Research and Innovation Action (RIA) H2020 – 957017

Stream Learning for Multilingual Knowledge Transfer

https://selma-project.eu/

D6.6 Final Impact Report

Work Package	6
Responsible Partner	DW
Author(s)	Ksenia Skriptchenko
Contributors	Kay Macquarrie, all Partners
Reviewer	Tugtekin Turan
Version	1.0
Contractual Date	31 March 2024
Delivery Date	28 March 2024
Dissemination Level	Public

Version History

Version	Date	Description
0.1	13/02/2024	Added List of Events & Publications
0.2	20/02/2024	Added Exploitation Components
0.3	13/03/2024	Internal review
1.0	25/03/2024	Finalization and Submission

1. Executive Summary

The SELMA mission statement shows a central impact objective:

Shaping AI speech and text technologies for media & the newsroom

A major focus is to convince potential user partners about SELMA output and exploit components and platforms.

SELMA is a "Research and Innovation Action" project. Therefore, the project had two impact focuses:

⇒ advancing the state-of-the-art in various NLP related tasks and technologies through improvements in research, and

⇒ bringing technology to the market through tools, components and an open-source platform.

SELMA has made considerable progress in several NLP technologies, raising the state of the art in areas such as abstractive summarization, speaker recognition, and subject labelling through translingual transfer.

One of the project's main goals was to create the foundation for the searching and analysis of large-scale multilingual data streams and content collections, which include both text and video content. SELMA's technology is currently scalable to handle up to 10 million items per day and can process 300,000 media pieces. The media production workflow (UC 2) has evolved into a product, which is now integrated and used within DW, with a user base of around 1000 people, has already been adopted by other clients, and is looking ahead to more potential users. SELMA created an open-source media creation platform that includes voice-over, transcription, and translation features. This platform emphasizes its open-source character by being available locally through a download or through a web link.

Table of Contents

Table of Figures

Table of Tables

2. Introduction

This deliverable, "D6.6 Final Impact Report," outlines SELMA's activities in connecting with the relevant players in the commercial, academic, and general public sectors as well as in achieving the intended impact on its target audience.

This report is divided into three main parts: Dissemination, Communication, and Exploitation.

Section 1 (Dissemination) outlines plans and shows efforts to inform, inspire and involve future SELMA users and the research community. SELMA's dissemination goal was to engage potential users, HLT (Human Language Technology) providers, and early adopters to establish a feedback culture and build a network with other projects, researchers, and technology users.

The communication strategy is laid out in **Section 2 (Communication)**. It identifies the primary channels and activities for communication. Additionally, it highlights the initiatives made to increase public knowledge of the project and its advancement.

A summary of exploitation activities is provided in **Section 3 (Exploitation)**. It includes an overview of the output of the project, including datasets, components / models and platforms, as well as descriptions on the IPR status of each of the 26 components and its licenses (of which the half of them have been released as open source).

3. Dissemination

The purpose of this section is to provide a project dissemination strategy by highlighting targeted groups and communities, defining internal dissemination/communication guidelines and procedures, outlining the foreseen channels, and reporting on the efforts in the first project year.

This section describes the efforts of defining, identifying, and reaching our target audience, focusing on its two primary target groups, i.e., the scientific communities and the media world. It provides the plan of selecting, setting up and supplying the right dissemination channels and a survey of the dissemination events to promote the results in the related fields of research.

Following the dissemination activities framework outlined in the DoA, SELMA adopts a multi-channel and multi-target approach. SELMA pursues a clearly defined strategy, which will be outlined and specified further below in the sub-chapters.

3.1 Target Groups

Overall, we divide our target audience for dissemination activities as follows, and will address primarily:

Target Groups	Reach out primarily via	Examples
Broadcasting and media world	Conferences	EBU 2024, EBDVF 2023
Industries using language processing technologies	Conferences, User Days	User Day II Avignon
Translation agencies	LinkedIn	Number of posts
Media monitoring organizations	Conferences, User Days	User Day I Bonn
Industries in need of monitoring multilingual content across the world media	Conferences, User Days	EBU 2024, EBDVF 2023

Scientific and research community	Conferences, Publications, GitHub	Number of conferences, Number of publications
Stakeholders and their networks	Conferences, User Group Meetings	User Group Meeting 2024
Policy makers and interest groups	Conferences and workshops	EBDVF
Human Language Technology users	GitHub, User Days, Linkedin	User Day II Avignon
General public (interested users)	Website, GitHub, YouTube, X	Number of blog posts

Table 1 *Target groups and reach out*

Network

All the SELMA consortium's partners have established networks and exert every effort to inform and engage with their intended markets, both on an individual basis and collectively.

The SELMA consortium also intends to support the Big Data Value Association in any actions related to the project's activities.

User Group

The User and Advisory Board serves as an advisor to the Innovation Manager and the project's Steering Board.

On March 24th, 2022, SELMA had its first User Group meeting. It was organized as an interactive online meeting and brought together NLP and media experts from many organizations including BBC, University of Edinburgh, COFINA, Visapress and AICEP. The status of SELMA developments and Use Cases were showcased and feedback from the user group was discussed and collected. As an outcome, the Advisory Board was formed, which is a smaller subset of the User Group and consists of representatives of these organizations: BBC, EBU and University of Edinburgh.

On March 21st, 2024, the second and final SELMA User Group meeting took place as an online event. A focus was on the SELMA outcome, including academic findings and concrete outcomes, such as the Monitio Media Monitoring tool, the translation and subtitling tool plain X, and the Podcast Creator. Also, the SELMA Open-Source NLP Platform was showcased. Feedback provided by attending media experts was collected.

User group activities include identifying the strong points and potential issues with respect to the objectives of the project (with emphasis on the innovation objectives) and providing recommendations. Furthermore, the members of the User Group helped us maximize our industry outreach, serving as links between the consortium and external key industry players. Around 15 representatives from research organizations, European media companies and technology providers have formed the SELMA user group.

3.2 Strategy

The plan identifies the main initiatives and channels that will support the project goals. Events, the creation of materials to describe and demonstrate project accomplishments, and targeted publications are all governed by the dissemination strategy. Three significant stages have been established by the project consortium.

<table>
<tr><td></td><td>Y1</td><td>Y2</td><td>Y3</td><td>Y4</td></tr>
<tr><td>Phase 1</td><td colspan="3">Inform & inspire</td><td></td></tr>
<tr><td>Phase 2</td><td></td><td colspan="2">Involve & contribute</td><td></td></tr>
<tr><td>Phase 3</td><td colspan="2"></td><td>Share & Convince</td><td></td></tr>
</table>

Table 2 *Three dissemination stages and progress at the end of the SELMA project*

Inform and inspire

The ***"inform and inspire"*** phase started in the first year and was active during the whole lifetime of the project. The focus in the first phase was on:

- Outlining the project's vision, aims and goals
- Setting up the dissemination channels and activities to spread the word and to inform target audiences about SELMA and its major objectives
- Introducing and interacting with relevant communities, which might differ in the scope of different use cases

Involve and contribute

The ***"involve and contribute"*** part started in the second year and continued until the end of the project. This phase was consisting of the following steps:

- Identifying key influencers to involve into the feedback loop, in testing early prototypes, sharing research results
- Utilizing early adopters as multipliers and to spread further awareness
- Obtaining user feedback on development and creating solutions for obstructions

Share and convince

The last part of the dissemination strategy ***"share and convince"*** started in the third year and mainly focused on:

- Demonstrating progress by making available a variety of open-source resources and project outcomes
- Showing prototypes and innovative features to targeted audiences and third parties
- Engaging with target groups and individual users to support the SELMA exploitation activities

3.3 Dissemination Approach

The general SELMA dissemination strategy and approach were discussed in the chapter above. The next part provides more information on how we intended to accomplish the predetermined objectives as well as what has been planned and accomplished throughout the project's timeframe.

3.4 Channels (Website, Social Media)

Website and Blog

An important channel of the SELMA communication and dissemination strategy is its website.

The SELMA homepage can be reached at *www.selma-project.eu.* It provides information on the project, main goals and project partner descriptions, and contact person information.

The blog feature of the homepage serves as the SELMA repository of the pertinent contributions to relevant artificial intelligence and human language technologies made by consortium partners. The public at large is the primary audience.

Figure 1 *"My year with plain X" blog article by Konstantin Klein, DW senior editor and plain X user*

The following content was published throughout the project:

#	Title	Partner	Link	Views
1	**Biases in AI**	DW	Link	206
2	**Spoken Languages – Learn like a child** Curriculum Learning Methods	LIA	Link	321
3	**Diving into Infoboxes & Knowledge Graphs**	IMCS	Link	418
4	**What we do with people, places, and organizations** Named Entity Recognition	Prib	Link	274
5	**A Yummy Piece of Cake** Machine Learning Algorithms	FhG	Link	377
6	**Why (Counting) Diversity Matters** Diversity in Media	DW	Link	561
7	**How to satisfy data-hungry machine learning** Self Supervised Learning	LIA	Link	399
8	**Need Computing Resources? Take a Queue Token!**	IMCS	Link	218
9	**Who Spoke When?** Speaker Diarization	FhG	Link	124
10	**On the Path to the Responsible AI** Trustworthy AI	Prib	Link	136
11	**Why does AI need labeled data?**	DW	Link	262

12	**Introducing LeBenchmark:** A Comprehensive Framework for Evaluating Self-Supervised Learning	LIA	Link	55
13	**The SELMA Open-Source Platform**	IMCS	Link	88
14	**My year with plain X** User Experience and Acceptance	DW	Link	43

Table 3 *Overview Blogposts on SELMA*

It was constantly encouraged for all SELMA partners to actively contribute content to the website, especially for the blog section.

In the final year we also posted:

- documentations on project activities and outcome:

  https://selma-project.eu/project-output/

- links to prototypes, demos and tutorials, code, datasets and resources:

  https://selma-project.eu/use-cases/

Both sections of the SELMA webpage generated a great deal of interest from our users and received 502 and 467 Views, respectively.

plain X Website

During the second half of 2023, a product website for plain X was developed and published jointly by Priberam and Deutsche Welle. Its aim is to address potential customers and to offer a direct entry for interested people and companies. Visitors can easily register and get to know the system for a limited period of time ("Try for free"). The plain X website can be reached under: https://www.plainx.com/.

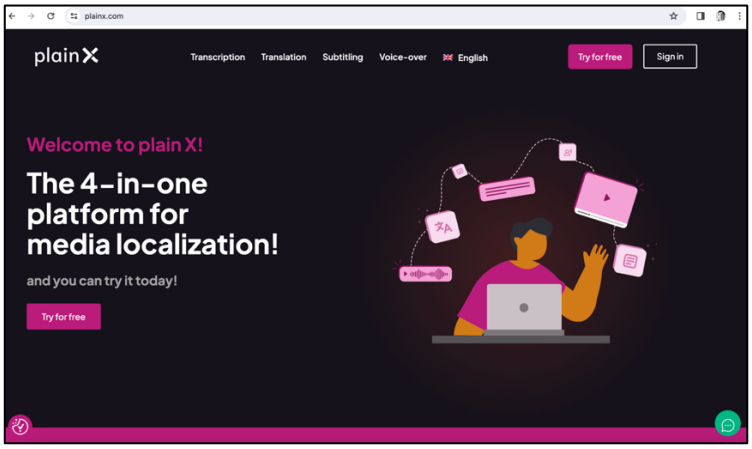

Figure 2 *plain X product Website*

Social Media

This section lists social media-based dissemination approaches and describes corresponding strategies.

YouTube

This platform was chosen to present our findings in the form of videos. Furthermore, we used our YouTube presence to reach out to new contacts and target new collaborations. The first YouTube contribution was an image video that gave our target audience a brief but in-depth overview of the project.

Figure 3 *SELMA's image video as seen on SELMA's YouTube channel*

This image video was made available in English. Part of the video was artificially dubbed using SELMA's own plain X application. Subtitles in English, German, and Russian were also created using plain X. The following table summarizes SELMA's video output in the second and the third project year and gives a brief overview of video topics.

Videos	Description	URL
SELMA Image Video	SELMA helps media monitors and journalists make sense of huge content streams, also making audiovisual output more accessible through transcription, translation, voice over and subtitling.	URL
New Podcast Creation (Demonstration)	One concrete application of the SELMA project is the Podcast Creator. Its purpose is to create a podcast almost on-the-fly using search, summarization and speech synthesis techniques with customized SELMA voices.	URL
"The Footprint of AI & NLP Technology in the Media - What Comes Next?" (User Day Panel Discussion)	Stephanie Bradford (DW Diversity), Kirsten Radtke (DW Informations), Guntis Bārzdiņš (University of Latvia), Afonso Mendes (Priberam), moderated by Olga Kisselmann (DW Research & Cooperation Projects), discussing the topic during the User Day.	URL
"Data Representation Matters - Counting Diversity" (User Day Panel Discussion)	Mirjam Gehrke, senior editor from Deutsche Welle highlights the needs of a broadcaster to meet strategic goals in this field. Kay Macquarrie, coordinator of the SELMA project, shows how SELMA technology can help to count diversity numbers based on NLP and data analysis.	URL
"plain X - The Four in One NLP Tool" (User Day Demonstration)	plainX in a nutshell: plainX is an integrated platform combining a task-based workflow with access to powerful HLT (human language technologies).	URL
"Monitio - A Multilingual Media Monitoring Tool (User Day Demonstration)	Monitio searches and analyses content (currently up to 200,000 articles a day from around the world) to deliver information and indications for better decision-making.	URL

"Artificial Intelligence (HLT) & The Newsroom" (User Day Panel Discussion)	Bird's eye perspective from Peggy van der Kreeft on how AI at Deutsche Welle evolved (featuring various EU projects), basically covering the past decade. Includes an outlook into the future.	URL
SELMA Open-Source Platform	In a nutshell walkthrough of the OSS	URL
Three SELMA language technology apps: DW Speaker DW Summarizer Podcast Creator	An overview of three macOS-based experimental apps that we prototyped as part of the SELMA EU project. The DW Speaker app synthesizes speech from text using a variety of speech engines. This includes an engine developed within SELMA, giving access to voices in Urdu and Brazilian Portuguese. The DW Summarizer app summarizes text using a various engines, amongst them the Priberam engine developed within SELMA. The Podcast Creator app allows a user to create audio news bulletins with synthetic speech. It groups the functionalities of the other two apps, adding an introduction and a good-bye section, with some some music mixed in. As an additional feature, the app imports suggested storylines from the Monitoring platform Monitio.	URL

Table 4 *Related videos, produced and posted throughout the project's lifecycle*

SELMA's last produced video on language technology apps was posted on the DW Innovation YouTube account. This was done to ensure the longevity of the video's dissemination and, thus, the sustainability of SELMA project results. We also decided not to post on SELMA's own YouTube channel to avoid duplicate content, as the YouTube algorithm punishes this practice.

GitHub

We have made several of the SELMA software and applications available on GitHub. It is therefore possible to use and try out various SELMA components under: https://selma-project.github.io/

3.5 Events

A highly important part of the dissemination strategy is being present at relevant events for the SELMA project work. It helped us to stay informed and up to date in the scientific areas, present project achievements and results, meet relevant stakeholders for future collaboration and cooperation, and prepare for exploitation activities.

A list of attended or organized conferences (academic & industry), workshops, and other dissemination events in the runtime of the SELMA project is provided in the table below. User Days / Events are listed separately.

