



MINISTERIO  
DE CIENCIA  
E INNOVACIÓN



Financiado por  
la Unión Europea  
NextGenerationEU



Plan de Recuperación,  
Transformación y  
Resiliencia



AGENCIA  
ESTATAL DE  
INVESTIGACIÓN

**HiTZ**

ahōLAB

Hizkuntza Teknologiako Zentroa  
Basque Center for Language Technology

# ASR Models EU

Asier Herranz

19/02/2024

*Este proyecto con referencia 2022/TL22/00215335 está financiado por el Ministerio de Transformación Digital y por el Plan de Recuperación, Transformación y Resiliencia – Financiado por la Unión Europea – NextGenerationEU.*

# Corpus EU

Common  
Voice

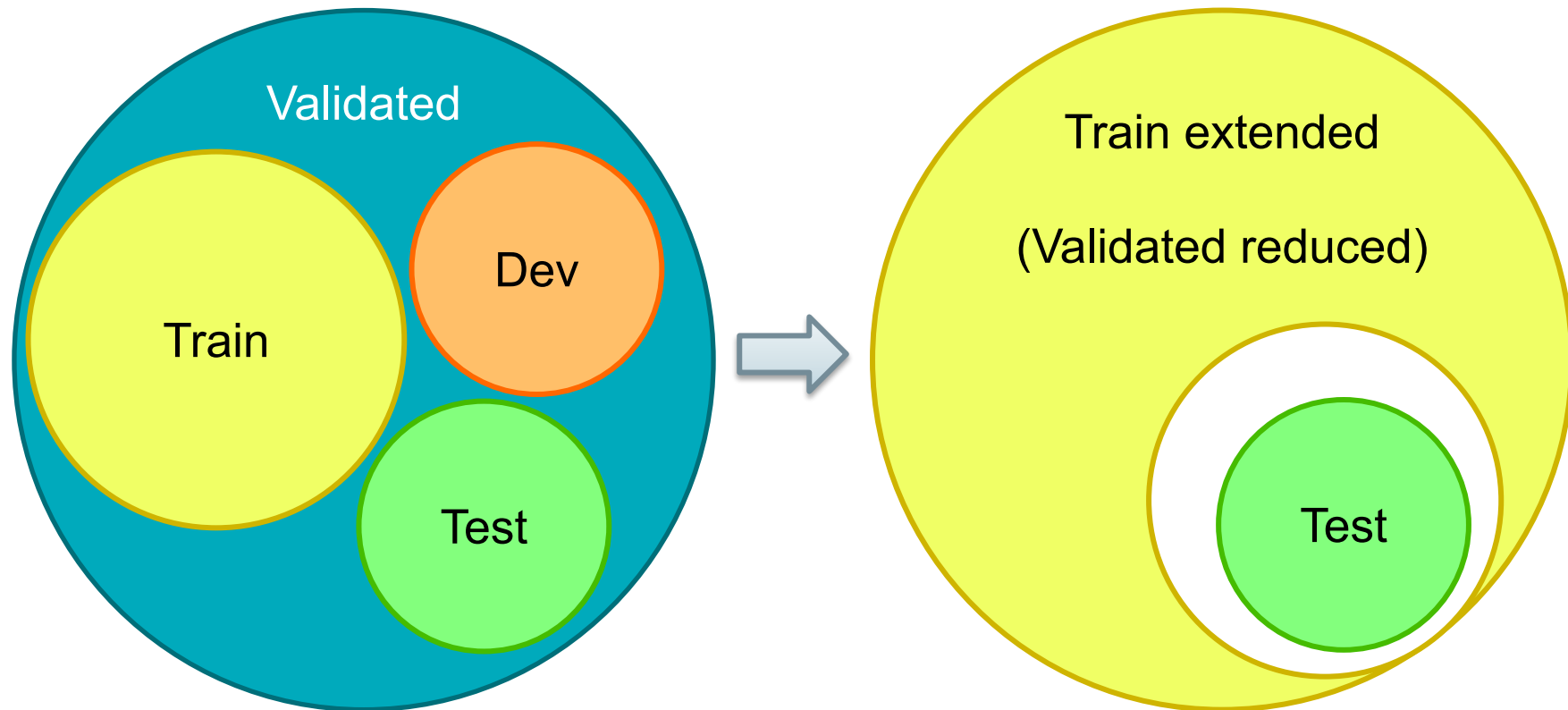
OpenSLR

Kiko  
Aintzane  
AhoMyTTS

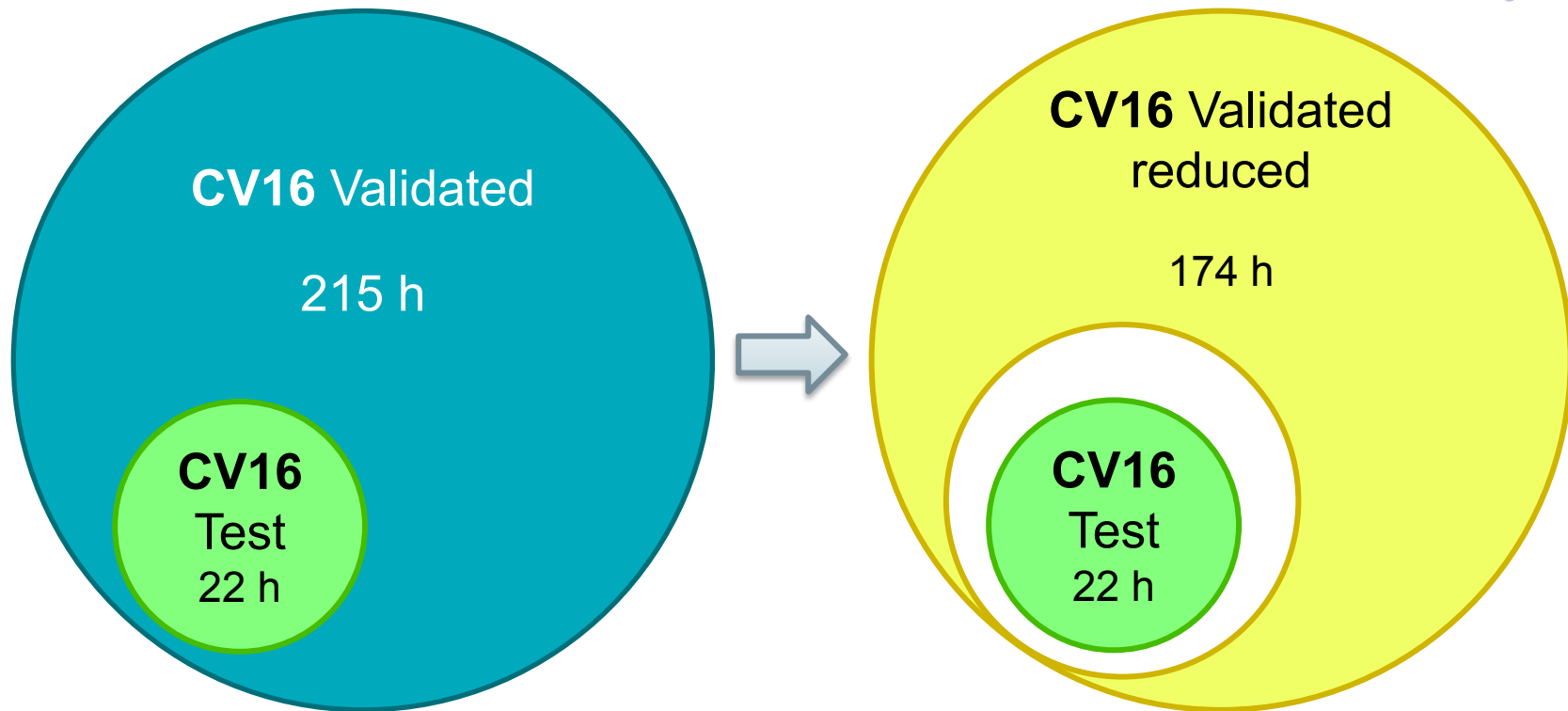
Parliament

- Only Basque (EU)
- No C&P (Capitalization & Punctuation)
- No acronyms
- No numerical characters
- $\sim 1 \text{ s} < \text{Audios} < \sim 20 \text{ s}$

