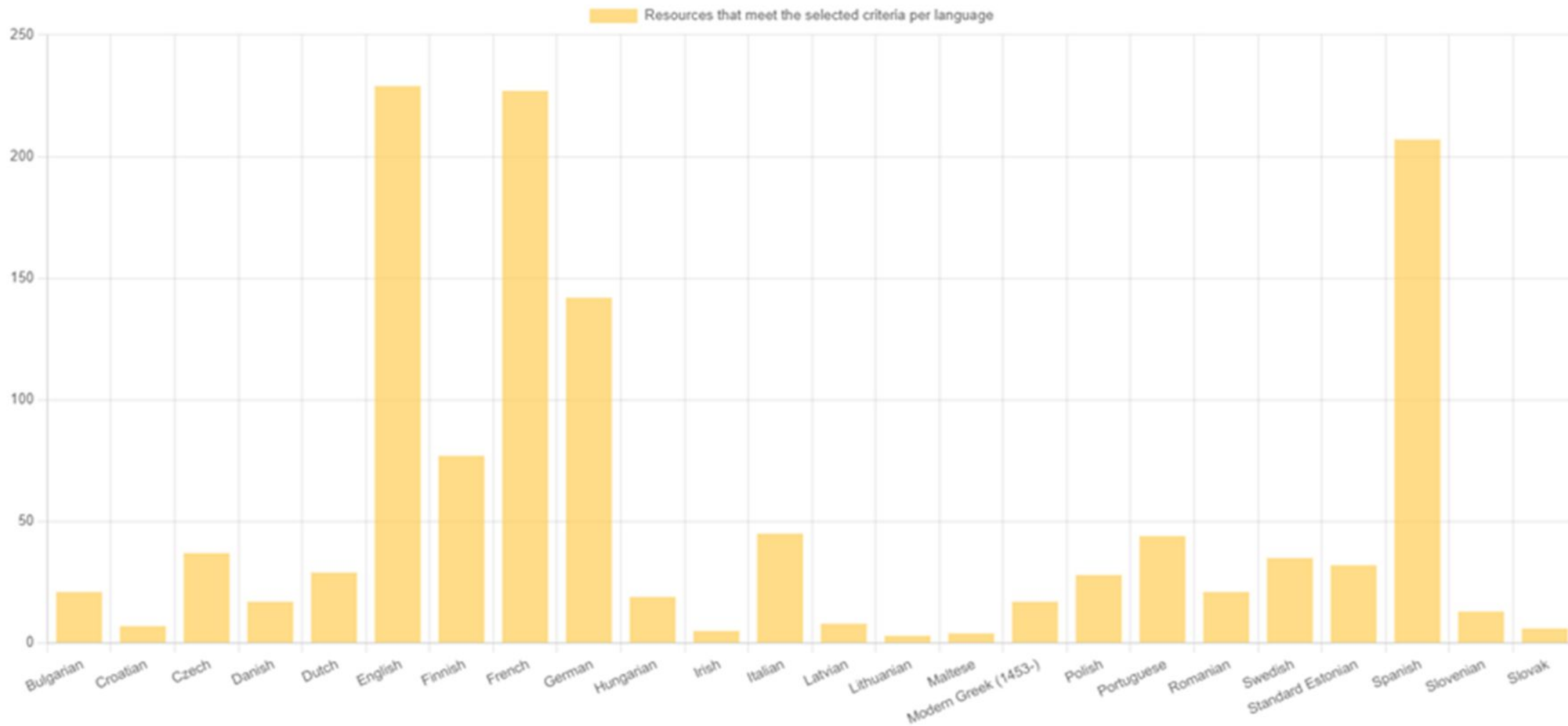
## Language Resources ... Some major sources

- ELRA, ELRA Language Resources Association, <http://www.elra.info/en/>
- LDC, Linguistic Data Consortium, <https://www ldc.upenn.edu/>
- CLARIN, the research infrastructure for language as social and cultural data, <https://www.clarin.eu/>
- SpeechOcean, <https://en.speechocean.com/>
- Data Tang, <https://www.datatang.ai/>
- ZENODO, <https://zenodo.org/>
- Appen, <https://appen.com/>
- META-SHARE, <http://www.meta-share.org/> and now the ELG platform, <https://live.european-language-grid.eu/>
- OLAC (a catalog of catalogs), the Open Language Archives Community, <http://www.language-archives.org/>
- but also Google Data Search, Huggingface etc.