#	Date	Title	Partner
	2021		
1	10.01.2021	EBU presentation: DW Benchmarking Efforts	DW
2	05.02.2021	79th International Conference of the University of Latvia	IMCS UL
3	18-19.03.2021	BDVA PPP “project meet-up”	DW
4	23.03.2021	Seminar "A very brief history of Spoken Language Understanding"	Priberam, LIA
5	20.04.2021	3rd Shared Task on SlavNER (part of the 8th Workshop on Balto-Slavic Natural Language Processing)	Priberam
6	26.05.2021	Open “NLP” Workshop of GoURMET H2020	Fraunhofer
7	21.06.2021	EBU Workshop: MT for low-resourced languages applied at DW	DW
8	28.06.2021	17th Baltic Conference on Intellectual Cooperation (BCIC)	IMCS UL
9	5-6.08.2021	SIGSLT - Special Interest Group on Spoken Language Translation (SIGSLT)	LIA
10	30.08-03.09.2021	Interspeech 2021	LIA

11	27.09.2021	EBU workshop on AI applications, with BBC, RAI, DW and other EBU members	DW
12	27-30.09.2021	SPECOM 2021	LIA
13	06.10.2021	Interview AI in language technology	DW
14	25.10.2021	EBU Workshop: AI at DW	DW
15	15.11.2021	Presentation at Annual Meeting Goethe - plain X Application	DW
16	17-19.11.2021	Festival IA 2021 – Avignon Université	LIA
17	18.11.2021	GoURMET User Event	DW
18	24.11.2021	Presentation at ARD: plain X Application Outlook	DW
	25.11.2021	EBU AIDA Workshop – AI at DW	DW
19	6-14.12.2021	NeurIPS conference 2021 (Thirty-fifth Conference on Neural Information Processing Systems)	LIA
	22.12.2021	Presentation at EBU – DW Benchmarking Efforts	DW
	2022		
1	15.02.2022	Hipeac Webinar	DW
2	21.-23.03.2022	Propor 2022	Priberam
3	22.03.2022	Online Workshop of the Interinstitutional Task Force on Speech Recognition (European Union)	LIA
4	28.04.2022	European Language DataSpace Meeting: DW Data & European Language DataSpace Challenges	DW
5	22.-27.05.2022	IEEE ICASSP 2022	LIA
6	22.-27.05.2022	ACL 2022 + IWSLT 2022	LIA
7	02.06.2022	EAMT2022 Poster Presentation	DW, Priberam

8	10.06.2022	Fifth Workshop on Narrative Extraction from Texts, held in conjunction with the 44th European Conference on Information Retrieval (ECIR 2022)	Priberam
9	13-17.06.2022	JEP 2022 Journées d'étude sur la parole	LIA
10	15.06.2022	27th International Conference on Natural Language & Information Systems (NLDB 2022)	IMCS
11	17.06.2022	EBU Workshop (EBU Access Services Expert Event)	DW
12	20-25.06.2022	LREC 2022 - International Conference on Language Resources and Evaluation	Fraunhofer, LIA
13	27.06-01.07.2022	TALN 2022 Conférence sur le Traitement Automatique des Langues Naturelles	LIA
14	27.06-05.08.2022	JSALT Workshop 2022	LIA
15	18-22.09.2022	ISCA Interspeech 2022	LIA
16	14.10.2022	Fête de la Science	LIA
17	14.10.2022	Presentation of SELMA DockerSpaces technology to AI-Lighthouse (HORIZON-CL4-2022-HUMAN-02-02) consortium meeting at Oulu University	IMCS
18	25.11.2022	Diversity Use Case presentation for ARD AI working group	DW
19	07.-12.12.2022	EMNLP 2022 (The 2022 Conference on Empirical Methods in Natural Language Processing)	Priberam
20	14.12.2022	Second BDVA workshop with ICT-51 projects	DW
21	15.12.2022	LOCDOC Master Class on plain X	DW, Priberam
	2023		
1	17.02.2023	81st International Conference of the University of Latvia, Presentation	IMCS
2	26.04.2023	Humanities in the digital age: Securing innovation and empowering democracy, Presentation	IMCS

3	04-10.06.2023	2023 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2023)	LIA
4	10.06.2023	SASB 2023, presentation & LIA Commitee Member	LIA
5	14.06.2023	BDVA Data Week, Presentation	DW
6	16.06.2023	EBU Access Service Meeting, Keynote	DW
7	20.06.2023	Global Media Forum, Presentation	DW
8	27.06.2023	MetaForum, Presentation	DW
9	29.06.2023	JSALT Presentations	DW, LIA
10	20-24.08.2023	Interspeech 2023, Presentation	LIA
11	12.09.2023	CLARIN-LV Conference, Presentation	IMCS
12	12.09.2023	Meeting the Nerds 2023	DW
13	25.10.2023 - 27.10.2023	European Big Data Value Forum (EBDVF) 2023, Booth	DW
14	02.11.2023	Seminar for Language School of National Armed Forces, Seminar	IMCS
15	14-15.11.2023	FestivalIA Avignon, 2nd User Day	LIA, all
16	27.09.2023	Conference at University of Latvia "Information, half-truths, lies - what is what?", Invited Talk	IMCS
17	16-20.12.2023	2023 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)	LIA
	2024		
1	30.01.2024-01.02.2024	EBU Production Technology Seminar	DW
2	09.02.2024	82nd International Conference of the University of Latvia	IMCS
3	13-15.3.2024	PROPOR 2024	Priberam

4	24-27.6.2024	EAMT2024 (planned)	DW, Priberam

Table 5 *List of Dissemination Events*

As shown in the table above, DW participated in various BDVA workshops and actively contributed to the collaboration of EU projects in the field of Big Data.

3.6 User Events (User Board & User Day)

For SELMA products and results to stay visible and accessible for our stakeholders in different target groups, we regularly organized user events.

As important as the attendance of other technology and research-related events is, the hosting of SELMA-related events gave us extensive opportunities to focus on the dissemination of the results of the SELMA project. By presenting demonstrators and by meeting face-to-face relevant stakeholders for future collaboration and cooperation, we were able to show the relevance of SELMA project achievements for the HLT community.

User & Advisory Board Meeting

On March 24th, 2022, SELMA had its first User Group meeting. It was organized as an interactive online meeting and brought together 31 NLP and media experts from many organizations, including BBC, the University of Edinburgh, COFINA, and AICEP.

Figure 4 *First SELMA User Group Meeting announcement*

The status of SELMA developments and Use Cases were showcased and feedback from the user group was discussed and collected. As an outcome, the Advisory Board was formed, which is a smaller subset of the User Group and consists of representatives of these organizations: BBC, EBU and University of Edinburgh.

First SELMA User Day

The first SELMA User Day had the maxim to "Simplify Life in the Newsroom" with the use of AI and Language Technology. The hybrid event (which also featured a live stream and remote attendants) brought together professionals from all sorts of European media organizations, including ARTE, the BBC, Lusa, and Priberam. It took place on October 13, 2022, in Bonn and was hosted by Deutsche Welle.

Figure 5 *First SELMA User Day in Bonn, Germany*

The User Day had three main parts: interactive presentation sessions, a panel discussion and a workshop/ poster presentation area. The morning block consisting of interactive presentations gave insights into the latest advances into SELMA-related multilingual language technology,

and how it's applied in the field of journalism and media production. At noon, a panel discussion on the "Footprint of AI & NLP Technology in the Media" followed. In the afternoon, participants were invited to test various demo versions, including the two main Use Cases plain X and Monitio, the Podcast Creator, the SELMA open-source tool and a demo on how to segment audio files. LIA invited participants to provide direct feedback on the latest developments on speech synthesis of specific voices. Valuable feedback was being collected from the meeting. One of the participants stated concerning the Monition Diversity Use Case: "This gives me an instant overview of our [editorial] output – and our weak spots. A lot to think about!" In the course of the afternoon, LIA was able to collect enough participant feedback to further improve work on the speech synthesis.

During the first User Day, SELMA could reach out to around 86 persons (up to 58 persons in the stream simultaneously and 28 people physically in the room). The presentations and the panel discussion are also available on demand and have received 182 views so far.

Second SELMA User Day

The SELMA User Day was a two-day meeting focusing on industrial uptake of SELMA outcomes and was a joined event with the FESTIVAL IA with its motto: Multilingualism and Sovereignty from LIA / University of Avignon in France. Language spoken at the event was primarily French. It took place on November 14-15, 2023, in Avignon and was hosted by LIA / University of Avignon.

Figure 6 *Flyer for the second SELMA User Day in Avignon, France*

The first day featured several SELMA presentations about various NLP demonstrators, technologies and tools as well as panel talk about «IA et multilinguisme : enjeux et état des lieux» (AI and Multilingualism: Issues and Current Status).

The second day was a demonstration day with hands-on experiences for NLP related data, tools and technological developments. In several stations, we demonstrated how language technology can be used for media production and news media monitoring and collected valuable feedback from users and potential clients.

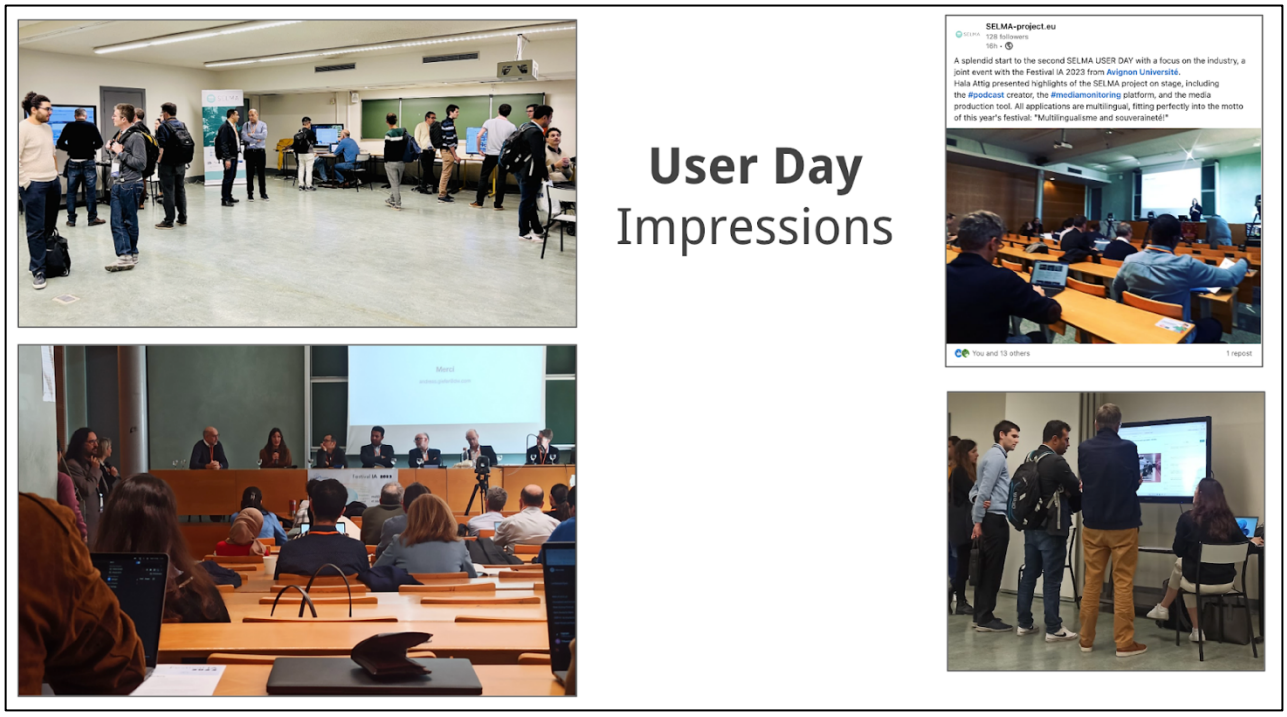

Figure 7 *Impressions from the second SELMA User Day in Avignon, France*

During the second User Day, SELMA could reach out to around 100 persons (40 persons during the first day, almost 60 persons during the second day). All in all, there were up to 20 companies including Airbus, INRIA, The French Defense Ministry, SMF and Facebook / Meta.

Final User Event including User Group Meeting

On March 21st, 2024, we held the Final User Event which also comprised the second User Group meeting. It was a compact 90-minute online session with around 40 participants from various organizations, institutes and companies. It featured a keynote about "AI & Media - What the newsroom needs today" by DW's Senior Technology Manager and AI circle leader Ruth Kühn, followed by key takeaways in SELMA's research and demonstrations of novel NLP prototypes / product advancements in the last three years. People were actively interacting verbally, through chat and through a feedback collection tool (Mentimeter).

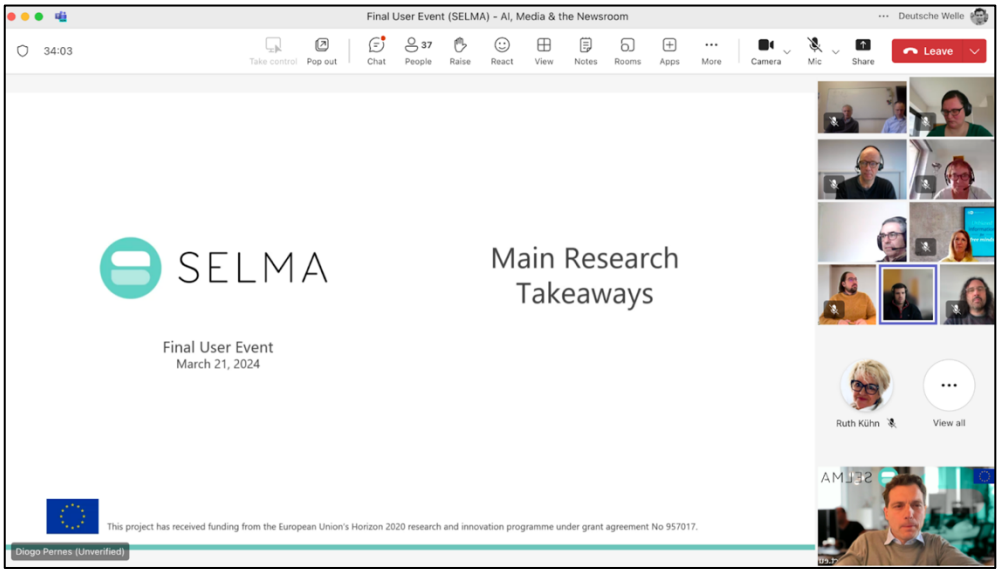

Figure 8 *SELMA's Final User Event held online*

Collecting feedback through different means, here with an interactive tool (Mentimeter).

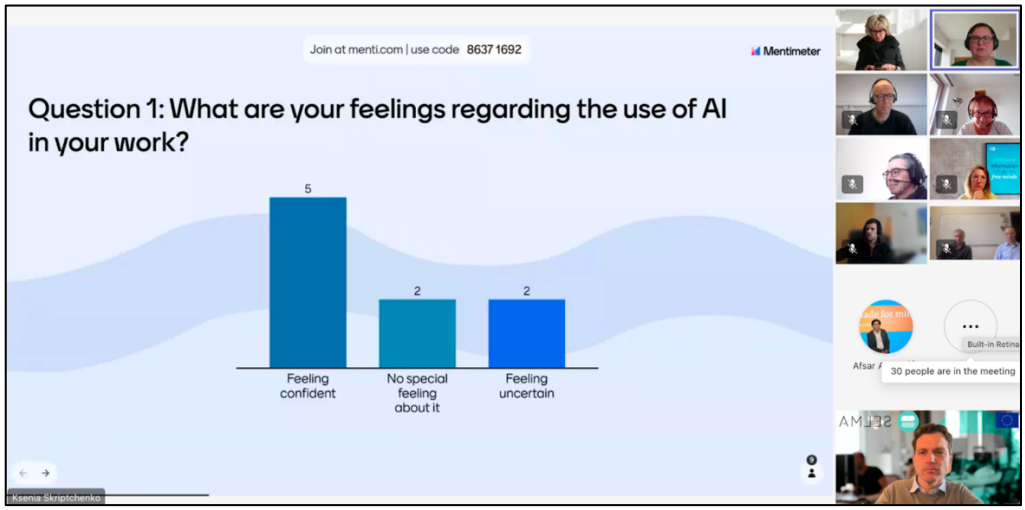

Figure 9 *Gathering Feedback on usage of AI*

3.7 Publications

Publications in the scientific and academic spheres are an essential component of SELMA dissemination activities. A list of 30 publications and research papers published in the runtime of the project is provided below.