# Corpus EU

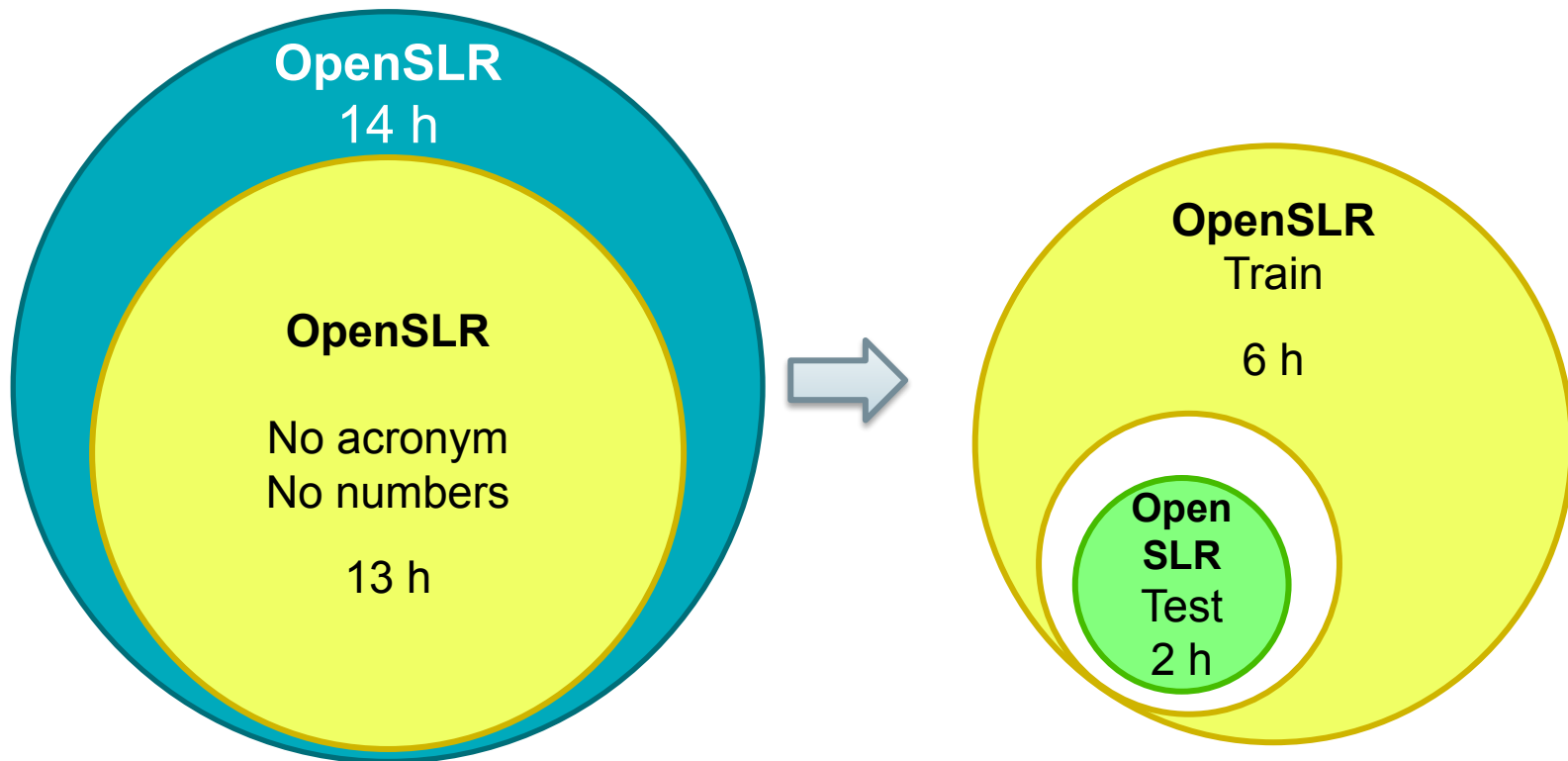


# Corpus EU - Common Voice 16



# Corpus EU - OpenSLR

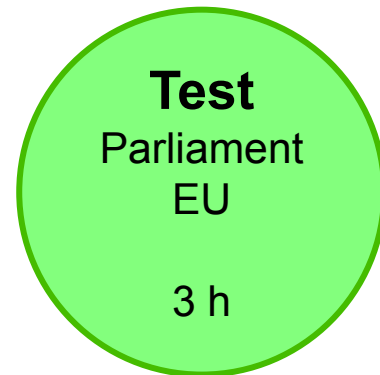
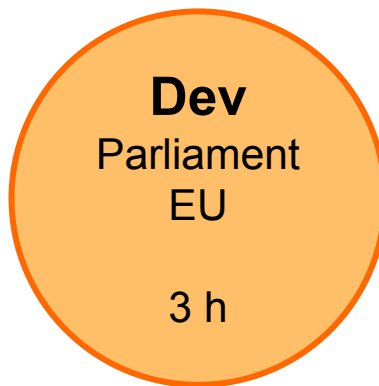
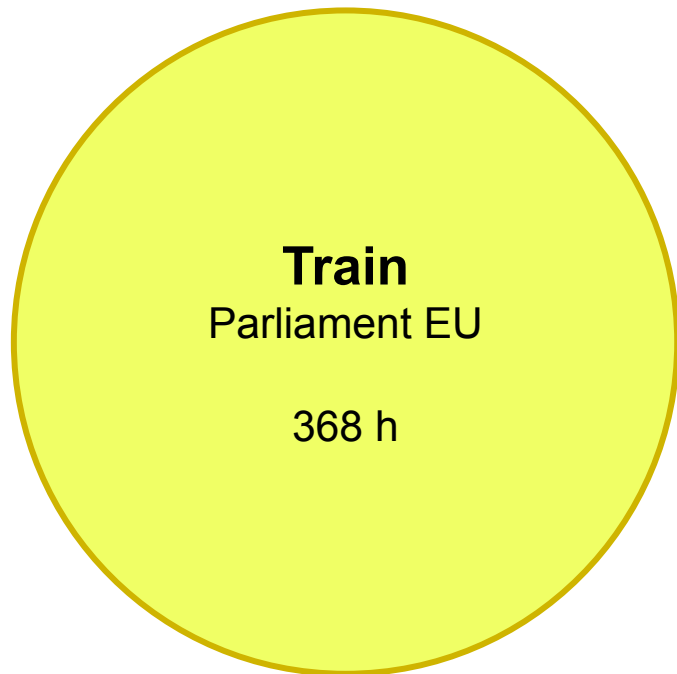
*Open*SLR