# ASR data & market 4 - data in EU

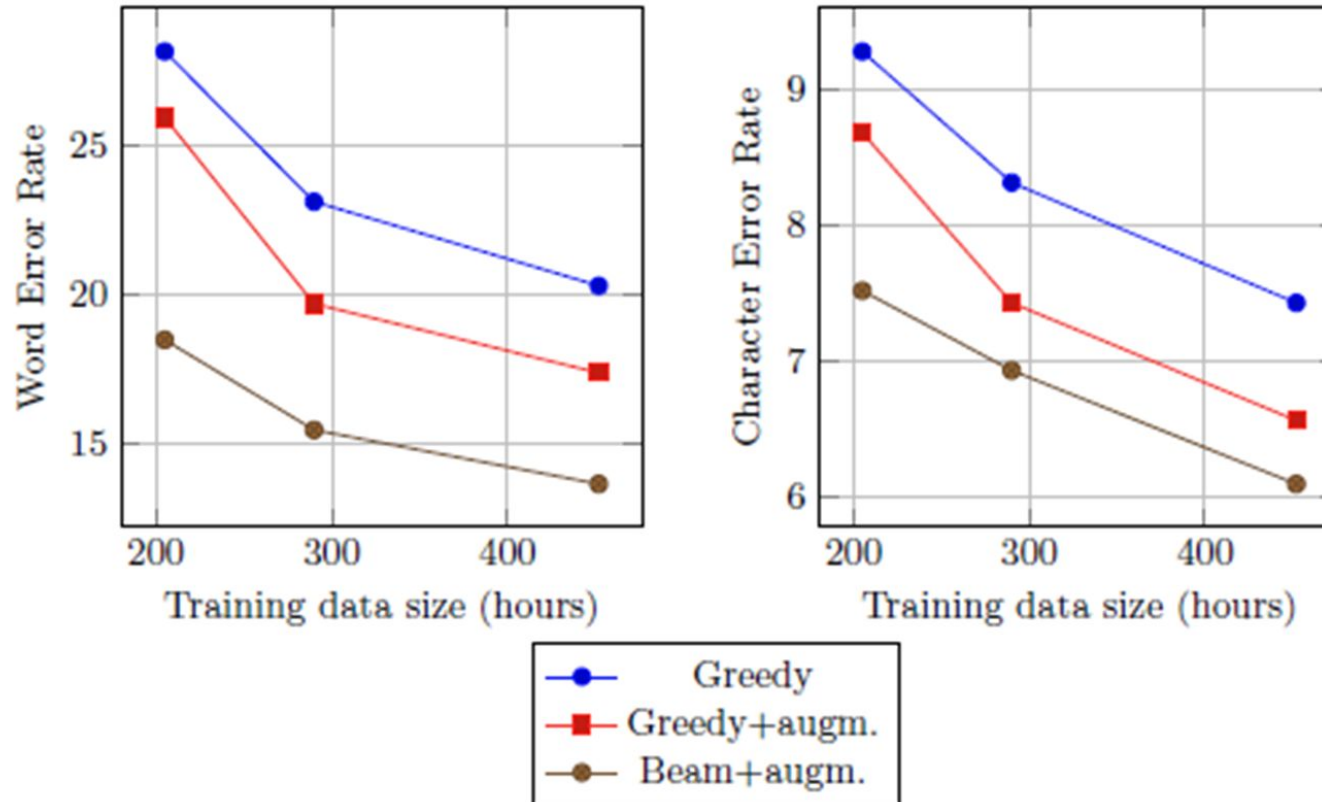
Number of resources





# ASR data & market 5

Quantity and quality of the data needed



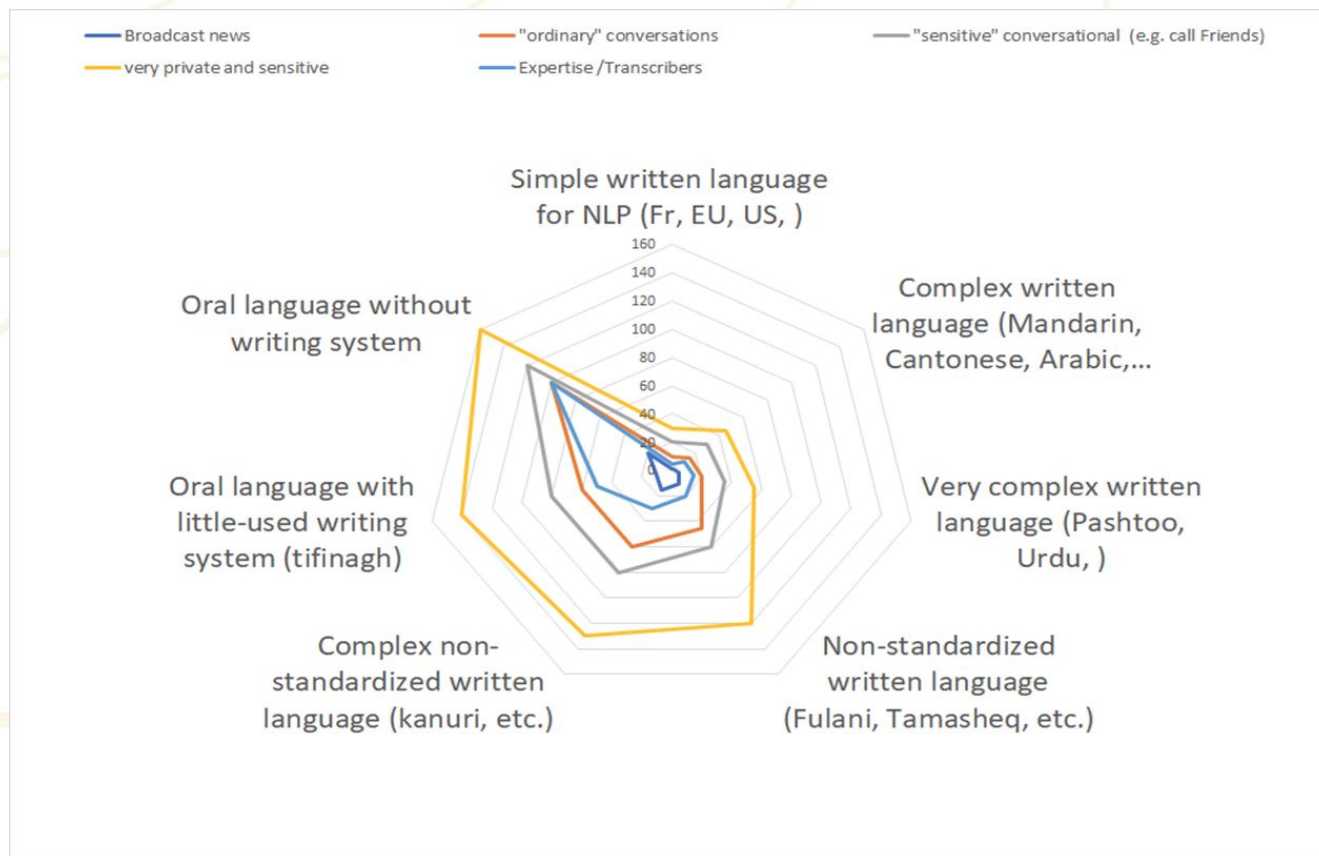
# ASR data & market 6

## Production process: Supervised data preparation

- Produce 100h of transcribed audio with high quality (low word error rate)
- Train the ASR modules (ASR v0.1)
- Use ASR v0.1 to pre-transcribe another 100h and assess quality
- Revise the transcriptions (Revisions)
- Redo transcription if quality is low
- retrain ASR (v0.2)
- assess performance (if adequate WER then stop)
- Otherwise redo the whole process till 1000h but usually more is needed
- Check performance and loop

# ASR data & market 7

## Production process: Supervised data preparation



### Complexity of data production depends on factors such as:

- Language
- Domain
- Media (interview, meeting, etc)
- Source noise, overlaps
- Data and privacy regulations
- Transcribers' proficiency
- If the writing system is standardised