#	Consortium Partner/ Names	Title	Published in	Open Access & Golden Standard (+URL)
	2021			
1	Priberam: Pedro Ferreira, Ruben Cardoso, Afonso Mendes	Priberam Labs at the 3rd Shared Task on SlavNER	Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing	Yes
2	LIA: Solène Evain, Ha Nguyen, Hang Le, Marcely Zanon Boito, Salima Mdhaffar, Sina Alisamir, Ziyi Tong, Natalia Tomashenko, Marco Dinarelli, Titouan Parcollet, Alexandre Allauzen, Yannick Estève, Benjamin Lecouteux, François Portet, Solange Rossato, Fabien Ringeval,Didier Schwab, Laurent Besacier	Task Agnostic and Task Specific Self-Supervised Learning from Speech with LeBenchmark	Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks	Yes
3	LIA: Solène Evain, Ha Nguyen, Hang Le, Marcely Zanon Boito, Salima Mdhaffar, Sina Alisamir, Ziyi Tong, Natalia Tomashenko, Marco Dinarelli, Titouan Parcollet, Alexandre Allauzen, Yannick Estève, Benjamin Lecouteux, François Portet, Solange Rossato, Fabien Ringeval, Didier Schwab, Laurent Besacier	LeBenchmark: A Reproducible Framework for Assessing Self-Supervised Representation Learning from Speech	Proc. Interspeech 2021, 1439-1443, doi: 10.21437/Interspeech.2021-556	Yes

4	LIA: Sahar Ghannay, Antoine Caubrière, Salima Mdhaffar, Gaëlle Laperrière, Bassam Jabaian, Yannick Estève	Where are we in semantic concept extraction for Spoken Language Understanding?	Speech and Computer: 23rd International Conference, SPECOM 2021, St. Petersburg, Russia, September 27–30, 2021, Proceedings	Yes
	2022			
1	IMCS: Paulis Barzdins, Audris Kalnins, Edgars Celms, Janis Barzdins, Arturs Sprogis, Mikus Grasmanis, Sergejs Rikacovs, Guntis Barzdins	Metamodel Specialisation based Tool Extension	Baltic Journal of Modern Computing	Yes
2	LIA: Salima Mdhaffar, Valentin Pelloin, Antoine Caubrière, Gaëlle Laperrière, Sahar Ghannay, Bassam Jabaian, Nathalie Camelin, Yannick Estève	Impact Analysis of the Use of Speech and Language Models Pretrained by Self-Supersivion for Spoken Language Understanding	Proceedings of the 13rd Language Resources and Evaluation Conference (LREC)	Yes
3	LIA: Gaëlle Laperrière, Valentin Pelloin, Antoine Caubrière, Salima Mdhaffar, Sahar Ghannay, Bassam Jabaian, Nathalie Camelin, Yannick Estève	The Spoken Language Understanding Media Benchmark Dataset in the Era of Deep Learning: data updates, training and evaluation tools	Proceedings of the 13rd Language Resources and Evaluation Conference (LREC)	Yes
4	LIA: Marcely Zanon Boito, Fethi Bougares, Florentin Barbier, Souhir Gahbiche, Loïc Barrault, Mickael Rouvier, Yannick Estève	Speech Resources in the Tamasheq Language	Proceedings of the 13rd Language Resources and Evaluation Conference (LREC)	Yes

5	LIA: Gaëlle Laperrière, Valentin Pelloin, Antoine Caubrière, Salima Mdhaffar, Nathalie Camelin, Sahar Ghannay, Bassam Jabaian, Yannick Estève	Le benchmark MEDIA revisité : données, outils et évaluation dans un contexte d'apprentissage profond	Journées d'Études sur la Parole - JEP2022	Yes
6	LIA: Hang Le, Sina Alisamir, Marco Dinarelli, Fabien Ringeval, Solène Evain, Ha Nguyen, Marcely Zanon Boito, Salima Mdhaffar, Ziyi Tong, Natalia Tomashenko, Titouan Parcollet, Alexandre Allauzen, Yannick Estève, Benjamin Lecouteux, François Portet, Solange Rossato, Didier Schwab and Laurent Besacier	LeBenchmark, un référentiel d'évaluation pour le français oral	Journées d'Études sur la Parole - JEP2022	Yes
7	LIA: Solène Evain, Ha Nguyen, Hang Le, Marcely Zanon Boito, Salima Mdhaffar, Sina Alisamir, Ziyi Tong, Natalia Tomashenko, Marco Dinarelli, Titouan Parcollet, Alexandre Allauzen, Yannick Estève, Benjamin Lecouteux, François Portet, Solange Rossato, Fabien Ringeval, Didier Schwab and Laurent Besacier	Modèles neuronaux pré-appris par auto-supervision sur des enregistrements de parole en français	Journées d'Études sur la Parole - JEP2022	Yes
8	LIA: Marcely Zanon Boito, John Ortega, Hugo Riguidel, Antoine Laurent, Loïc Barrault, Fethi Bougares, Firas Chaabani, Ha Nguyen, Florentin Barbier, Souhir Gahbiche, Yannick Estève	ON-TRAC Consortium Systems for the IWSLT 2022 Dialect and Low-resource Speech Translation Tasks	Proceedings of the IWSLT 2022 (ACL Anthology)	Yes
9	LIA: Antonios Anastasopoulos, Loïc Barrault, Luisa Bentivogli, Marcely Zanon Boito, Ondřej Bojar, Roldano Cattoni, Anna Currey, Georgiana Dinu, Kevin Duh, Maha Elbayad, Clara Emmanuel, Yannick Estève, Marcello Federico, Christian Federmann, Souhir Gahbiche, Hongyu Gong, Roman Grundkiewicz, Barry Haddow, Benjamin Hsu, Dávid Javorský, Věra Kloudová, Surafel Lakew, Xutai Ma, Prashant Mathur, Paul McNamee, Kenton Murray, Maria Nădejde,	Findings of the IWSLT 2022 Evaluation Campaign	Proceedings of the IWSLT 2022 (ACL Anthology)	Yes

	Satoshi Nakamura, Matteo Negri, Jan Niehues, Xing Niu, John Ortega, Juan Pino, Elizabeth Salesky, Jiatong Shi, Matthias Sperber, Sebastian Stüker, Katsuhito Sudoh, Marco Turchi, Yogesh Virkar, Alexander Waibel, Changhan Wang, Shinji Watanabe			
10	Priberam: João Santos, Afonso Mendes, Sebastião Miranda	Simplifying Multilingual News Clustering Through Projection from a Shared Space	Proceedings of the Text2Story (Fifth Workshop on Narrative Extraction from Texts held in conjunction with the 44th ECIR, 2022)	Yes
11	LIA: Marcely Zanon Boito, Laurent Besacier, Natalia Tomashenko, Yannick Estève	A Study of Gender Impact in Self-supervised Models for Speech-to-Text Systems	Proceedings of the INTERSPEECH 2022	Yes
12	LIA: Salima Mdhaffar, Jarod Duret, Titouan Parcollet, Yannick Estève	End-to-end model for named entity recognition from speech without paired training data	Proceedings of the INTERSPEECH 2022	Yes
13	Priberam: Pedro Henrique Martins, Zita Marinho, Andre Martins	∞-former: Infinite Memory Transformer	Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)	Yes
14	LIA: Gaëlle Laperrière, Valentin Pelloin, Mickaël Rouvier, Themos Stafylakis, Yannick Estève	On the Use of Semantically Aligned Speech Representations for Spoken Language Understanding	Publication in the SLT 2022: IEEE Workshop on Speech and Language Technology 2022	Yes

15	IMCS: Eduards Mukans, Gus Strazds, Guntis Barzdins	RIGA at SemEval-2022 Task 1: Scaling Recurrent Neural Networks for CODWOE Dictionary Modeling	Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022), July 14-15, 2022. Seattle, United States	Yes
16	IMCS: Arturs Znotins, Roberts Dargis, Normunds Gruzitis, Guntis Barzdins, Didzis Gosko	RUTA: MED - Dual Workflow Medical Speech Transcription Pipeline and Editor	Proceedings of the NLDB 2022: Natural Language Processing and Information Systems. Lecture Notes in Computer Science book series (LNCS, volume 13286)	Yes
17	Priberam: Diogo Pernes, Afonso Mendes, André F.T. Martins	Improving abstractive summarization with energy-based re-ranking	Proceedings of the GEM workshop at EMNLP 2022	Yes
	2023			
1	Priberam: Raul Monteiro and Diogo Pernes	Towards End-to-end Speech-to-text Summarization	Publication in a Conference Proceeding (TSD2023)	Yes
2	IMCS: Eduards Mukans and Guntis Barzdins	RIGA at SemEval-2023 Task 2: NER Enhanced with GPT-3	Proceedings of the The 17th International Workshop on Semantic Evaluation (SemEval-2023)	Yes
3	LIA:	LeBenchmark 2.0: a Standardized, Replicable and Enhanced Framework	Computer Speech & Language	Yes

	Titouan Parcollet, Ha Nguyen, Solène Evain, Marcely Zanon Boito, Adrien Pupier, Salima Mdhaffar, Hang Le, Sina Alisamir, Natalia Tomashenko, Marco Dinarelli, Shucong Zhang, Alexandre Allauzen, Maximin Coavoux, Yannick Estève, Mickael Rouvier, Jerôme Goulian, Benjamin Lecouteux, Francois Portet, Solange Rossato, Fa- bien Ringeval, Didier Schwab, Laurent Besacier	for Self-supervised Representations of French Speech		
4	LIA: Jarod Duret, Benjamin O'Brien, Yannick Estève, Titouan Parcollet	Enhancing expressivity transfer in textless speech-to-speech translation	2023 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)	Yes
5	LIA: Antoine Laurent, Souhir Gahbiche, Ha Nguyen, Haroun Elleuch, Fethi Bougares, Antoine Thiol, Hugo Riguidel, Salima Mdhaffar, Gaëlle Laperrière, Lucas Maison, Sameer Khurana, Yannick Esteve	ON-TRAC consortium systems for the IWSLT 2023 dialectal and low-resource speech translation tasks	International Conference on Spoken Language Translation (IWSLT) 2023	Yes
6	LIA: Gaëlle Laperrière, Ha Nguyen, Sahar Ghannay, Bassam Jabaian, Yannick Estève	Specialized Semantic Enrichment of Speech Representations	IEEE ICASSP 2023 workshop on Self-supervision in Audio, Speech and Beyond	Yes
7	LIA: Tuan Nguyen, Salima Mdhaffar, Natalia Tomashenko, Jean-François Bonastre, Yannick Estève	Federated Learning for ASR based on Wav2vec 2.0	2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)	Yes
8	LIA: Jarod Duret, Titouan Parcollet, Yannick Estève	Learning multilingual expressive speech representation for prosody prediction without parallel data	12th ISCA Speech Synthesis Workshop (SSW2023)	Yes

9	Priberam: Simão Gonçalves, Gonçalo Correia, Diogo Pernes, and Afonso Mendes	Supervising the Centroid Baseline for Extractive Multi-Document Summarization	In Proceedings of the 4th New Frontiers in Summarization Workshop, pages 87–96, Singapore. Association for Computational Linguistics.	Yes
	2024			
1	LIA: Salima Mdhaffar, Fethi Bougares, Renato De Mori, Salah Zaiem, Mirco Ravanelli, Yannick Estève	TARIC-SLU: A Tunisian Benchmark Dataset for Spoken Language Understanding	LREC-COLING 2024	Not yet (paper will be available in the LREC proceedings in May)

Table 6 *List of Publications*

3.8 Publication Highlights/Awards

SELMA Partner LIA received an award/prize of 2000 USD from Cambridge University (founder of the Flower toolkit) for the paper ICASSP 2023 “Federated Learning for ASR based on Wav2vec 2.0”.

Innovation Radar

Three SELMA project outcomes (as listed in the table below) were submitted and selected by the EU Innovation Radar Platform. In early 2023, the applications were added to a list of EU-based innovations.

Application	Description
plain X	A novel tool for easy transcription, translation, subtitling & voice-over
Scalability Module	Highly scalable SELMA NLP orchestration platform for processing extreme volumes
Podcast Creation	Create News Podcasts using NLP analysis and summarization

Table 7 *List of Innovations submitted to the EU Innovation Radar*

4. Communication

SELMA's communication initiatives are described in this section. There are three sections to it. The communication strategy establishes the project identity and stipulates rules for internal and external communication.

This section's "Evaluation" subsection gauges and assesses the project's performance.

Finally, the "Sustainability" chapter offers details on how the project's outcomes and results can be accessed and made available via publicly accessible channels beyond its active phase.

4.1 Strategy

SELMA’s communication strategy described the channels and methods used to efficiently implement communication:

- within the project (internally),
- with peer researchers, related projects,
- with end users and the public at large.

It was important to ensure that the entire consortium was informed of the development status and achievements, also of the challenges to overcome, and the efforts invested. Equally important was the communication towards the relevant research community, related projects, and our main target group consisting of media and broadcasting professionals to inform and ensure that the SELMA results can be used in professional communities outside of the consortium and relevant feedback can be provided during the project’s lifespan.

The SELMA output can be used by a wide range of users interested in consuming content in other languages, using subtitles, full-text transcripts, or voice-over applications. To ensure the continued use of the outcomes and results, it was success-critical to inform the possible end users and general public about advantages, limitations, and overall progress of the SELMA’s outcome.

4.2 Evaluation & Reporting

The table below evaluates the progress on the communication means and outreach:

Activity	Planned (in total)	Description	Achieved in total
Website Visits	24.000	Y1: 7.000 Y2: 7.000 Y3: 10.000	41,409
Twitter Followers	180	Y1: 60 Y2: 60 Y3: 60 (updated 0, negative tendency)	102
Tweets	150	Y1: 50 Y2: 50 Y3: 50 (updated 5)	105
LinkedIn Followers	90	Y1: 30 Y2: 30 Y3: 30 (updated 70)	137
LinkedIn Posts	50	Y1: 15 Y2: 15 Y3: 20	50
GitHub Stars	30	Y1: 0 Y2: 0 Y3: 30	9
Videos	10	Y1: 0 Y2: 5 Y3: 5	9
Poster	1	(updated version if required)	1
Flyer	1	(updated version if required)	1

Roll-Up	1	(updated version if required)	1
Events (participated and organized)	55	Y1: 15 Y2: 20 Y3: 20	56
Publications	11	Y1: 3 Y2: 4 Y3: 4	30
User Events	4	Y1: 0 Y2: 2 Y3: 2	4

Table 8 *Overview Communications Means & Progress*

As shown in the table above, almost all previously set targets were met or even exceeded. With more than 20 publications in the second year, we have significantly more scientific output than anticipated. The numbers on X (formerly Twitter), which was previously seen as our primary communication channel, fell short of expectations. This is due to the current situation with X's takeover and its ongoing internal strategy changes. We closely monitored developments in order to respond quickly. Starting the third year of the project, we placed a greater emphasis on our LinkedIn channel and shifted our communication there.

SELMA's GitHub channel was made accessible to the public relatively late in the project's timeline, leading to an inability to achieve the anticipated number of stars by the project's conclusion. Nonetheless, based on the positive reception of the project outcomes, we anticipate a surge in interest in the coming months. This expectation is supported by the observed trend of accumulating interest, exemplified by the attainment of 10 stars within the last few months following the channel's activation and promotion.

4.3 Sustainability

To ensure the availability of the work performed by the SELMA project, the consortium aims to keep dissemination channels available for a period of at least three years after the end of the project. This will include the website and the social media channels.

Although the editorial input will be kept to a minimum after the project, all channels of distribution shall be made and kept accessible when the project is completed.

The SELMA Open-Source platform is publicly available up to two years after the ending of the project. It can be accessed under https://selma.ailab.lv/#

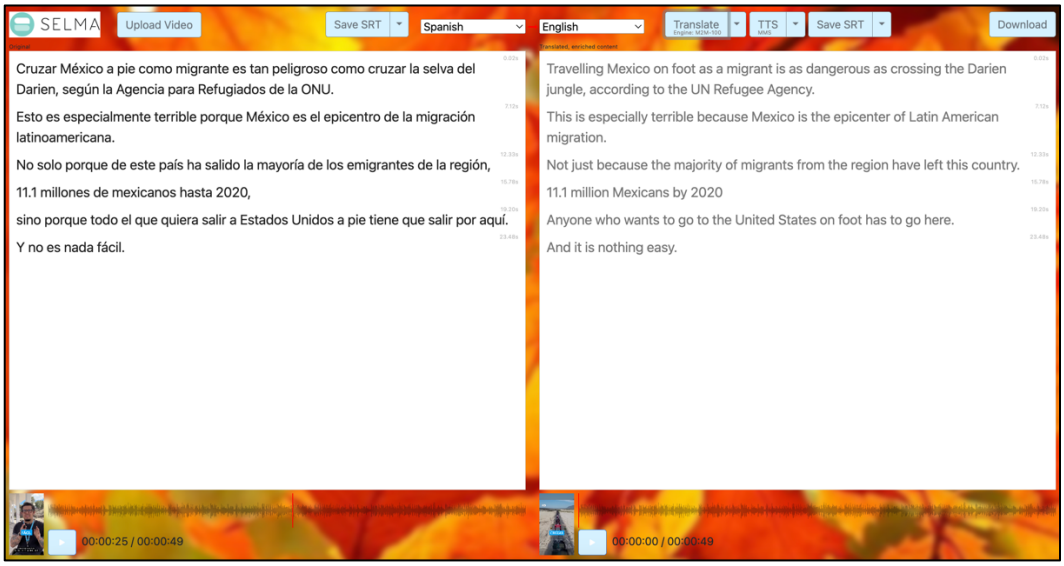

Figure 10 *SELMA's open source platform*

5. Exploitation

This section of the deliverable gathers the information required and explores options to ensure the output of SELMA is further exploited by consortium partners and others. The exploitation activities started in the second year and were a focus activity in the final part of the SELMA project, including IPR Management (see 4.4).