# Corpus EU - Parliament - GTTS

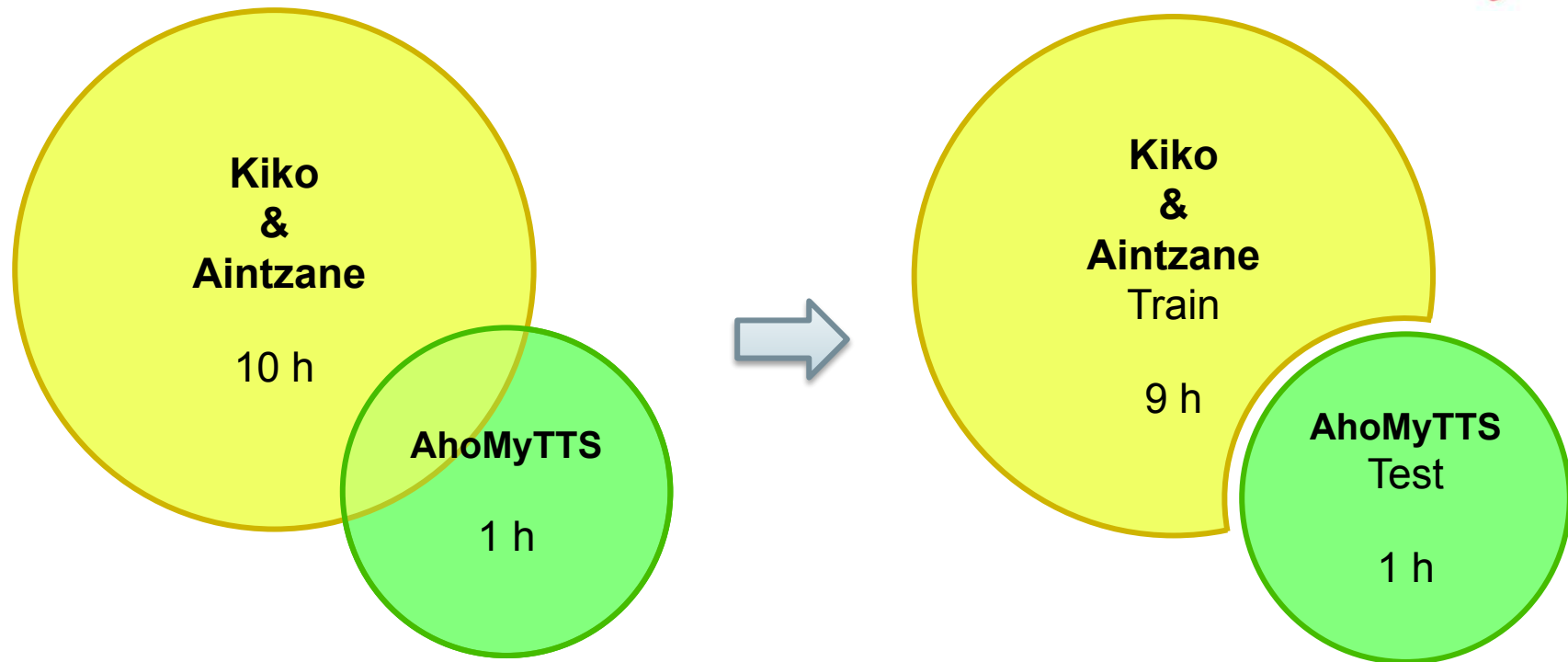


**Hugging Face**

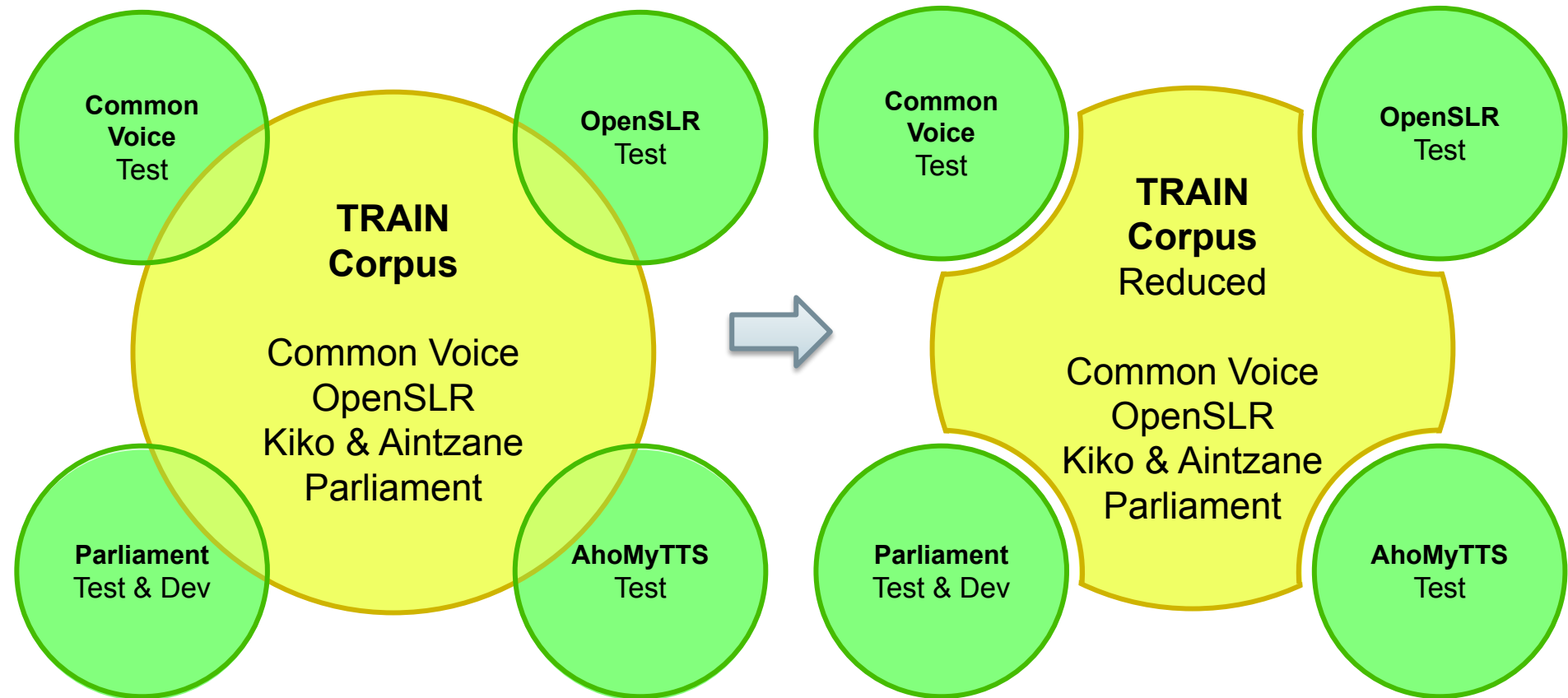