### 1h of data requires about:

- **French:** 12-20 hours of labor
- **Tamasheq:** 60-70 hours of labor

# ASR data & market 8 - challenges

## Privacy, legal, and security issues

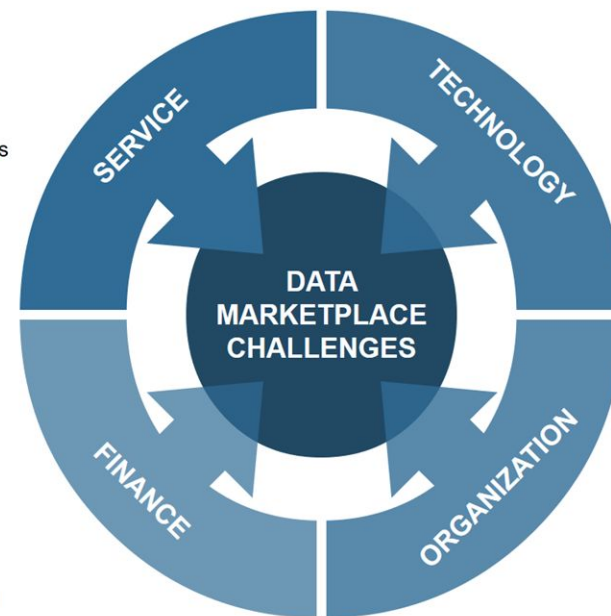
- Comply with all EU regulations (IPR, GDPR, Data Act, ....)
- Ensure compliance with the requirements of the technology developers:
  - Biases handling, gender/age/accent/etc. specifics wrt applications
  - Domains and modalities (e.g. language genre)

## Major challenges as depicted by the ELE project

### CHALLENGES OF DATA MARKETPLACES


- **Data ownership definition**
- Ensuring data integrity
- Assessing data quality
- Ensuring contractual compliances
- Loss of control over data
- Lack of transparency

- Pricing mechanism
- Data valuation
- Profit maximization



- Privacy protection
- Security
- **Technical efficiencies and scalabilities**
- Data placement cost
- **User-friendly applications and interfaces**

- **The absences of legal frameworks**
- **Lack of resources and technical knowledge**
- **Unclear organizational structure**
- **Ethical concern**

 Under-researched area



# Text-to-speech 1

Main LT category	Text-to-speech (TTS) including AI dubbing
Market size estimate (2023)	EUR 2-5 billion
Growth potential	Very high; \$5-10 billion in 2024
Investment interest	High
Market character	Three levels: 1. Core tech is well-established, big-tech dominated, commoditised 2. Innovative startups for new features such as voice cloning 3. Base of composite LTs and services (Chatbots, MI/S2ST)
AI / ML adoption / disruption level	Fundamental
Technology maturity level	Mature (core TTS), Evolving (AI dubbing)

## Introduction. Demand.

TTS is a “final output” producer - the speech created is almost exclusively for direct consumption. While single speaker settings (such as elearning or marketing videos/audios) are most suitable, AI dubbing takes TTS to the next level with additional engineering. Demand for TTS technology has been increasing and will continue to do so:

- Direct: Entertainment, e-learning, training, social media / content creation, voice-enabled devices.
- Indirect: LTs such as S2ST, machine interpreting, chatbots, and AI avatars.

## Market size and character.

We estimate the current market size to be close to EUR 5 billion. Although the market is dominated by big-tech solutions, as the vast majority of audio and video content is created in a single language in exponentially increasing volumes, the market for these technologies has a very high potential to grow in the next few years, also due to the the compound need for TTS (in voicebots, dubbing, machine interpreting, etc).

# Text-to-speech 2

## Main actors.

Big-tech dominated, commoditized (incl. Chinese big-tech).

Challengers: Adobe and OpenAI and AppTek, Colossyan, Elevenlabs, Papercup, and various dubbing-focused companies.

Differentiators are core targeted markets, synthetic vs human (cloned) voice.

Languages coverage are similar as with ASR, with few solutions supporting EU24 languages.

## Quality and technology outlook.

The assessment of the quality of speech output is not trivial to standardise, especially because the perception of quality of speech is highly subjective. The most commonly used quality metric for end-to-end speech translation is user feedback (Likert scale).

TTS coined an “easier problem than ASR”, Improvements are needed and expected in the area of prosody, intonation, and emotional expression, especially for synthetic voice acceptance. Lip syncing to video is the next frontier in AI dubbing.