5.1 Exploitation Ways

There are four main ways in which SELMA is exploited:

- **Orchestration platform**: The focus was to provide an open-source big-data platform able to ingest and orchestrate the pipeline graph of NLP modules and apply stream learning and user feedback.
  - → The platform has been released as SELMA open-source platform and is currently available at https://selma.ailab.lv/; it can also be downloaded and can be run locally
  - → Additionally, the two Use Cases /platforms Monitio and plain X are now leveraging the SELMA orchestration platform core. Monitio and plain X are available and being exploited as two separate products (for more information, see section 5.4 Monitio exploitation and 5.5 plain X exploitation)
- **Component-based/individual system modules:** The SELMA components/ modules have a high potential value as improvements for existing services, or as the basis for other new services.
  - → Including the platforms, SELMA developed 27 components, which are either available via on open-source license (6) or proprietary (21). For more information, see section 5.8 IPR Management.
- **Integration into other projects/products:** Results have been integrated in the Monitio and plain X products.

→ During SELMA, plain X has been developed and improved (UI, improved models, use case development. See D1.4). This Use Case has emerged into a product. The product plain X is being exploited jointly by DW and Priberam and taken to market, having already real clients outside the consortium (Section 5.5). Within DW, plain X is also customized and exploited as a major DW HLT platform, a one-stop shop for different language processes. Section 5.5 DW / Media production (plain X).

→ SELMA Research results have been integrated in the Monitio product (Improved models and related UI, See D1.4), a Media Monitoring platform under development and commercialization by Priberam, which was productized during the MONITIO project (an H2020 FTI project for AI powered Media Monitoring lead by Priberam).

- **Knowledge**: Through continuous research, each consortium partner is learning and developing new techniques that can be applied to succeeding projects. It can also be utilized to design new approaches and new services.

  → Every SELMA partner brought knowledge into their organizations / companies, either through technology, workshops, or tools. For more information, see Section 5.5 SELMA exploitation (per Partner).

5.2 Exploitation: Open-source & commercial approach

SELMA worked on various output formats including software components, platforms and datasets. An overview of the exploitation approach (open-source versus proprietary) is shown in the following two tables.

Overview of Technology and Components

#	ID	Technology	Component	Partner	OS	Propr.
1	1.1	ASR	Speech Recognition French	LIA		**X**
2	1.2		Speech Recognition Urdu	LIA		**X**
3	1.3		Speech Recognition Latvian	IMCS	**X**	
4	1.4		Speech Recognition German	FhG	**X**	
5	2.1	Text, Speech MT	M2M-100 Machine Translation (Integration)	IMCS / Priberam	**X**	
6	3.1	News Summarization	English Monolingual Abstractive Summarization	Priberam	**X**	
7	3.2		Crosslingual Abstractive Summarization	Priberam	**X**	
8	3.3		Crosslingual Multidocument Extractive Summarization	Priberam	**X**	
9	3.4		Speech Summarization	Priberam	**X**	
10	4.1	NER & NEL	PiniTree Ontology Editor	IMCS		**X**
11	4.2		Multilingual Hierarchical nested NER	Priberam		**X**
12	4.3		Entity representations for 20M Wikidata entities	Priberam		**X**
13	4.4		Entity Linking	Priberam		**X**
14	5.1	(Autom.) Post Editing	Automatic Post-Editing	FhG	**X**	
15	5.2		Speech2Text PostEditor From User Feedback	Priberam		**X**
16	6.1	Clustering	Online Crosslingual Clustering	Priberam	**X**	
17	7.1	Topic Detection	Multilingual IPTC Topic Classification	Priberam		**X**
18	8.1	Speech Synthesis	Text To Speech for Latvian	IMCS, LIA		**X**
19	8.2		Text To Speech for Brazilian	IMCS, LIA	**X**	
20	8.3		Text To Speech for Urdu	IMCS, LIA	**X**	
21	9.1	Story Segmentation	Story Segmentation	FhG	**X**	
22	10.1	Punctuation & Truecasing	Punctuation and Casing Recovery	FhG	**X**	
23	11.1	Speaker Diarization	Speaker Diarization	FhG		**X**

24	12.1	Speaker Recognition	Speaker Recognition (Identification)	FhG	**X**	
25	13.1	Graph Orchestrator platform (Maestro)	Graph Orchestrator (Maestro)	Priberam	**X**	
26	14.1	Monitio platform (UC1)	Monitio platform	Priberam		**X**
27	15.1	plain X platform (UC2)	plain X platform	Priberam		**X**
28	16.1	SELMA OSS platform (UC0)	Use Case 0 – SELMA Open Source Platform	IMCS	**X**	

Table 9 *Technology / Components Overview - Open-Source vs Proprietary*

Overview of SELMA Datasets and Datastreams

#	SELMA Datasets & Datastreams	Main Purpose	Format*				Volume - hours, # of documents	Release Level**			Partner
			T	A	S	V		L1	L2	L3	
1	Turkish	NER	x				100 docs	x			DW / Prib
2	Dutch	NER	x				50 docs	x			DW / Prib
3	Ukrainian	NER	x				300 docs	x			IMCS / DW / Prib
4	Russian	NER	x				160 docs	x			IMCS / DW / Prib
5	Latvian	NER	x				740 docs	x	x	x	IMCS / Prib
6	Amharic	LR scripts for ASR		x	x	x	10 hrs	x			DW
7	Bengali	LR scripts for ASR		x	x	x	5 hrs	x			DW
8	Urdu	News Training data Voices		x	x		10 hrs	x			DW
9	Brazilian Portuguese	News Training data Voices		x	x		96 hrs	x			DW
10	DW AV (19 lang.)	SELMA Foundation	x	x		x	15000 hrs	x			LIA / DW
11	Wikipedia / Wikidata	Enitity representation	x				40 Mio docs	x			Priberam
12	Monitio News	Datastream	x		x		300.000 / day	x			Priberam

* TASV = Text, Audio,Script,Video, ** L1 = Project internal, L2 = For research, L3 = Public domain

Table 10 *Datasets Overview and Release Level*

The Latvian Dataset comprising 741 Texts of hierarchically annotated named entities in Latvian news articles was published in the public domain and is available on Clarin: https://repository.clarin.lv/repository/xmlui/handle/20.500.12574/98.

5.3 Monitio exploitation of SELMA contributions

SELMA contributes to the Monitio platform with several components to enhance its multilingualism, changing the paradigm of translating to English and then analyzing, on which the first versions from SUMMA used to operate. The new orchestration component allows it to scale, be more versatile and configurable for different NLP pipelines. The new multilingual clustering and the new topic detection model significantly expanded the quality and number of languages. The integration of the ASR models into the pipeline allows the platform to ingest and analyze multimedia content. SELMA online representations allow discovery and disambiguation of entities in over 30 languages. The new multilingual Named Entity Recognition models allow detection of mentions in more than 100 languages. The feedback model for correcting entities and mentions enables users to correct entities and apply those corrections on future documents. In addition, the UI has evolved to incorporate these new capabilities.

During SELMA, the platform has been used by DW in its internal workflows for monitoring its content production, to evaluate the diversity of its publications, to build podcasts based on the trending stories and to monitor the news externally. During this time, DW has delivered feedback and new requirements, many of which are now available to the users.

The new storyline clustering and the platform have been extensively used by the “Barómetro” a publication made by ISCTE a Portuguese University to report on the published news in Portugal (https://medialab.iscte-iul.pt/barometro/).

AICEP, the Portuguese agency for external commerce, was the first paying client. They use the platform to observe how the news on selected external markets and verticals affect Portuguese external commerce and inform Portuguese companies about opportunities or risks.

PMG, a German media monitoring company, has been evaluating the platform both from the point of view of its processing capabilities to integrate in their systems and as a new service to their customers.

The above are examples of a continued effort at Priberam to exploit the platform as a product for media monitoring.

As of this year, Priberam is starting to move its product LegiX to use Monitio as the product platform. LegiX is the major legal research tool in Portugal with a large client base, covering most of the top Law Firms and a large penetration in the public sector.

The platform is now ingesting open-source scientific publications from the pharmaceutical and medical sector and is under evaluation at BIAL the biggest pharma company in Portugal. For this new use case, the models developed in SELMA have been used to train new modules to extract medical entities (diseases, genes, proteins, etc..), to cluster scientific publications and to do topic detection using the medical coding ontology ICD9 (International Classification of Diseases).

Monitio is designed for professional media monitoring and goes beyond today's solutions. The core idea of Monitio is to enable advanced filtering of content, keywords or concepts across different languages. As such, Monitio is a logical next step in a world where media offerings are increasingly consumed not via fixed channels but through various digital platforms and social media in multiple languages. Key features of Monitio enable refined media monitoring well beyond simple keyword tracking. Instead, Monitio can search through vast content repositories in various languages, and filters can be used to refine the output further. In addition, based on evolving needs in the field of media monitoring, the platform simplifies reporting. It is possible to generate reports fully automatically based on sophisticated search and filter options. For instance, Monitio can be used for instant alerts should certain keywords be mentioned, even in other languages. Using the platform, templates can be used to automatically generate media clippings, topical reports, or even regular newsletters with a minimum of effort and a wider coverage of content in multiple languages.

5.4 plain X exploitation

plain X is a 4-in-1 content adaptation platform. The software integrates four steps of the process: Transcription, translation, subtitling and (synthetic) voice-over. The software uses AI technology to enable faster content adaptation by saving many manual steps. At the same time, plain X relies on the „human in the loop“ concept, meaning that after each step, there are options for human experts to check, correct and optimize the results. This is not mandatory, we have use cases that process the content in a fully automated manner (for instance, subtitled news provided to Frankfurt airport) and other use cases that require detailed reviewing and corrections (for example long-form documentaries).

plain X has been implemented into DW's technical and editorial workflows based on an extensive refinement process that took over a year to cover all the insights and needs of a multilingual newsroom. For example, for illustration, if there is an editorially controlled transcript, using this often results in higher quality than even the most advanced speech-to-text technologies. Accordingly, it is possible to upload such transcripts.

plain X is engine agnostic, which means that users can select from a variety of powerful platforms such as Google, Facebook, Open AI, DeepL or Speechmatics. Should new language engines become available from established or new players, these can be integrated into plain X. This variety results in a considerably large offering of languages which can be handled for transcription and translation. In addition, plain X is well positioned as a platform for new research and features, in the very dynamic and evolving AI field.

All engines are connected through an API to the plain X user interface so that plain X has a key role as a controllable and enhanceable user interface to access all these engines productively. Users have multiple options to use specific engines for certain steps. Additional options are to set preferences for one specific engine or to re-run a segment if too many things could be corrected. A particular area of development is the integration of low-resource languages through API integration of engines covering more low-resourced languages such as Amharic or Galician. In the future, the expected outcome is that small teams might use HLT specifically to enhance the quality of languages, which is not as important to large language providers.

Even at this early stage, the plain X software allows for considerable automation. For example, if the user needs Spanish subtitles for a German original video, the software will automatically create a transcript, do the translation, and then generate the subtitles based on specific rules of large platforms such as Netflix or the BBC and in a design style that matches the brand of a specific media organization.

The automation has already been put to use for another use case: At Frankfurt Airport, hundreds of screens are showing DW news content, but with subtitles because audio would not be an option in the busy hallways. To produce these subtitles, DW videos are uploaded with editorially controlled transcripts (not based on speech-to-text). Based on the quality input, the following steps can be done automatically so that new videos are transmitted to the airport day by day.

Another more minor but exciting use case: Researchers at the University of Bremen used plain X to transcribe extended research interviews of up to two hours and reported a straightforward workflow with excellent results. In the future, additional use cases like this are likely to evolve. The critical task is to put new HLT features to use, in order to achieve the highest possible quality of language adaptations.

To grow the user base further, plain X is presented at conferences and in media business contexts. Feedback so far is very positive, specifically based on the options for humans to check the output and the perspective that new engines, new features can be implemented to enhance the quality even further. In this way, the tool ensures that work done in research is not lost, but put to good use.

The current status for plain X is as follows: The software is technically implemented at DW, with some 1,000 accounts, the graph in the Figure below shows the evolution of usage at DW. Hundreds of videos are used monthly. Other active clients are LUSA, the news agency of Portugal, Publico a Portuguese newspaper, Media Livre and Impresa.

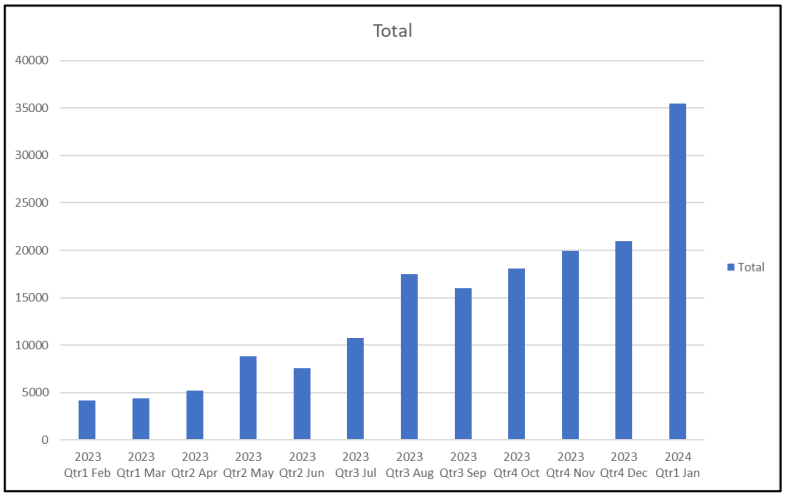

Figure 11 DW usage by minutes of ingested video since the beginning of 2023

Users can start using plain X in several steps: The first step is usually a presentation and demo, followed by initial testing by selected users to identify the specific needs and potential use cases. The second step is a paid three-month trial in order to find out which users will want to work with the software and the average number of language tasks (text, audio, video). Based on this assessment, plain X can be used based on the number of minutes/hours in certain tiers. In summary, this unique approach combines commercial usage with research activities that promote sustainable use.

5.5 SELMA exploitation (per Partner)

5.5.1 LIA

The LIA has significantly benefited from its participation in the European project SELMA. This collaborative endeavor has not only bolstered the visibility of the LIA but has also catalyzed advancements in the field of speech processing, particularly through the development of the SpeechBrain open-source project.

One of the primary advantages derived from the SELMA project is the enhanced visibility of the LIA achieved through numerous scientific publications. These publications not only showcase the expertise and contributions of the LIA team but also serve to amplify its presence within the academic and research communities.

Furthermore, the resources allocated to the development of SpeechBrain under the SELMA project have yielded substantial benefits. SpeechBrain, being an open-source initiative, has empowered the broader scientific and technical community engaged in automatic speech processing. By providing accessible tools and frameworks, SpeechBrain fosters collaboration and innovation across diverse research endeavors. The strengthening of SpeechBrain through the SELMA project has also garnered recognition for the LIA among industrial partners. The practical applications and innovations stemming from SpeechBrain's development have positioned the LIA as a valuable collaborator within the industrial landscape, thereby enhancing its standing and influence in relevant sectors.

Moreover, the LIA's involvement in the SELMA project has facilitated the support of two doctoral theses. This support not only enables the advancement of knowledge within the domain of speech processing but also provides valuable opportunities for emerging researchers to contribute to cutting-edge research initiatives.