# Corpus EU - Kiko & Aintzane, AhoMyTTS

ahō LAB  
ahō myTTS

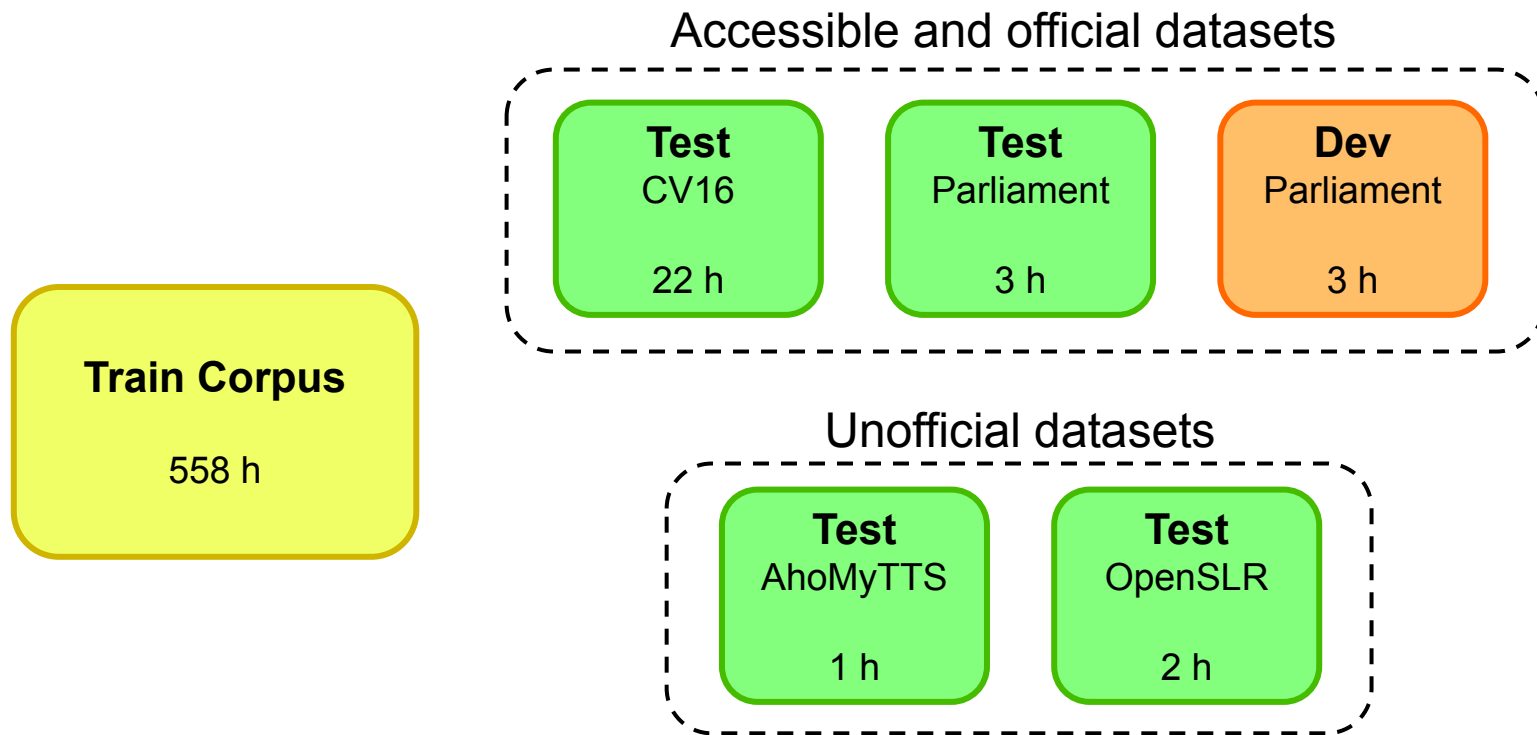