Company	Country of origin	Languages supported	Estimate ASR revenue (2023)	Investment / funding (till March, 2024)
Microsoft	US	130, all EU24	n/a - non-core	n/a - big tech
IBM	US	14 dialects of 9 languages	n/a - non-core	n/a - big tech
Amazon	US	38, non-EU24	n/a - non-core	n/a - big tech
Google	US	>50, non-EU24	n/a - non-core	n/a - big tech
Nvidia	US	6, non-EU24	n/a - non-core	n/a - big tech
Apptek	US	13	EUR 20 million	n/a
Colossyan	HU, UK	>70, all EU24	EUR 7 million	EUR 25 million
Deepdub	IL, US	>80 languages, 130 dialects	EUR 10-15 million	EUR 20 million
Dubdub	IN	55, (likely) non-EU24	EUR 1 million	
Dubverse	IN	32, non-EU24	<EUR 1 million	<EUR 1 million
ElevenLabs	US	28, non-EU24	EUR 10 million	>EUR 100 million
Murf.ai	US	20, non-EU24	EUR <5 million	EUR 10 million
Ollang	US	60+ (details unavailable)	EUR 6 million	
Papercup	UK	“Any language” (details unavailable)	EUR 25 million	>EUR 20 million
Resemble.ai	CA	>60, non-EU24	EUR 2-3 million	EUR 8 million
Speechify	US	28	EUR 10 million	-
Voiseed	IT	5-10	<EUR 1 million	<EUR 1 million
WellSaid Labs	US	1 (English)	EUR 10 million	EUR 10 million
Wavel.ai	SG	>70	EUR 5 Million	
XL8 MediaCat	US		EUR 5-10 million	EUR 11 million

# Machine interpreting and S2ST 1

Main LT category	MI and S2ST
Market size estimate (2023)	<EUR 1 billion (both MI and S2ST)
Growth potential	High; EUR 1-2 billion in 2024 (both MI and S2ST)
Investment interest	Moderate
Market character	Fundamental
AI / ML adoption / disruption level	Fundamental
Technology maturity level	Emerging (direct S2ST) / evolving (cascade MI/S2ST)

## Introduction. Demand.

Machine interpreting (also called real-time speech translation) and speech-to-speech translation are sister LTs. Both are cascade systems of ASR>MT>TTS, but MI is real-time (simultaneous), which requires additional engineering for pipeline management, accuracy, and latency.

MI and S2ST keep opening new markets as the option for no-MI / no-S2ST, with different audiences:

- MI - devices for in-person MI- consumers, tourism, small group training
- MI - software: B2B for internal corporate communications, e-learning, training (often unidirectionally).
- S2ST: marketing and PR agencies and departments, elearning and training organisations for external / pre-recorded purposes.

## Market size and character.

We estimate the current market size to be close to EUR 5 billion. Although the market is dominated by big-tech solutions, as the vast majority of audio and video content is created in a single language in exponentially increasing volumes, the market for these technologies has a very high potential to grow in the next few years, also due to the the compound need for TTS (in voicebots, dubbing, machine interpreting, etc).

# Machine interpreting and S2ST 2

## Main actors.

MI is heralded primarily by interpreting-background companies, whereas S2ST is big-tech driven, as they have the necessary components that can be integrated together in an automated system.

**Language coverage is restricted to high-resource languages**, and (for MI) the input and output speech language coverage may differ (depending on market demand).

## Quality and technology outlook.

MI and S2ST carry the compound quality challenges of the cascade technologies (ASR, MT, TTS). While interim quality can be checked, ultimate quality evaluation is similar to TTS: user feedback.

While they rely heavily on 3rd party core tech for the cascade, **MI actors' added value is in the orchestration of the pipeline.**

**In S2ST (and AV2AV in general) the next frontier is direct hybrid models** that sidestep the cascade (Seamless by Meta, AudioPaLM by Google).