5.5.2 DW

DW strongly benefits from the SELMA project and its outcomes. Mainly through the two tools / Use Cases (Monitio, plain X), and its use case application prototypes, but also through the freely accessible SELMA OS platform, where basic concepts were easily demonstrable, although we are not describing SELMA OS in more detail in this section. DW also was able to showcase various NLP components in an integrated manner (Prototypes, Apps) to many editorial departments and initiated discussions and ongoing impact (see section Prototypes).

Tools

DW successfully managed to enhance AI and speech-driven work within the company based on the two Use Cases, i.e., Media Production (plain X) and Media Monitoring (Monitio). At the beginning of the SELMA project, an HLT platform was already under implementation at DW (resulting from funding projects). During the course of the project, this platform benefited from various technological developments (e.g. in terms of improved ASR, UI and speech generation) which were eventually integrated into the platform. Also, the Monitio Monitoring platform has been introduced and shown to several editorial departments. Through various evaluation and feedback cycles, the UI was improved and geared towards journalistic needs. Although full use and integration into the DW environment has not been achieved yet, pending trialing of the recently launched enhanced version and transcription and translation of all DW languages, the Monitio platform is expected to be used as a monitoring tool at Deutsche Welle, based on feedback following tests and demonstrations.

Media Production (plain X)

For DW, a media production tool such as plain X is very valuable since it can serve as a one-stop shop for media localization, adaption (translation into other languages), subtitling and voiceover.

At the end of the SELMA project - but already in the course of it - various technologies, engines and User Interaction findings / improvements were integrated into the DW HLT platform such as:

- Improved ASR in many languages including Low- and High-Resource Languages (LRL / HRL)
- Speaker diarization
- Improved Voice-Over capabilities (tune voices)
- Improved User Interface including personalization

SELMA was very important for LT (Language Technology) activities at Deutsche Welle. In conjunction with a new strategy for (semi-automatic) subtitling, all of its current DW content by 2025, the DW HLT platform experienced a major boost through SELMA. At the time being, the DW tool has registered almost 1000 accounts. It is also a crucial tool for Deutsche Welle to reach its accessibility objective of subtitling 100 percent of its on-demand content, in all of its publication languages (currently 32), by the end of 2025.

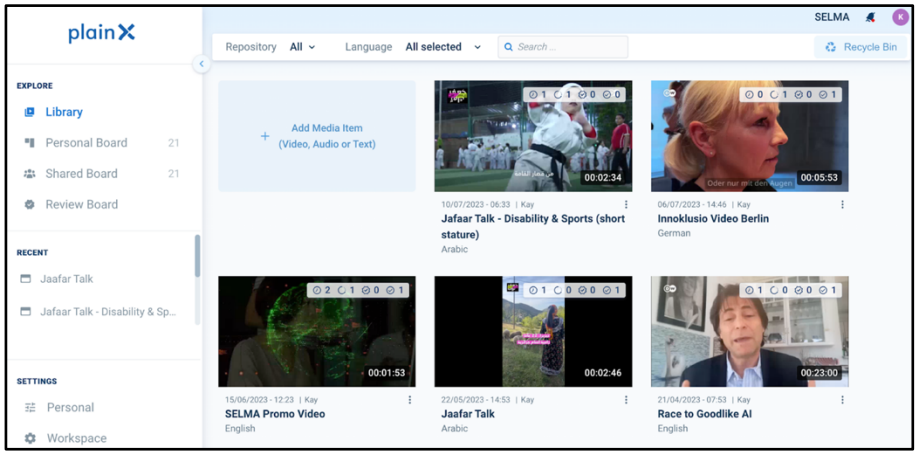

Figure 11 *Screenshot of the plain X Media Production platform with DW content*

Media Monitoring (Monitio)

Monitio was introduced and discussed at DW in many departments (e.g. language departments, archive, management). For DW, it is of high interest to be able to monitor its own output in all its 32 languages across its various channels and media formats (text, audio and video). With

the Monitio Monitoring tool, enabling content collection and language-agnostic topic analysis and clustering, many of these tasks can be done in an efficient manner.

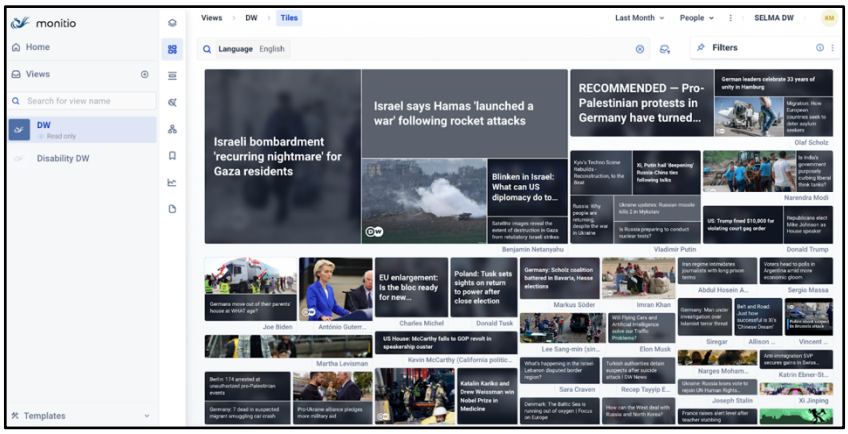

Figure 12 *Screenshot of the Monitio Media Monitoring platform showing DW content in the SELMA user area*

The Monitio platform has been tested and assessed by several departments who have shown interest in continuing to use it in their daily operations. With the new release of March 2024, after a major update of the platform, it is anticipated it will be applied gradually in DW for both internal and external monitoring. Several use cases have been identified.

Several editorial departments, such as the Hindi department, find this a very useful tool for external monitoring, whereby they set up and run a saved search on a daily basis to get an immediate overview of trending topics in selected sources from a certain region, for instance India. This helps them to determine what topics need to be covered that day/week and what stories to write. Crucial here is to get access to external sources in the region of coverage.

Another use case is that of internal monitoring, in which management as well as editorial departments can easily get an overview of what has been published in a specific region or language/group of languages. This has been a major challenge up to now, due to the large number of languages in which DW publishes material. Having access to a tool that can efficiently analyze content in all required languages, in all formats (text, video and audio) and from various publication channels (web, social media, feeds), and present the results in the original language of publication as well as a language of choice, opens up numerous paths of exploitation. DW content is already continuously being ingested into Monitio and analyzed,

and especially the transcription and translation of video and audio material into a common language, is extremely compelling. A pilot phase is expected to be launched in the near future to explore this further.

Prototypes

During the SELMA project, DW worked on and improved 6 Use Case Applications / prototypes. The development of these prototypes was used to test the impact of NLP technology and tools in a journalistic setting such as Deutsche Welle: Podcast Creator, Diversity Monitoring, DW Benchmarking, DW Summarizer, DW Voices and DW Avatar.

#	Use Case Application / Prototype	What it does	Main features / technologies / models
1	**Podcast Creator**	Creates a news podcast on the fly	Includes workflow (music, structure). Imports story clusters from Monitio. Uses summarization and speech synthesis
2	**Diversity Monitoring**	Analyzes Binary Gender, Age and Regional Origin	Uses Wikidata information (connects to Monitio Use Case)
3	**DW Benchmarking**	Compares ASR, MT & VO	Test tool for various NLP models including SELMA outcomes; automatic and manual ingestions
4	**DW Summarizer**	Summarizes Text	API connections to OpenAI, Alpaca, LLaMA and SELMA summarization
5	**DW Speaker App**	Generates Speech	Many languages, many voices, including customized SELMA voices
6	**DW Avatar**	Creates adaptions with animated avatar	Uses SELMA voices to create various language versions

Table 12 *Overview of prototypes introduced to DW*

Podcast Creator

The Podcast Creator is an application under development. It is an app to (semi-)automatically create Podcasts with synthetic voices while being connected via an API to the Monitio platform. Articles can be "fetched" from Monitio and can be summarized – either manually or automatically. Then the system weaves everything together, including company-specific jingles, and creates a downloadable Podcast within seconds. At DW, two editorial departments are working with the app in a pilot phase. The feedback gathered was very positive. Whether or not an actual use will be realized is also dependent on the general AI strategy at DW which currently considers which AI technologies to use.

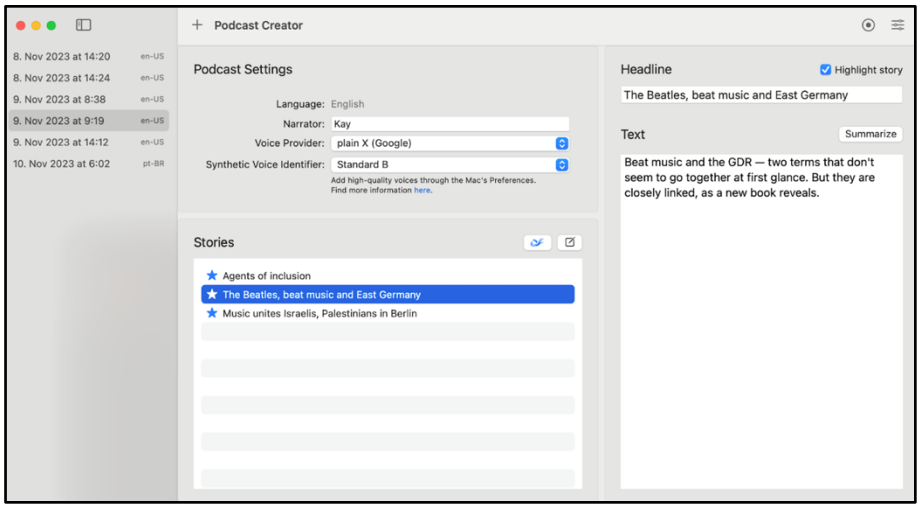

Figure 13 *Screenshot of the SELMA Podcast Creator*

Diversity Monitoring Prototype

DW serves a diverse audience in 32 languages across many regions of the world. It aims to create content which is relevant and relatable to people reading, watching or interacting with it. The Diversity Monitoring Prototype aims to measure the diversity of content along three axes: binary gender, age and regional distribution by counting appearances ("mentions") based on Wikidata labeling. It serves as a sensitivity tool to improve gender-balanced media output and has been proven valuable for one editorial department which has used it as a daily tool for

editorial decision making for several months. DW is currently promoting it to spread its use across more DW units – also as a complementary to the 50:50 The Equality Project initiated by the BBC, which manually counts representations of under- or low-represented groups (such as women, people of color, people with disabilities).

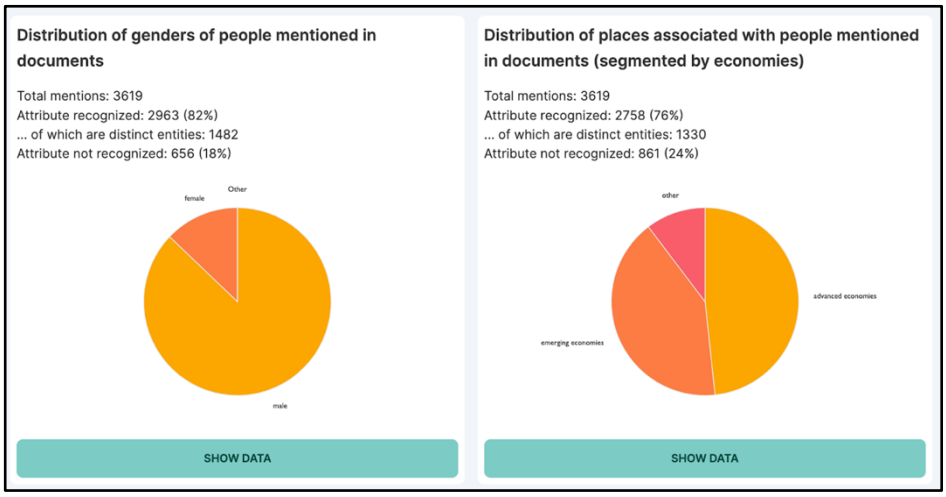

Figure 14 Screenshot of the SELMA / Diversity monitoring prototype

DW Benchmarking

The benchmarking activities at Deutsche Welle aim to improve the use of technology and tools for users by identifying and selecting the best engines and machines. They have significantly benefited from SELMA. The current focus is on finding the best engine for transcription (ASR), Translation (MT) and Voice-Over (VO). In the course of the project, many engines have been added and insights have been collected, based on manual as well as automatic testing. The results were used to feed the platforms (especially plain X) and prototypes and to simplify interactions (providing shortcuts) with the tools. Next phases include the enhancement with a hallucination matrix through generative AI (e.g. summarizations). There are current drafts to establish and integrate the continuous development of the benchmarking platform into the infrastructure of Deutsche Welle.

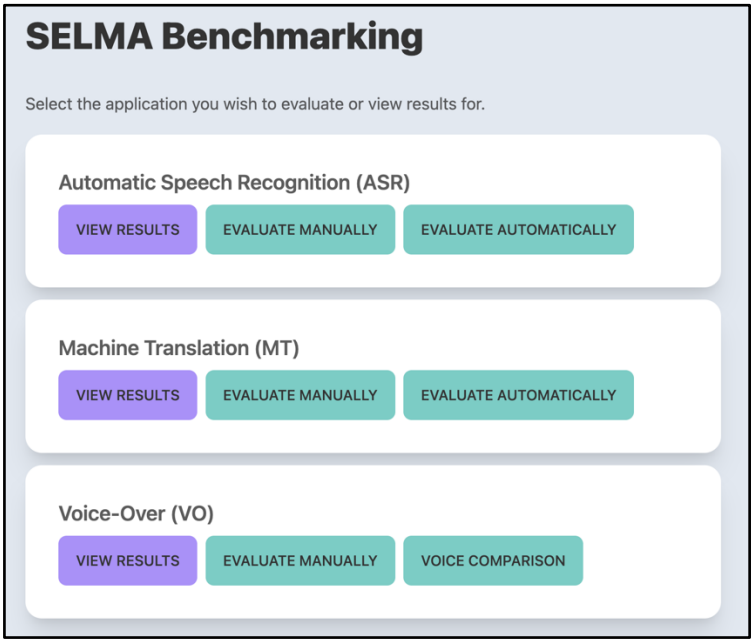

Figure 15 Screenshot of the SELMA / DW Benchmarking platform

Summarizer App

The Summarizer is a beta app fully developed within SELMA. It serves a single purpose: DW journalists can summarize and compare texts using four different engines: OpenAI, Alpaca, LLaMA and Priberam (SELMA). The aim is to get an impression of the quality of the output for different languages. Results can be fed back into available prototypes (such as the Benchmarking system or the Podcast creator). Initial test feedback has shown that for certain purposes abstractive summarization can already be used – e.g. for summarization of text and articles. The aim is to gain a quick overview of what a text might be about or, alternatively, to serve as a basis for a shortened version of a text for a different channel. An example would be journalistic use where a large text needs to be reduced to make it fit into a news podcast format. Feedback of users have pointed out the need to always integrate the “human-in-the-loop" quality principle and to be fully transparent about production processes, especially when content is semiautomatically published.

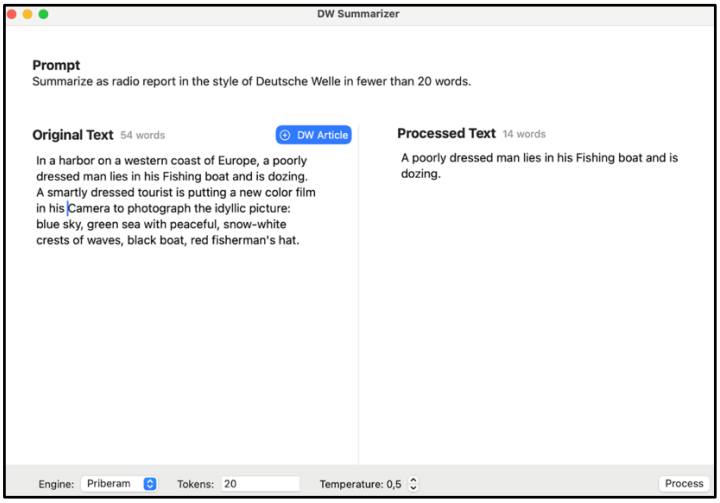

Figure 16 Screenshot of the SELMA Summarizer App

Speaker App

As the summarization app above, the Speaker App is fully developed within SELMA. It integrates 30+ languages using different voice providers and voices (including SELMA voices, e.g. Portuguese for Brazil and Urdu) and generates speech. It serves as a simple tool to check and test different voices and contributes to a higher acceptance of voice over technology. The speaker app is used to demonstrate the status of voices and interact with editorial departments to identify potential new use cases and suitable scenarios for further adoption.