# Corpus EU - Train & Test

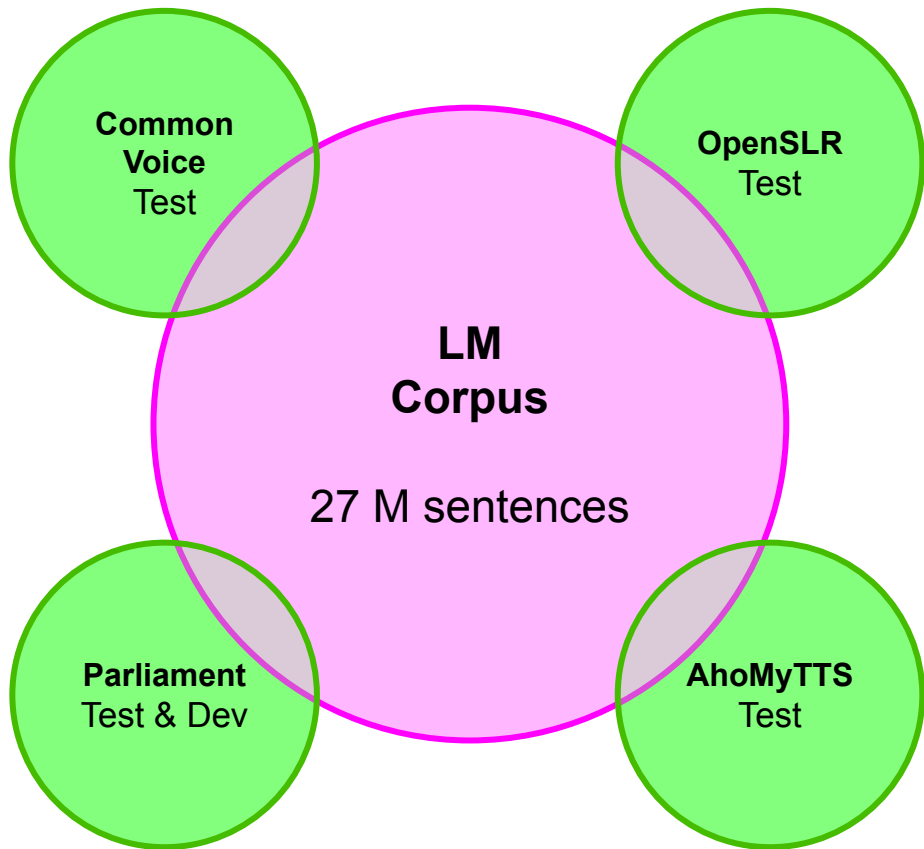




# Corpus EU - Train & Test