Company	Country of origin	Product	Languages supported	Estimate ASR revenue (2023)	Investment / funding (till March, 2024)
<b>Machine interpreting</b>					
<b>Kudo</b>	US	KUDO AI	6 source, >25 output	EUR 20-50 million	EUR 25 million
<b>Interprefy</b>	CH	Aivia	80, all EU24	EUR 20-50 million	undisclosed
<b>Wordly</b>	US	Wordly	>50	undisclosed	undisclosed
<b>S2ST</b>					
<b>Big tech</b>	Microsoft, Google, Amazon, Nvidia, Baidu, ByteDance, Alibaba				
<b>Contenders</b>	OpenAI, ElevenLabs, Deepdub, Apptek, or Resemble.ai				

# Chatbots and virtual assistants 1

Main LT category	Chatbots and virtual assistants
Market size estimate (2023)	EUR 5-10 billion
Growth potential	Moderate
Investment interest	High
Market character	Well-established Core tech is big-tech dominated, application platforms are various Typically custom-built solutions heavy with professional services, often through implementation partners. Heavily fragmented market for country-specific small and medium business (SMB) clients.
AI / ML adoption / disruption level	High
Technology maturity level	Maturing / evolving

## Introduction. Demand.

Chatbots are a B2B industry, involving buyers from key sectors such as e-commerce, retail, banking, healthcare, IT services, software companies, SaaS cloud platforms, telecommunications, automotive, travel and hospitality, and generally customer service and support, where speed and efficiency of response are critical.

As a compound language technology, chatbots utilise STT, NLU/NLP, MT, and TTS solutions. The traditional predefined, rule-based responses model is being disrupted by LLMs with a more open-ended natural language interface for question answering, information retrieval, and other NLU/NLP tasks, with all their risks.

## Market size and character.

We estimate the market size to be between EUR 5 and 10 billion.

The market potential continues to grow as these systems become more sophisticated by integration with emerging technologies such as AI avatars and RAG, and more reliable via the evolution of robust NLP tools such as semantic search and sentiment analysis.

# Chatbots and virtual assistants 2

## Main actors.

Both concentrated (for big buyers) and fragmented (for SME customers).

Big tech companies developed their own chatbot platforms, while various players compete with them on the market as SaaS platforms, both locally and globally, for enterprise customers.

SMEs are also supported by various local actors, including Poly.ai in the UK, or companies like Tilde.

## Quality and technology outlook.

Chatbots quality and performance as productivity tools are measured by cost savings and by customer satisfaction, while the accuracy may be assessed post-chat with NLP tools, or as ticket reopens / recall metrics. Accuracy demand depends on verticals and cultures (error tolerance).

Technology-wise, low-code and no-code platforms are proliferating, pre-built models as products are holding ground against DIY solutions, and GenAI solutions (including LLMs) are disruptive although with potential latency challenges.

Company	Country of origin	Product	Languages supported	Estimate chatbot revenue (2023)	Investment / funding (till March, 2024)
Amazon	US	Lex	27, non-EU24	n/a (big tech)	n/a (big tech)
Google	US	DialogFlow	120, all EU24	n/a (big tech)	n/a (big tech)
Microsoft	US	LUIS / CLU	95, all EU24	n/a (big tech)	n/a (big tech)
IBM	US	IBM watsonx Assistant	13	n/a (big tech)	n/a (big tech)
Avaamo	IN	LLaMB and IVA	>100, all EU24	EUR 20 million	EUR 27 million
Amelia	US	Amelia	>100	EUR 100 million	~EUR 200 million
Boost.ai	Norway	Boost.ai	n/a	EUR 20-30 million	
Cognigy	Germany	Cognigy.AI	>100	EUR 20 million	EUR 70 million
Kore.ai	US	Kore.ai	>130, all EU24	EUR 200 million	EUR 200 million
OneReach.ai	US	GSX platform	>30 languages for transcription, >100 for translation	EUR 15 million	
OpenStream	US	EVA	n/a	EUR 10 million	
Omilia	Greece	Omilia	n/a	EUR 20 million	EUR 20 million
Yellow.ai	US / IN			EUR 100-200 million	EUR 80 million
Laiye	Canada		n/a	EUR 40 million	EUR 160 million (Series C)
[24]7.ai	US		n/a	EUR 1.2 billion (total)	EUR 20 million
Inbenta	Spain		35	EUR 30 million	EUR 40 million
Rasa	US		"Any language"	EUR 20 million	EUR 30 million
MindMeld	US		"Any tokenizable language"		



LANGUAGE  
TECHNOLOGY  
LANDSCAPE  
CONFERENCE

# Q&A

Thank you.

Have more questions later?  
Find me on LinkedIn or reach out to the project team at  
[Itsurvey@nimdzi.com](mailto:Itsurvey@nimdzi.com).

Feedback survey



SCAN ME