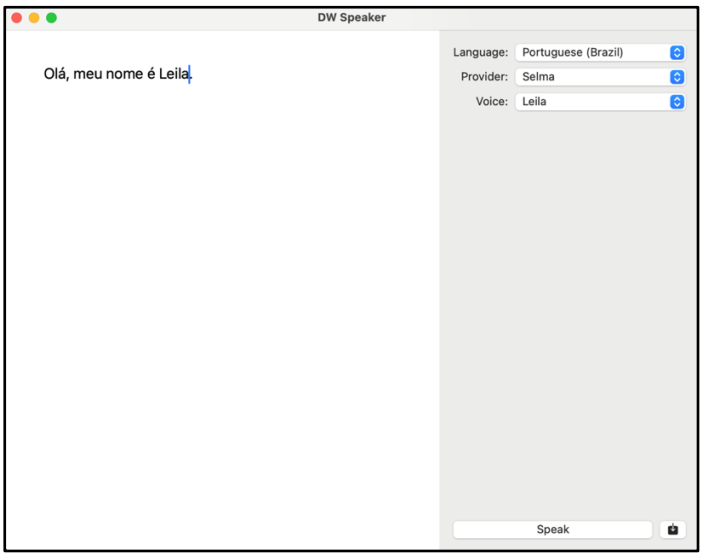

Figure 17 Screenshot of the SELMA Speaker App

DW Avatar

DW Avatar is a recent lab prototype developed at Deutsche Welle. The aim is to see how content in certain languages can be semi-automatically adapted to also be published in more languages. Synthetic speech is being generated and – together with an animated avatar –- is added as an additional layer on top of the video. Seeing the avatar, the viewer immediately gets an idea that the voice is being generated. Also, SELMA voices (generated by LIA) are part of the prototype.

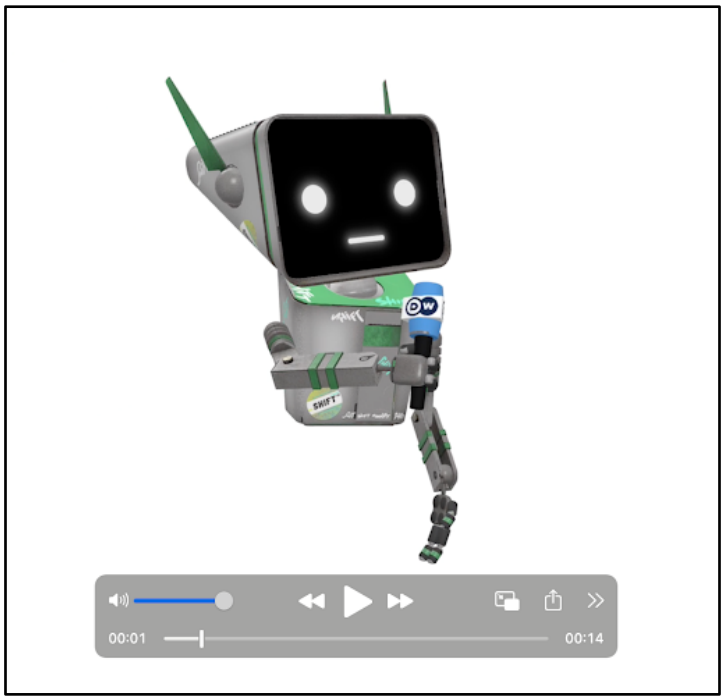

Figure 18 *Screenshot of DW Avatar aimed for adaption of content in many languages*

5.5.3 Fraunhofer

At Fraunhofer, our speech and language technologies are part of a comprehensive service package designed for seamless integration into users' workflows. Our unified media interface showcases this adaptability where it features user-friendly components like multilingual punctuation recovery, speaker diarization, and speech recognition, encapsulated within dockerized containers for modular deployment. This approach ensures our technologies can be incorporated into existing systems, particularly enhancing efficiency in media production by streamlining content transcription and refinement processes across languages.

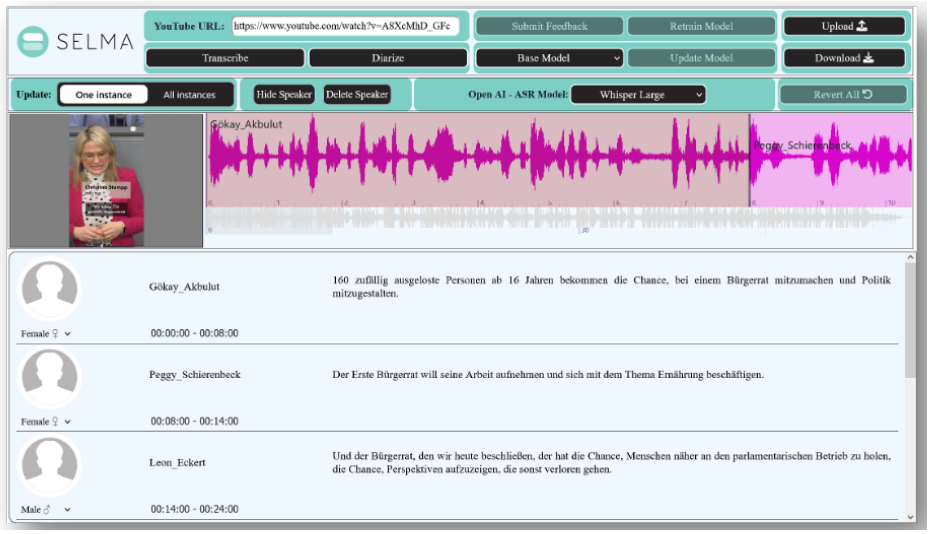

Figure 17 *Integrated Speech Processing Interface for Multilingual Audio Analysis*

Fraunhofer's approach to commercialization centers on understanding and meeting user needs with technology that fits right into their current workflows. Our solutions are not just advanced; they're practical and ready to deploy, ensuring the broadcast sector can quickly benefit from improved efficiency and effectiveness in its operations.

Figure 18 *User-Driven feedback can enhance accuracy and adaptability*

5.5.4 IMCS

Exploitation of SELMA results on IMCS part was primarily focused around Use Case 0 (UC0), conceived in the deliverable D4.2 and elaborated in the deliverables D4.3 (Docker Spaces cloud scaling technology) and D4.4 (SELMA-OSS Open Source Software release). A particularly valuable commercial add-on to the SELMA-OSS release is a high-quality Latvian Text-To-Speech (TTS) module developed within the SELMA project, but omitted from the Open-Source Software release due to ethical, privacy, and commercial reasons of the voice owner (a Latvian actor). Latvian News Agency LETA has expressed interest into using this high-quality TTS voice, but the commercial arrangements are still pending. In the meantime, this high-quality voice has been used for voicing video lectures at University of Latvia (https://github.com/guntisx/DeepLearningCourse). The Named Entity Recognition and Linking (NER-NEL) technologies advanced in the SELMA project (see deliverable D1.4) for the Latvian language have been successfully integrated into the PiniTree.com products and are already in the commercial use by the LETA news agency.

5.5.5 Priberam

Priberam's exploitation of the SELMA outcomes is based on three vectors, one is by disseminating its results in the media, the scientific community and the HLT community to create awareness of its research and development of NLP and big data technologies, thus building a strong image. By making open source parts of its research, Priberam is creating a better engagement with the community and attracting possible investors.

The second vector is the productization of the outcomes of the project. As of the end of the project, plain X and Monitio already have a few clients and a co-development and co-ownership agreement has been established with Deutsche Welle for the case of plain X. A framework contract has also been signed with DW for the usage of the platform.

The third vector is to exploit the outcomes by enlarging and diversifying our line of products with derived products. At the end of the project, we are moving LegiX, our legal research product, to use the Monitio platform and engaging with the pharma sector to create a new Technology Watch platform for the medicine and pharmaceutical industry.

Technology exploitation

The technology, developed by Priberam, in the form of its standalone modules and APIs for content enrichment have been integrated in our existing offering for the Media sector and is now being used by some of the existing clients. Also, the technology for NER/EL has been applied at Hospital de São João to anonymize patient clinical data.

Fraunhofer has been contacted to license the SELMA diarization module and offer its functionality on plain X.

Monitio exploitation

Monitio is now part of our product line, and our commercial efforts are being focused on the Portuguese, Spanish and German markets. A Go-to-Market plan has been defined and is being pursued. Priberam signed copyright agreements with media content organizations in Portugal, Spain and the UK. The website for the product is available at www.monitio.com.

plain X exploitation

plain X exploitation as a product is already under way and the first clients beside DW are already on-board. We already invited 231 possible clients to the trial period. The website has already generated 138 free trials for the product and a total of 566 media items have been uploaded to the platform by trial users, each trial lasts 7 days. From the trial with potential users, four major media companies in Portugal have signed contracts. Two major media international companies are using the trial period. Integrations of the platform with the CMS of some of the clients are being discussed. Since 2023, we've verified a steady growth in number of real active users in the platform, as seen in the graph below.

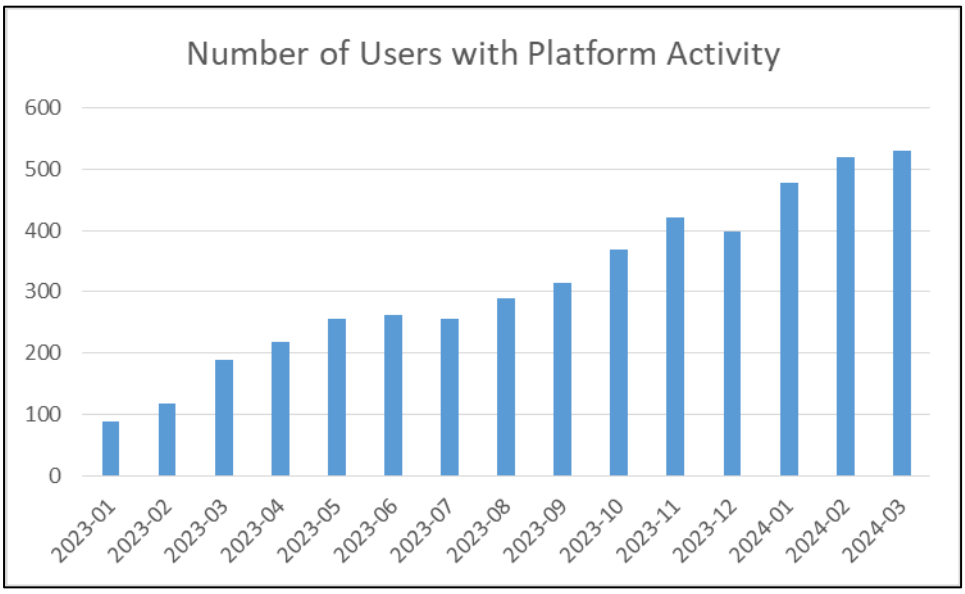

Figure 20 *Number of plain X Users with Platform Activity per month*

During the current year (2024), a self-onboarding will be implemented and made available via the website and the platform at the end of the trial period.

A thorough internal analysis of the competition has been made and a Go-to-Market plan has been established. The following shows a list of some competitors analyzed with different types of offerings. Globally, plain X outstands for the workflow, the versatile integration with different service providers and a more comprehensive offering.

3playmedia.com beey.io captionhub.com checksub.com deepdub.ai descript.com dubverse.ai eztitles.com flixier.com getsubly.com happyscribe.com limecraft.com maestra.ai	papercup.com rev.com sonix.ai subtitlebee.com trint.com veed.io verbit.ai voiceinteraction.ai voizer.net zubtitle.com www.kapwing.com

The commercial team is being expanded and two new hires will start in the next two months, one for the global market and another focused on the Spanish market. The product website is now available at www.plainx.com.

5.6 Technology Impacts and KPI's

SELMA's output includes a platform with improved research and tools for various NLP technologies. The following table sums up the expected technology / component achievements as listed in the DoA.

#	Technology / Component	TRL (2020)	TRL (2024)	Comment
1	Punctuation Recovery	5	7	as planned
2	Speaker Diarization	6	8	as planned
3	Speaker Recognition	6	9****	better than planned
4	Rich Automatic Speech Recognition (including named entities)	6	8*	as planned
5	Text Machine Translation	7	9**	external integration
6	Expressive and Personalized Voice Synthesis	5	7	as planned
7	Speech Machine Translation	5	7	as planned
8	Topic Labeling (from crosslingual transfer)	4	(6), 8***	better than planned
9	Named Entity Recognition and Linking	6	8	as planned
10	Abstractive Summarization	3	8*****	better than planned
11	Integration Platform (NLP Components and UX)	6	8	as planned
12	Integration Platform (Learning/Training of NLP and Automatic Redeployment)	3	7	as planned

	*depending on the target languages and language pairs			
	** external integration			
	*** TRL raised from 6 to 8 in Y2 due to significant progress			
	**** TRL raised from 8 to 9 in Y3/4 due to significant progress			
	***** TRL raised from 6 to 8 in Y3/4 due to significant progress			

Table 13 *Technology / Component Improvements Overview*

The quality of SELMA's output can partly be derived by the KPI's of the components as written in the DoA.

Component	Promise	Status
Platform - Ingestion Scalability	Able to scale up to processing around 10M news articles/segmented video transcripts per day, given the computational resources.	OK - Orchestration
Platform – User Scalability	The system should handle 500 simultaneous users for an installation of the platform.	OK - plain X OK - Monitio OK - SELMA OSS
Platform – Language Coverage	Processing models for Albanian, Amharic, Arabic, Bengali, Bosnian, Bulgarian, Chinese, Croatian, Dari, English, French, German, Greek, Hausa, Hindi, Indonesian, Kiswahili, Latvian, Macedonian, Pashto, Persian, Polish, Portuguese for Africa, Portuguese for Brazil, Romanian, Russian, Serbian, Spanish, Turkish, Ukrainian, Urdu.	OK - Except Dari
Component - Crosslingual Representations and Entity Linking	We will evaluate the representations on downstream tasks that make use of them, such as entity linking and text classification. We seek to outperform scores of the state of the art on standard offline multi-pass methods, improving over 86.6 micro F1 scores	OK

	on end-to-end linking on the CoNLL YAGo dataset (Kolitsas et al., 2018).	
Component - Summarization	We expect gains of at least 5% in ROUGE-1, ROUGE-2 and ROUGE-L F-scores over the state-of-the-art scores and improvements over 5% on factual correctness for abstractive summarization. A common dataset for evaluation is the CNN/DailyMail (Hermann et al., 2015) and New York Times dataset (NYT) (Sandhaus, 2008).	OK
Component - Clustering	We expect improvements of at least 5% in the F1 metric.	OK
Component - Segmentation	We aim at an improvement of 10% relative of the Diarization Error Rate (Bredin 2017) over state-of-the-art segmentation systems such as pyannote or the segmentation provided by Kaldi (Povey 2011), measured on the in-domain data provided by Deutsche Welle.	OK
Component - Speech Machine Translation	The system will be able to offer 2 new language pairs thanks to the use of state-of-the-art end-2-end spoken machine translation approaches, especially for low resource speech translation.	OK
Component - Automatic Transcription	The SELMA ASR system will outperform state-of-the-art results on some high resource language benchmarks. For instance, for French, on the ETAPE benchmark dataset, a reduction of at least 5% of word error rate is expected.	OK

Figure 19 *KPI's per Component*

5.7 IPR Management

For clear IPR Management, we set up an overview table, in which each component is described including licenses.

