Extractive & Abstractive Auto Summarization (NL)

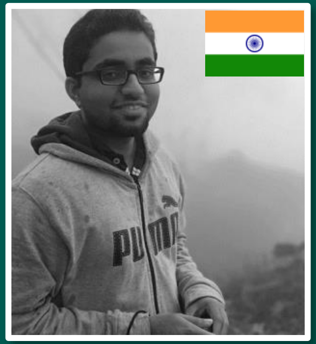
vrt REACH

Amberscript

Meet the team

Peter-Paul de Leeuw
Business Lead

Timo Behrens
Technical Lead



Nithin Holla
ML Engineer

Jolien de Louw
Project Manager

Peter-Paul holds a MBA from INSEAD and cum laude master's in Economics from the Erasmus University of Rotterdam. He has several years' experience working in strategy consultancy. Within Amberscript Peter-Paul is responsible for Strategy and product development.

Timo holds a degree in Electrical Engineering from TH Köln. As technical lead Timo is responsible for overall development, making sure Amberscript deploys state of the art technology with optimal reliability and quality.

Nithin holds a master's degree (cum laude) in Artificial Intelligence from the University of Amsterdam. He is responsible for research and development of Amberscript's natural language processing and speech recognition components.

Jolien holds a master in Economics from Utrecht University and has 8 years of experience as a management consultant at Accenture. She is responsible for operations management and customer support at Amberscript.

Expected project involvement:

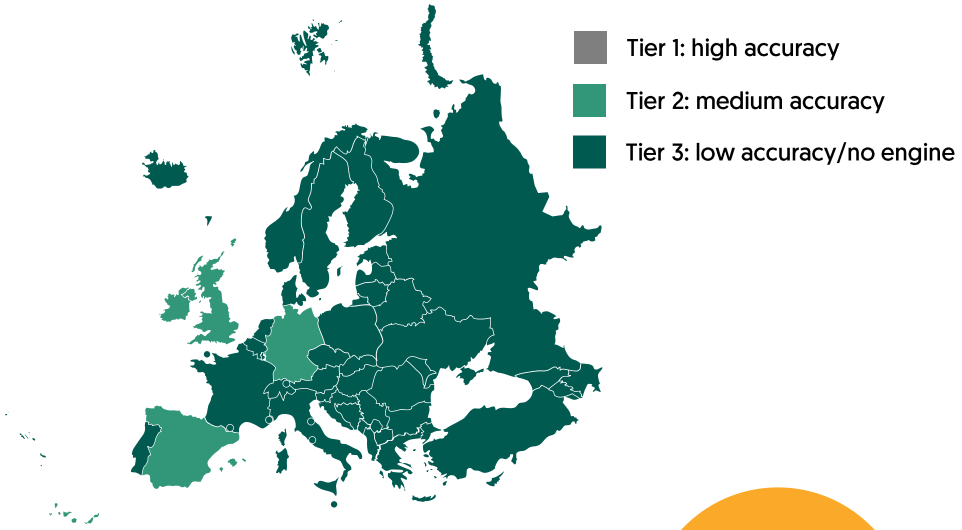
Peter-Paul de Leeuw	Timo Behrens	Nithin Holla	Jolien de Louw
~8 hours per week	~12 hours per week	~24 hours per week*	~12 hours per week**

** Next to Nithin (ML Engineer), we expect to require 12 hours per week from our NLP Engineer and ~12 hours per week from a Software Engineer for the Editor integration.*
*** During the evalution steps we also require ~12 hours per week form our freelance network of native language experts.*

Our milestones in the past years

Our growth journey

Europe is behind compared on speech-to-text engines accuracy and accompanying services compared to the USA & China.



Amberscript identified the opportunity and was founded in Amsterdam, 2017, focusing mostly on European languages.

Founded 2017

2018

2018: We launched our **manual transcription and subtitling services** in all European languages.

2019

2019: We launched **Europe's first speech recognition engine**, made specifically for political debates.

2020

2020: We captioned Germany's first political debate **in real-time and launched our subtitle editor**.

NOW

AMBERSCRIPT NOW:

- ✓ We serve over 1.000.000 customers worldwide
- ✓ Available in 38 language
- ✓ Offer (live) transcriptions, captions & subtitles
- ✓ User intuitive transcript and subtitle editor
- ✓ API and custom engines
- ✓ Data annotation

With whom we work with

A selection of our clients

SKYHIGHTV

VUB VRIJE UNIVERSITEIT BRUSSEL

ABN·AMRO

TU Delft

Erasmus MC
Universitair Medisch Centrum Rotterdam

COMPANY WEBCAST
A EURONEXT COMPANY

G-STAR RAW

Rijksoverheid

European Commission

HUMBOLDT-UNIVERSITÄT ZU BERLIN

UNIVERSITEIT VAN AMSTERDAM

amazon

BNN

anwb

npo

SAP

ORF

Ministerie van Justitie
en Veiligheid

fd.
hetfinancieeledagblad

UNIKASSEL VERSITÄT

WB

ZDF

WAGENINGEN
UNIVERSITY & RESEARCH

Our offerings

Transcriptions

Convert your audio and video to text using our high-end AI engines.

Automatic: Improve the text in our Editor yourself.

Manual: Let a native language expert perfect your text.

Subtitles

Add captions to your videos automatically using our subtitle generator.

Automatic: Improve the text in our Editor yourself.

Manual: Let a native language expert perfect or translate your captions.

API & Custom Engines

Add our Speech-to-text API to your stack with the option to request a tailored model for the world's highest accuracy.

Data Annotation

We can create your training material with annotators trained according to your guidelines, running annotations through strict quality control checks.

What's next?

- Create an **analytical layer** on top of speech-recognition and offer customers auto summarization.
- We are currently running a **proof of concept for the Ministry of Justice** to summarize court hearings.

Proof of concept

Ministerie van Justitie en Veiligheid

Experiment phase approach in a nutshell

Month 1	Month 2	Month 3	Month 4

1 PREPARATION

2 DATA ANALYSIS

3 EXPERIMENT

4 EVALUATE

4 BUILD

1 & 2. Prep & Data Analysis Activities

- ❑ Onboarding of the project team
- ❑ Create detailed Experiment phase project plan
- ❑ Agree on Experiment phase project goal & deliverables
- ❑ Determine & validate product MVP features
- ❑ Analyse VRT news data set
- ❑ Analyse English summarization dataset and the potential to translate these to Dutch

3. Experiment Activities

- ❑ Pre-processing of VRT news data set
- ❑ Create training & test sets
- ❑ Train extractive summarization model in an unsupervised fashion (using TextRank algorithm)
- ❑ Train abstractive summarization model (using mT5 model)
- ❑ Evaluate according the ROUGE metric
- ❑ Experiment with two-stage summarization

4. Evaluate Activities

- ❑ Gather and log incidents raised
- ❑ Organize feedback sessions to improve summarization model(s)
- ❑ Human evaluation by our network of professional language experts
- ❑ Redefine models based on the findings (in (3) experiment step)

5. Build Activities

- ❑ Meet all pre-defined MVP features
- ❑ Build Editor integration to manually modify generated summaries for users
- ❑ Create detailed project plan for Evolve phase

QUESTIONS

Amberscript B.V.
Keizersgracht 668B
1017 ET Amsterdam
the Netherlands

Commercial Register (KvK): 70427585
VAT (BTW) nr: NL858314071B01

Telephone: +31 20 24 40 992
Availability: Mon – Fri 09:00 – 18:00
Contact by email: info@amberscript.com