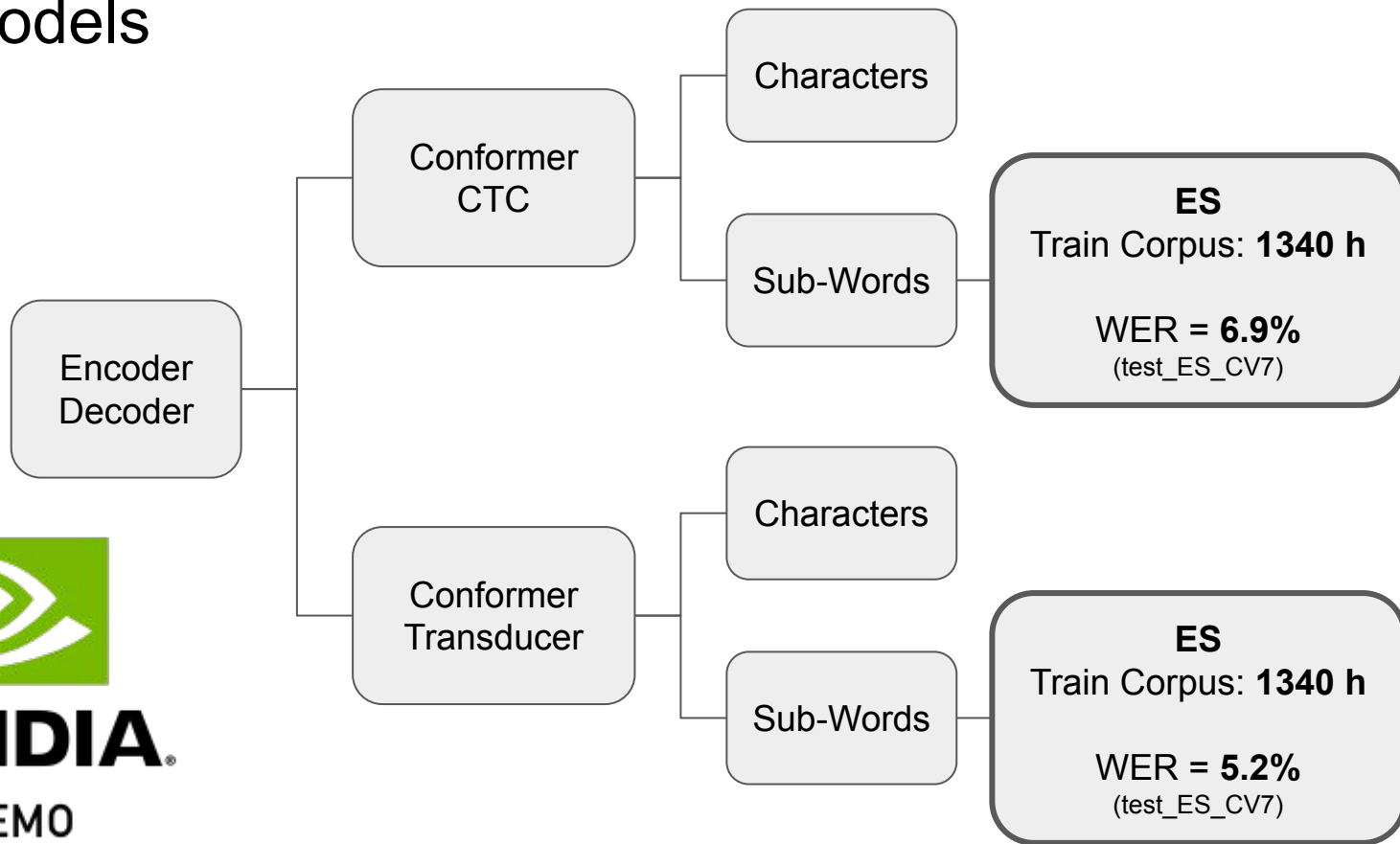
# Corpus EU - Language Model



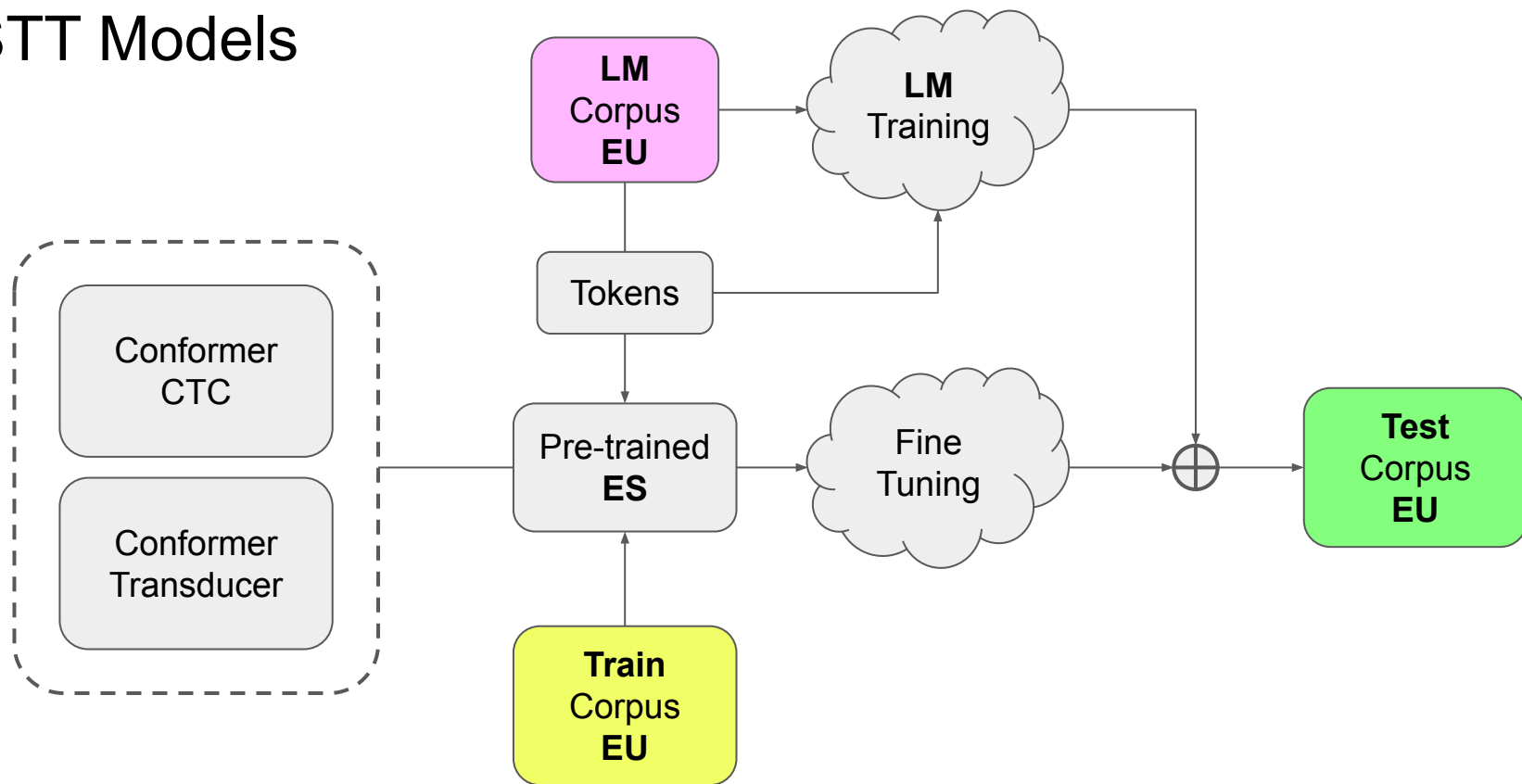
## Leakage:

- **CV16 Test:**
  - 43.88% (5591 / 12742)
- **AhoMyTTS Test:**
  - 13.39% (79 / 590)
- **OpenSLR Test:**
  - 0.29% (3 / 1034)
- **Parliament Test:**
  - 0% (0 / 1521)
- **Parliament Dev:**
  - 0% (0 / 1397)

# STT Models



# STT Models



# STT Models - Results

Train Corpus 114 h CV15 + Others (No Parliament)		Common Voice		AhoMyTTS		OpenSLR		Parliament	
Model Name Conformer-CTC	Tokens	Greedy	With LM	Greedy	With LM	Greedy	With LM	Greedy	With LM
stt_eu_conformer_ctc_v0.10	29	7,87%	5,91%	11,04%	8,20%	16,00%	11,89%	10,80%	9,33%
stt_eu_conformer_ctc_v0.6	256	9,21%	2,98%	14,23%	7,44%	18,21%	9,82%	10,76%	9,29%
stt_eu_conformer_ctc_v0.9	512	9,43%	2,99%	13,52%	7,38%	18,92%	10,12%	10,33%	8,72%
stt_eu_conformer_ctc_v0.8	1024	9,50%	2,43%	13,83%	7,83%	18,70%	10,63%	10,82%	9,34%

## Train Corpus 230 h CV16 + Others (No Parliament)

stt_eu_conformer_ctc_v0.1	256	4,85%	2,37%	10,98%	6,06%	16,27%	9,17%	10,20%	8,55%
---------------------------	-----	-------	-------	--------	-------	--------	-------	--------	-------

## Train Corpus 558 h CV16 + Others + Parliament

stt_eu_conformer_ctc_v0.4	256	4,53%	2,22%	9,60%	5,67%	16,56%	9,31%	4,49%	4,22%
---------------------------	-----	-------	-------	-------	-------	--------	-------	-------	-------

# STT Models - Results

Train Corpus 114 h CV15 + Others (No Parliament)		Common Voice		AhoMyTTS		OpenSLR		Parliament	
Model Name Conformer-Transducer	Tokens	Greedy	With LM	Greedy	With LM	Greedy	With LM	Greedy	With LM
stt_eu_conformer_transducer_v0.16	29	6,05%		9,03%		13,96%		8,61%	
stt_eu_conformer_transducer_v0.14	256	5,94%		10,33%		13,52%		8,40%	
stt_eu_conformer_transducer_v0.15	512	5,76%	3,25%	9,53%	7,73%	13,28%	10,35%	8,31%	8,97%

Train Corpus 230 h CV16 + Others (No Parliament)									
stt_eu_conformer_transducer_v0.2	256	3,04%	2,20%	7,88%	5,92%	12,59%	9,95%	7,82%	9,04%

Train Corpus 558 h CV16 + Others + Parliament									
stt_eu_conformer_ctc_v0.4	256	2,88%		6,81%		11,51%		3,94%	

# STT Models - Results

Model	AhoMyTTS	OpenSLR (100 sentences)
Conformer - CTC	9.60%   5.67%	19.02%   9.95%
Conformer - Transducer	6.81%   -	12.72%   -
Itzuli ASR	11.35%	9.45%

# STT Models - Comparisons

## Conformer - CTC

- Worst Greedy WER
- Faster Training
- Faster inference using LM
- WER using LM similar to Transducer

## Conformer - Transducer

- Better Greedy WER
- Slower Training
- Slower Inference using LM
- WER using LM similar to CTC



# STT Models - **Future**

- Tests with Capitalization & Punctuation
- Bilingual model ES/EU?
- New Architectures from Scratch
- Fine-tuning Whisper + LM