5.7.1 ASR

LIA 1.1 French

Component Name	LIA-SpeechBrain Speech Recognition Docker Module for French
ID	1.1
WP	3
Status	Core
Lead partner	LIA
Key contact(s)	Yannick Estève, Salima Mdhaffar
Inputs from	Speech audio recordings
Output to	Text
Brief description	A docker image containing an API endpoint to run a neural end-to-end model engined by SpeechBrain and dedicated to inference
What it does (more detail)	Transcribes incoming speech to a sequence of words
How it works (more detail)	The API uses a fine-tuned LeBenchmark wav2vec2.0 model to transcribe speech
Key innovative aspect	LeBenchmark wav2vec2.0 model pretrained by self-supervision on French data
Potential application	Applications using speech recognition
Performance requirements	
License	
Further documentation	

LIA 1.2 Urdu

Component Name	LIA-SpeechBrain Speech Recognition Docker Module for Urdu
ID	1.2
WP	3
Status	Core
Lead partner	LIA
Key contact(s)	Yannick Estève, Salima Mdhaffar
Inputs from	Speech audio recordings
Output to	Text
Brief description	A docker image containing an API endpoint to run a neural end-to-end model engine by SpeechBrain and dedicated to inference
What it does (more detail)	Transcribes incoming speech to a sequence of words
How it works (more detail)	The API uses a fine-tuned wav2vec2.0 model to transcribe speech
Key innovative aspect	SELMA-19 wav2vec model fine-tuned by self-supervision on journalistic data
Potential application	Applications using speech recognition
Performance requirements	
License	
Further documentation	

IMCS 1.3

Component Name	Speech Recognition Docker Module for Latvian
ID	1.3
WP	3
Status	Core
Lead partner	IMCS
Key contact(s)	Arturs Znotins, Roberts Dargis
Inputs from	Transcribed media from LETA, IMCS
Output to	Punctuation prediction, Machine translation
Brief description	A Docker image which runs an API endpoint for automatic speech recognition
What it does (more detail)	Transcribes incoming speech to a sequence of words
How it works (more detail)	The API uses fine-tuned Whisper models to transcribe speech
Key innovative aspect	With previous Kaldi or Wave2Vec approaches Latvian ASR was trained from scratch. With Whisper the base model already supports Latvian, our finetuning achieves WER 12% on Whisper-medium compared to WER 25% on Whisper-large original
Potential application	Applications using speech recognition
Performance requirements	One container can handle one request at the time. Memory, CPU, GPU requirements are the same as for WhisperV2-medium
License	GPL3.0 license
Further documentation	https://hub.docker.com/layers/selmaproject/whisper/medium_lv_v25.14.8400/images/sha256-c4cf00ef517b7c64812c65a2212cf4a453bc19c1bfbd52418047d56ce654f1bd?context=repo

FhG 1.4

Component Name	Automatic Speech Recognition for German
ID	*1.4*
WP	3
Status	Core
Lead partner	FhG
Key contact(s)	Tugtekin Turan
Inputs from	Broadcast audio streams
Output to	Punctuation & casing recovery, speaker diarization Module and speaker identification modules.
Brief description	A Docker container for German Automatic Speech Recognition, utilizing a fine-tuned version of the Whisper large model optimized for media and broadcast content, with rapid inference powered by CTranslate2.
What it does (more detail)	Transcribes German speech with high accuracy, specifically tailored for media and broadcasting. The fine-tuned Whisper model captures the nuances of German as spoken in various media contexts, enhancing transcription quality especially under robust environments.
How it works (more detail)	The module incorporates the Whisper model's robust training on diverse datasets and fine-tune it over German media speech. CTranslate2, a highly efficient inference engine for Transformer models, re-implements the Whisper model to deliver faster transcription times.
Key innovative aspect	The combination of Whisper's state-of-the-art recognition capabilities with the speed and efficiency of CTranslate2 represents a significant advancement in real-time, accurate ASR for German applications.
Potential application	Essential for media use-cases and content creators needing accessibility features such as subtitles, and any application where rapid and accurate German speech transcription is necessary.
Performance requirements	It operates on both CPU and GPU platforms. While fully functional on CPU, the model achieves substantially faster processing speeds on GPU hardware. For instance, the Whisper large model, running on a GPU, can transcribe 15 minutes of audio around 1.5 minute with 3.5 GB max. VRAM, showcasing its capability for efficient large-scale transcriptions.
License	Apache License 2.0
Further documentation	https://hub.docker.com/repository/docker/selmaproject/iais-asr-services

5.8.2 MT

IMCS, Priberam 2.1

Component Name	M2M-100 Machine Translation Docker Module
ID	2.1
WP	3
Status	Core
Lead partner	IMCS, Priberam
Key contact(s)	Arturs Znotins, Didzis Gosko
Inputs from	List of languages used in DW
Output to	NER, NEL, and TTS modules
Brief description	A Docker image which runs an API endpoint for machine translation between 100 languages
What it does (more detail)	Translates text from source language to target language
How it works (more detail)	The API uses Facebook (Meta) M2M-100 multilingual machine translation (MMT) model that can translate between any pair of 100 languages without relying on English data
Key innovative aspect	The M2M-100 model as open-sourced by Meta does not work well out-of-box for all 100 language pairs, as it groups languages in the language families and pivots via largest language in that group. We debugged the language mappings for the 30 languages used by DW
Potential application	Applications using machine translation for text
License	GPL3.0 license
Performance requirements	One container can handle one request at the time. Memory, CPU, GPU requirements are the same as for the original M2M-100 model. Achieves 30 paragraphs/min on GPU and 1 paragraph/min on CPU
Further documentation	https://hub.docker.com/layers/selmaproject/uc0/transl-100/images/sha256-0fda385c1955846d4d4f7c68b9a22994d3054bdc00ae98c154b794b82738b8a0?context=repo

LIA 2.2

Component Name	Textless speech-to-speech translation French-to-English
ID	2.2
WP	3
Status	Core
Lead partner	LIA
Key contact(s)	Jarod Duret, Yannick Estève
Inputs from	SpeechMatrix French-to-English data
Output to	Post-production video editing
Brief description	A Docker image which runs an API endpoint for textless speech-to-speech machine translation: French to English
What it does (more detail)	Translates text from source language to target language
How it works (more detail)	Approach based on the use of discrete speech units
Key innovative aspect	Use of discrete speech units, and baseline for speech-to-speech translation preserving the expressivity
Potential application	Applications using machine translation for speech
License	Open source
Performance requirements	One container can handle one request at the time.
Further documentation	https://hub.docker.com/layers/selmaproject/selma-tts-avignon/s2st/images/sha256-1d3e98b325c7d83274ff097a9fe6649cc0bcc8ba050c5ba24d845dfb88ed21e7?context=explore

5.8.3 Summarization

Priberam 3.1

Component Name	English Monolingual Abstractive Summarization
ID	3.1
WP	2
Status	Core
Lead partner	Priberam
Key contact(s)	amm@priberam.pt , ssm@priberam.pt
Inputs from	Monitio UI, Podcast Creator, other
Output to	Monitio UI, Podcast Creator, other
Brief description	A Docker image which runs an API endpoint for abstractive summarization, supports English.
What it does (more detail)	Generates text corresponding to a summary of a provided input text.
How it works (more detail)	Uses the EBR method proposed to output English abstractive summaries, exposes a REST API with a swagger documentation.
Key innovative aspect	We propose an energy-based model that learns to re-rank summaries according to one or a combination of recently proposed metrics for summarization.
Potential application	Media Monitoring, other
Performance requirements	Needs a GPU for acceptable decoding speed with at least 12GB free.
License	3-Clause BSD License for research only
Further documentation	https://github.com/Priberam/SummEBR https://aclanthology.org/2022.gem-1.1/

Priberam 3.2

Component Name	Crosslingual Abstractive Summarization
ID	3.2
WP	2
Status	Core
Lead partner	Priberam
Key contact(s)	amm@priberam.pt , ssm@priberam.pt
Inputs from	Monitio UI, Podcast Creator, other
Output to	Monitio UI, Podcast Creator, other
Brief description	A Docker image which runs an API endpoint for cross-lingual abstractive summarization, supports Portuguese, English, French, Italian, Spanish
What it does (more detail)	Generates texts corresponding to a summary of a provided input text in all the aforementioned output languages
How it works (more detail)	First, it generates a summary in the same language as the original document. It then generates summaries for each of the remaining languages, taking into account the cross-lingual similarity with the summary generated in the source language to ensure semantic consistency between the summaries generated in all languages
Key innovative aspect	The task of cross-lingual summarization has mostly been addressed in the setting where one wants to generate a summary in a given target language. In our setting, we want to summarize the same document in multiple target languages, and therefore semantic consistency across all generated summaries is a concern
Potential application	Media Monitoring, other
Performance requirements	GPU-8GB
License	3-Clause BSD License for research only
Further documentation	D2.7

Priberam 3.3

Component Name	Crosslingual Multidocument Extractive Summarization (CeRAI)
ID	3.3
WP	2
Status	Core
Lead partner	Priberam
Key contact(s)	amm@priberam.pt , ssm@priberam.pt
Inputs from	Monitio UI, other
Output to	Monitio UI, other
Brief description	A Docker image which runs an API endpoint for cross-lingual extractive multi-document summarization
What it does (more detail)	It creates a summary by extracting sentences from a set of related input documents which can be of several different languages. The output can be in the original languages or if a translation endpoint is provided, all translated to a target language
How it works (more detail)	Provides a REST API with the respective swagger documentation. Given a set of related documents extracts a set of sentences that summarize the given content. Uses an additional service for translation
Key innovative aspect	Uses a multilingual approach to calculate a single dense representation for the summary that guides the selection of the output sentences
Potential application	Media Monitoring, other
Performance requirements	CPU or GPU-3GB depending of the performance required
License	3-Clause BSD License for research only
Further documentation	https://github.com/Priberam/cera-summ https://aclanthology.org/2023.newsum-1.9/

Priberam 3.4

Component Name	Speech Summarization
ID	3.4
WP	2
Status	Core
Lead partner	Priberam
Key contact(s)	amm@priberam.pt , ssm@priberam.pt
Inputs from	Video, Audio
Output to	Text
Brief description	An and-to-end summarization model directly from speech to text
What it does (more detail)	The system takes an audio of a broadcast news in French and generates an abstractive summary of its content on the same language in an end-to-end fashion
How it works (more detail)	A Wav2vec 2.0 model pre-trained on French audio data is used to extract self-supervised features from the audio waveform, which are then fed to an encoder-decoder summarizer that autoregressively generates the summary
Key innovative aspect	Transfer learning with a decoder trained on textual French summarization data and a smart pre-training of the encoder to be able to map self-supervised features from speech to the domain of the decoder
Potential application	Media Monitoring
Performance requirements	Reasonable computing times for a real-word application would require the use of a GPU of at least 8GB
License	3-Clause BSD License for research only
Further documentation	https://link.springer.com/chapter/10.1007/978-3-031-40498-6_27 https://github.com/Priberam/S2TSumm

5.8.4 NER & NEL

IMCS 4.1

Component Name	PiniTree Ontology Editor
ID	4.1
WP	2
Status	Non-core
Lead partner	IMCS
Key contact(s)	Guntis Barzdins
Inputs from	Content from LETA news, and from IMCS Tezaurs.lv
Output to	NER and NEL modules
Brief description	A Docker image running PiniTree database API and GUI backend
What it does (more detail)	A Docker image which runs PiniTree API and GUI backend for semi-automatic human-in-the-loop Named Entity Linking (NEL) and Word Sense Disambiguation (WSD)
How it works (more detail)	The API uses SQLite database to store and query semantic ontological structure of the text documents stored in the database
Key innovative aspect	Semi-automatic human-in-the-loop NEL and WSD annotation methodology and tooling
Potential application	Applications requiring human-in-the-loop verification of automatically generated NEL and WSD annotations
Performance requirements	Linux or MacOS
License	Proprietary, see pinitree.com
Further documentation	https://hub.docker.com/layers/selmaproject/uc0/matrix20/images/sha256-2a5627da43d49b76ec09adcf3f52ef6d4f7688eae439eeb8e37270324358149e?context=repo

Priberam 4.2

Component Name	Multilingual Hierarchical nested NER (HNNER)
ID	4.2
WP	2
Status	Core
Lead partner	Priberam
Key contact(s)	amm@priberam.pt , ssm@priberam.pt
Inputs from	Maestro
Output to	NEL, other
Brief description	A Docker image which runs an API endpoint for multilingual NER
What it does (more detail)	Given an input text document outputs a json with the NER results with the list of mention span and class according to the Named Entity Ontology used in SELMA
How it works (more detail)	The model uses a transition based deep neural network and a xlm-roberta contextual model for NER classification in multiple languages
Key innovative aspect	The model can generalize and zero shot to several unseen languages during training
Potential application	Named Entity Recognition, Media Monitoring, other
Performance requirements	CPU or GPU-4GB
License	Proprietary
Further documentation	D2.7 and D2.8

Priberam 4.3

Component Name	Entity representations for 20M Wikidata entities
ID	4.3
WP	2
Status	Core
Lead partner	Priberam
Key contact(s)	amm@priberam.pt , ssm@priberam.pt
Inputs from	Wikipedia and Wikidada
Output to	NEL, clustering and retrieval
Brief description	A Database with Wikidata IDs and Properties together with a dense vector for each entity independent of the language
What it does (more detail)	The representations for the entities that can be used in several different contexts and plugged on different models
How it works (more detail)	Given a contextual representation of a text using xlm-roberta the database can be queried for the most similar entities
Key innovative aspect	The representations are computed on Wiipedia using dense representations of the contexts where they appear, using a new model based on the Ganea&Hofman algorithm
Potential application	NEL, Clustering, Classification, Retrieval
Performance requirements	10GB of memory
License	Proprietary
Further documentation	D2.7 and D2.8

Priberam 4.4

Component Name	Entity Linking V1, V2
ID	4.4
WP	2
Status	Core
Lead partner	Priberam
Key contact(s)	amm@priberam.pt , ssm@priberam.pt
Inputs from	Maestro
Output to	Maestro
Brief description	A Docker image that plugs into Maestro and detects and disambiguates mentions against Wikidata
What it does (more detail)	From an input document it calls the NER module and link and disambiguates the mentions provided by the NER module. The output is a list of mentions and its Wikidata Ids
How it works (more detail)	The code implements our contextual EL model derived from DCA.(see D2.7)
Key innovative aspect	Multilingual EL, linking to 20M Wikidata entities
Potential application	Entity Linking, Monitio, Health discovery etc.
Performance requirements	16 GB GPU
License	`Proprietary`
Further documentation	D2.7 and D2.8

5.8.5 Postediting & User Feedback

FhG 5.1

Component Name	Automatic Post-Editing
ID	5.1
WP	2
Status	Core
Lead partner	FhG
Key contact(s)	Tugtekin Turan
Inputs from	Speech translation and story segmentation modules
Output to	Directly to end-users for reviewing and approving edits
Brief description	A Dockerized automatic post-editing module that refines ASR/MT outputs by applying contextual correction based on user-specific vocabulary.
What it does (more detail)	Enhances the quality of transcriptions or translations by correcting context-based spelling errors, leveraging a large user vocabulary to align the text closer to user expectations and domain-specific entities
How it works (more detail)	Introduces a retrieval algorithm that maps misspelled n-grams to user phrases, overcoming the limitations of edit-distance methods. It then uses a non-autoregressive model to evaluate the initial transcript alongside retrieved candidates, selecting most contextually appropriate corrections
Key innovative aspect	The unique retrieval algorithm significantly increases the recall of candidate phrases by focusing on n-gram mappings rather than just common letters, paired with a robust BERT-based correction model that processes multiple candidates in one pass for efficient post-editing
Potential appl.	This module is pivotal for any platform that requires high-fidelity text output from speech, such as transcription services, media production and particularly when domain-specific vocabulary is frequently used
Performance req.	The module is designed for optimal performance on GPU. While capable of running on CPU, utilizing a GPU will enhance processing speed and efficiency, allowing for real-time post-editing on large transcripts
License	Apache License 2.0
Further doc.	https://github.com/SELMA-project/Error-Correction

Priberam 5.2

Component Name	Speech2Text PostEditor From User Feedback (M-PHANTOM)
ID	5.2
WP	2
Status	Core
Lead partner	Priberam
Key contact(s)	amm@priberam.pt , ssm@priberam.pt
Inputs from	plain-X, audio other
Output to	plain-X
Brief description	A Docker image which runs an API that given a set of user stored correction and their respective audios(optional) corrects the output of an ASR system.
What it does (more detail)	The model has two possible applications: multilingual keyword detection and ASR postediting. The first works as a standalone model and the second use Whispher prompting to apply the correction of the spotted keywords.
How it works (more detail)	TODO – work is ongoing
Key innovative aspect	A keyword-spotting model that can generalize to languages and keywords that were not seen during training. This model should be later integrated in a system that takes human feedback from ASR misspellings to aid in generating better transcripts.
Potential application	ASR transcription correction and keyword spotting.
Performance requirements	30 GB GPU
License	3-Clause BSD License for research only
Further documentation	https://github.com/SELMA-project/crosslingual-clustering https://ceur-ws.org/Vol-3117/paper2.pdf

5.8.6 Clustering

Priberam 6.1

Component Name	Online Crosslingual Clustering
ID	6.1
WP	2
Status	Core
Lead partner	Priberam
Key contact(s)	amm@priberam.pt , ssm@priberam.pt
Inputs from	News Stream
Output to	Monitio Maestro
Brief description	A Docker image which runs an API endpoint for online crosslingual clustering
What it does (more detail)	Receives a document at the time and generates a cluster ID identifying the story it belongs to independently of the language
How it works (more detail)	Keeps an internal state of the currently detected stories and checks the incoming document with the current state
Key innovative aspect	The underlying model simplifies Multilingual News Clustering through a projection from a shared space, demonstrating that the use of multilingual contextual embeddings as the document representation significantly improves clustering quality
Potential application	Aggregate news documents that follow the same story for media Monitoring
Performance requirements	GPU 3GB
License	3-Clause BSD License for research only
Further documentation	https://github.com/SELMA-project/crosslingual-clustering https://ceur-ws.org/Vol-3117/paper2.pdf

5.8.7 Topic detection

Priberam 7.1

Component Name	Multilingual IPTC Topic Classification
ID	7.1
WP	2
Status	Core
Lead partner	Priberam
Key contact(s)	amm@priberam.pt , ssm@priberam.pt
Inputs from	News Stream
Output to	Monitio Maestro
Brief description	A Docker image which runs an API endpoint for text classification with IPTC subject codes
What it does (more detail)	Given a document ouputs a set of IPTC subject codes organized hierarchichaly (e.g Politics->Goverment) It can also output a set of spans that justify each of the seleted labels
How it works (more detail)	Given a json input with a document it produces a json with a list of topics
Key innovative aspect	The model extends the AttentionXML model for the multingual scenario leveraging on contextual multilingual embeddings, it also proposes a new method for explaining its decisions
Potential application	Topic classification for Media Monitoring
Performance requirements	CPU or GPU-3GB
License	Proprietary
Further documentation	http://pbagit.interno.priberam.pt:3000/jtf/NewsIPTC

5.8.8. Speech Synthesis

IMCS, LIA 8.1

Component Name	Text To Speech for Latvian
ID	8.1
WP	3
Status	Core
Lead partner	IMCS, LIA
Key contact(s)	Roberts Dargis, Didzis Gosko
Inputs from	Transcribed Latvian corpora from LETA, IMCS, Neredzigo Biedriba
Output to	Post-production video editing
Brief description	A Docker image which runs an API endpoint for Latvian TTS
What it does (more detail)	Provided with the text in Latvian, API produces a wav file with corresponding audio
How it works (more detail)	Model is trained using VITS: Conditional Variational Autoencoder with Adversarial Learning for End-to-End Text-to-Speech
Key innovative aspect	Latvian phonetic alphabet is used for the actual speech synthesis. A list of rules for conversion between the Latvian standard text and phonetic alphabet have been developed and finetuned
Potential application	High quality Latvian text synthesis
Performance requirements	One container can handle one request at the time. Memory, CPU, GPU requirements are the same as for the original VITS model
License	Proprietary, IMCS
Further documentation	https://hub.docker.com/layers/selmaproject/tts/lv-selma3/images/sha256-30a532672094c28c87ffaaa5be9b455378cc90568dba6af6f10deaf6010e2ce7?context=repo

LIA, IMCS 8.2

Component Name	Text To Speech for Brazilian
ID	8.2
WP	3
Status	Core
Lead partner	LIA, IMCS
Key contact(s)	Jarod Duret, Yannick Estève
Inputs from	Transcribed Brazilian corpus from Deutsche Welle
Output to	Post-production video editing
Brief description	A Docker image which runs an API endpoint for Brazilian TTS
What it does (more detail)	Provided with the text in Brazilian Portuguese, API produces a wav file with corresponding audio
How it works (more detail)	Tacotron 2 + WaveRNN as a vocoder
Key innovative aspect	The use of DW data
Potential application	High quality Brazilian Portuguese text synthesis
Performance requirements	
License	Open source, LIA
Further documentation	https://hub.docker.com/layers/selmaproject/selma-tts-avignon/pt_br-v2/images/sha256-d1b0d09f6ddcc627b44cdf5672dd8854049e7329a6e51cb3737472c991913ea1?context=explore

LIA, IMCS 8.3

Component Name	Text To Speech for Urdu
ID	8.3
WP	3
Status	Core
Lead partner	LIA, IMCS
Key contact(s)	Jarod Duret, Yannick Estève
Inputs from	Transcribed Urdu corpus from Deutsche Welle
Output to	Post-production video editing
Brief description	A Docker image which runs an API endpoint for Brazilian TTS
What it does (more detail)	Provided with the text in Urdu, API produces a wav file with corresponding audio
How it works (more detail)	Tacotron 2 + WaveRNN as a vocoder
Key innovative aspect	The use of DW data
Potential application	High quality Urdu text synthesis
Performance requirements	
License	Open source, LIA
Further documentation	https://hub.docker.com/layers/selmaproject/selma-tts-avignon/dw-tts-urdu/images/sha256-a8311550da36d4ac32c5082275591f6117477db409a8198a0ae2bc61f0798eba?context=explore

5.8.9 Story Segmentation

FhG 9.1

Component Name	Story Segmentation
ID	9.1
WP	2
Status	Core
Lead partner	FhG
Key contact(s)	Tugtekin Turan
Inputs from	News clustering and punctuation & casing recovery modules
Output to	Outputs to search & visualization for segment-based text exploration, and indexation if the segments are indexed for search purposes.
Brief description	A Docker module for semantic story segmentation in documents, using LSTM-based neural networks to structure text by topic changes.
What it does (more detail)	Divides large text documents into semantically coherent segments, facilitating better understanding and management of content, especially in long-form multilingual news articles or transcripts.
How it works (more detail)	Employs a hierarchical network composed of two layers of bidirectional LSTM sub-networks. The lower-level network generates representations for each sentence by processing words and applying max-pooling over the LSTM outputs. The higher-level network then evaluates these sentence representations to determine segment boundaries as a supervised learning task, with binary classification for segment endings.
Key innovative aspect	The utilization of a hierarchical architecture for text segmentation presents a novel approach to understanding document structure, enabling more accurate topic segmentation based on linguistic patterns within the text.
Potential application	Ideal for news aggregators, content management systems and any platform that requires the automated structuring of text for enhanced navigation and comprehension
Performance requirements	The module is fully functional on CPU with the capability to segment a standard-length news article efficiently. Significantly better performance and speed can be achieved when utilizing GPU hardware
License	Apache License 2.0
Further documentation	https://hub.docker.com/repository/docker/selmaproject/text-segmentation

5.8.10 Punctuation & Truecasing

FhG 10.1

Component Name	Punctuation and Casing Recovery
ID	*10.1*
WP	3
Status	Core
Lead partner	FhG
Key contact(s)	Tugtekin Turan
Inputs from	Speech recognition/translation and story segmentation modules.
Output to	Post-editing, speaker diarization and speech synthesis modules.
Brief description	A dual-functionality Docker module that utilizes a pre-trained BERT model to add punctuation and correct casing in a given input
What it does (detail)	Inserts punctuation and corrects casing for transcribed texts across multiple languages, enabling better comprehension and display of text
How it works (more detail)	It utilizes a BERT model with dual token classification heads, one for punctuation and another for capitalization. Each head processes the encoded representation from the BERT [CLS] token to perform its task. The model is uniquely fine-tuned on a joint task for multiple languages, including Amharic, Bengali, German, English, Spanish, French, Hindi, Italian, Latvian, Pashto, Portuguese, Russian, and Tamil
Key innovative asp.	The model's multilingual design allows it to apply accurate punctuation and capitalization to a variety of languages within a single framework, maximizing the utility of the module across diverse linguistic datasets
Potential app.	Can be employed in publishing tools to enhance text clarity in numerous languages. Ideal for any application requiring high-quality text outputs
Performance req.	Process at least 100 tokens per second per language, maintaining an accuracy rate above 75% on a standard CPU setup
License	Apache License 2.0
Further doc.	https://hub.docker.com/repository/docker/selmaproject/punctuation-casing

5.8.11 Speaker Diarization

FhG 11.1

Component Name	Speaker Diarization
ID	*11.1*
WP	2
Status	Core
Lead partner	FhG
Key contact(s)	Tugtekin Turan
Inputs from	Automatic speech recognition module.
Output to	Story segmentation, post-editing and NER, NEL, discovery modules.
Brief description	A proprietary Docker module for speaker diarization that identifies and differentiates speakers within an audio stream.
What it does (more detail)	Analyzes audio to segment and label each part of the stream by the speaker, enabling the identification of who is speaking at any given time.
How it works (more detail)	Utilizes a hidden Markov model alongside a sequence of x-vectors extracted from the audio to cluster speaker segments. It enables the model to effectively differentiate speakers in complex audio scenarios.
Key innovative aspect	This approach represents a significant advancement in speaker diarization accuracy, particularly in scenarios with overlapping speech and a large number of speakers.
Potential application	Essential for transcription services, news analysis, and any application requiring the separation of speakers in multi-party conversations, such as journalistic interviews, and media interviews.
Performance req.	The module is optimized for CPU usage, capable of diarizing 30 minutes of audio within 2 minutes when running on a decent hardware.
License	Proprietary
Further documentation	https://hub.docker.com/r/selmaproject/diarization

5.8.12 Speaker Recognition

FhG 12.1

Component Name	Speaker Recognition (Identification)
ID	*12.1*
WP	2
Status	Core
Lead partner	FhG
Key contact(s)	Tugtekin Turan
Inputs from	Automatic speech reconginiton and speaker diarization modules
Output to	Can output to speech synthesis module for new model trainings
Brief description	A Docker module for text-independent speaker identification that generates speaker embeddings to recognize and differentiate speakers based on their unique vocal attributes
What it does (more detail)	The system identifies individuals by analyzing the characteristics of speech, irrespective of the spoken content. It operates on variable-length speech utterances, converting them into fixed-length vectors, known as speaker embeddings, which capture the speaker's unique vocal traits
How it works (more detail)	The module processes speech to produce speaker embeddings that serve as a compact representation of a speaker's identity. It uses traditional machine learning models trained on a set of known speakers to classify speaker identities accurately. The system also allows for user feedback, where users can update the speaker set by adding or removing speakers with just a short duration of audio sample from each speaker
Key innov. asp.	The ability to integrate user feedback dynamically is a key innovation, allowing the system to adapt to new speakers quickly and efficiently. The use of fixed-size embeddings ensures consistent performance regardless of the length of the input audio
Potential appl.	This module can be utilized in personalized user interfaces, media monitoring systems requiring speaker identification, and other applications where distinguishing between known speakers is crucial
Performance req.	The system is optimized to process speaker embeddings quickly and accurately, with the capacity to enroll a new speaker with minimal audio input and adapt to user feedback in a short notice
License	Apache License 2.0
Further doc.	https://github.com/SELMA-project/Speaker-Recognition

5.8.13 Graph Orchestrator platform (Maestro)

Priberam 13.1

Component Name	Graph Orchestrator (Maestro)
ID	13.1
WP	4
Status	Core
Lead partner	Priberam
Key contact(s)	amm@priberam.pt , ssm@priberam.pt
Inputs from	scrapper
Output to	Monitio search, Plain X DB
Brief description	Orchestrates the NLP pipeline according to a job graph
What it does (more detail)	Maestro Orchestrator is a task-agnostic DAG (directed acyclic graph) workflow execution engine designed for running NLP pipelines and handling job queues.
How it works (more detail)	Given a graph with the jobs as nodes and the dependencies as directed edges, Maestro Orchestrator maximizes the number of jobs running in parallel, while ensuring each job has the necessary input.
Key innovative aspect	
Potential application	Any NLP pipeline that can be configured as a DAG
Performance requirements	2 CPU cores and a Postgres database
License	3-Clause BSD License for research only
Further documentation	D4.1

5.8.14 Monitio platform - SELMA extensions (UC1)

Priberam 14.1

Component Name	Use Case 1 – SELMA Monitio Platform - SELMA extensions
ID	14.1
WP	4
Status	Core
Lead partner	Priberam
Key contact(s)	amm@priberam.pt , ssm@priberam.pt
Inputs from	The Web
Output to	User
Brief description	The Monitio platform, a media monitoring tool powered by AI
What it does (more detail)	Ingests news documents from the Web and provides a set of tools and visualizations for the task of media monitoring
How it works (more detail)	see deliverables
Key innovative aspect	Probably the first multilingual media-monitoring platform available with integrating the AI analysis as developed in SELMA
Potential application	Media Monitoring Technology Watch, Horizon Scanning, Legal Research
Performance requirements	
License	Proprietary
Further documentation	www.monitio.com

5.8.15 plain X platform (UC2)

Priberam 15.1

Component Name	Use Case 2 – plain X Platform
ID	15.1
WP	4
Status	Core
Lead partner	Priberam
Key contact(s)	amm@priberam.pt , ssm@priberam.pt
Inputs from	The user
Output to	User
Brief description	A content production tool for A/V localization
What it does (more detail)	The software integrates four steps of the localization process: Transcription, translation, subtitling and (synthetic) voice-over. The software uses AI technology to enable faster content adaptation by saving many manual steps
How it works (more detail)	see deliverables
Key innovative aspect	The workflow, the ability to integrate with new engines, the orchestration, the user feedback, etc.
Potential application	Media localization
Performance requirements	see documentation
License	Proprietary
Further documentation	D4.6 , www.plainx.com

5.8.16 SELMA Open-Source platform (UC0)

IMCS 16.1

Component Name	Use Case 0 – SELMA Open-Source Platform
ID	16.1
WP	4
Status	Core
Lead partner	IMCS
Key contact(s)	Guntis Barzdins, Didzis Gosko
Inputs from	NLP Componets from all partners
Output to	GUI for yesying
Brief description	Open-Source Platform for testing NLP compnents developed within SELMA
What it does (more detail)	Use Case 0 GUI has multiple versions developed for testing various combinations of the NLP components developed within SELMA
How it works (more detail)	Docker Spaces scheduling system mediates queueing of jobs from the GUI towards the NLP components running as Docker containers in the backend. Also handles format conversion between the GUI and NLP components
Key innovative aspect	Docker Spaces were developed within the SELMA project. Docker Spaces allow to handle all format conversion between the GUI and NLP components directly in frontend completely eliminating the need for any specialised state-full backend. This makes Docker Spaces architecture highly scalable to 10M requests/day with additional state-less load-sharing module running on ARM M2 CPU architecture
Potential application	Testing any NLP component. Simple NLP pipeline for ASR, translation, TTS
Performance req.	Docker backend environment suitable for NLP Docker containers to be tested
License	GPL3.0 license
Further documentation	https://hub.docker.com/layers/selmaproject/docker-spaces/ligo-next4/images/sha256-9a8bb012bbf80b34662536c765ef0ca964f3b3a02b7c57ad80393b152b40fc7a?context=repo

6. Conclusion & Outlook

SELMA has created big footprints in terms of advancing the state of the art of Natural Language Processing - both in technical development as in applying results into various platforms and prototypes. In many publications - 31 in total (planned: 11) - and numerous events - 66 in total (planned 55) - we demonstrated NLP findings and outcomes to a diverse audience ranging from Sweden to Avignon and Riga to Amsterdam. Also, SELMA partners have been present at events beyond Europe, e.g. at Interspeech in Korea and ICASSP in Singapore.

We organized 4 User Day and User Group events with more than 150 participants and gathered valuable feedback which went into the further development of the project. Through our participation in many EU / BDVA Events both physically as well as virtually we could significantly extend our networks. Our submissions to the Innovation Radar showed that the project is also getting attention at EU level: All our 3 submissions were selected to be showcased on the Innovation Radar websites.

A major focus during the final phase of the project was set on how to exploit the SELMA outcome with special regard to its components. Together with the three platforms, we worked on publishing 28 components, which are either available via an open-source license (16) or proprietary (12). Many of the components can be tested in the freely available SELMA open-source platform under: https://selma.ailab.lv.

The project was also about developing a scalable system to process extremely large amount of data (10 million news items / day). That goal also was successfully reached.

SELMA has impressively shown that many NLP technologies are ready to be picked up by (media) industries.