



Co-funded by the
European Union



OpenEuroLLM: Open European Family
of Large Language Models

Communication, Dissemination and Exploitation Strategy

Deliverable number: 6.1

Version 1.0



This project has received funding from the Digital Europe Programme under
grant agreement No 101195233.

Project details

Project Acronym:	OpenEuroLLM
Project Full Title:	OpenEuroLLM: Open European Family of Large Language Models
Year of the Call:	2024
Type of Action:	Digital SME Support Actions
Grant Number:	101195233
Project URL:	https://openeurollm.eu

Report details

Deliverable title	
Lead author:	Gema Ramírez Sánchez (Prompsit)
Internal reviewers:	Tea Vojtěchová (CU), Erika Halonen (AMD Silo), Lev Berezhnoy (Prompsit)
Deliverable number:	6.1
Dissemination level:	Sensitive (SEN)
Contractual Delivery Date:	July 31, 2025
Actual Delivery Date:	July 31, 2025
Number of pages:	14

Document history

Version	Date	Changes
1.0	July 31, 2025	Original Submission

Abstract

This report provides a description of the strategy for communication, dissemination and exploitation activities in the OpenEuroLLM (OELLM) project. It is divided into several sections. This strategy was defined during the first 5 months of the project and will be slightly modified as the project progresses to accommodate to its development.

Contents

1	Executive summary	2
1.1	Introduction	2
1.2	Brief summary of the OpenEuroLLM project	2
2	Communication, dissemination and exploitation strategy for OpenEuroLLM	3
2.1	Objectives	3
2.2	Expected project results	3
2.2.1	Data	3
2.2.2	Software, tools and pipelines	4
2.2.3	Models	4
2.3	Target audience	4
2.4	Promotion materials	5
2.4.1	Visual Identity	5
2.4.2	Website design and contents	6
2.4.3	Conference papers, public abstracts, and publications	7
2.4.4	Social media, blogging and press content	7
2.5	Activities	8
2.5.1	Communication activities	8
2.5.2	Dissemination activities	9
2.5.3	Exploitation activities	10
2.6	Impact	10
2.7	Scheduling	10
3	Conclusions	12

1 Executive summary

1.1 Introduction

This deliverable, named: *Communication, Dissemination and Exploitation Strategy*, is linked to the *WP6: Communication*. The primary objective of this task is to outline an initial strategy for activities related to communicating, disseminating, and exploiting the project's results.

1.2 Brief summary of the OpenEuroLLM project

OpenEuroLLM, a consortium of 20 European research institutions, companies and EuroHPC centres, is building a family of performant, open-source, multilingual, large language foundation models for commercial, industrial and public services. The goal is to lower thresholds for European AI product development and refinement, increasing Europe's competitiveness and digital sovereignty. The project also demonstrates the strength of transparency, openness and community involvement; an Open Strategic Partnership Board has been established to ensure that the models, software, data and evaluation are fully open and can be adjusted to meet specific industry and public sector needs.

The models will comply with Europe's regulatory framework, ensuring alignment with European values while maintaining technological excellence.

OpenEuroLLM leverages support from previous European projects and the experience of the partners and their results, including large repositories of high-quality data and pilot LLMs developed previously. OpenEuroLLM has been awarded the STEP (Strategic Technologies for Europe Platform) seal.



All partners are involved in communication and dissemination activities; therefore, all contribute to the objectives of this deliverable.

2 Communication, dissemination and exploitation strategy for OpenEuroLLM

The OpenEuroLLM project aims to develop *transparent, multilingual, and open-source large language models (LLMs)* that align with European values and regulatory frameworks (e.g., the AI Act) and that serve the needs of European research and industrial stakeholders in developing strong and secure AI solutions.

This deliverable sets the strategic plan for the dissemination, communication and exploitation activities to support the rest of the activities in the project, i.e. management, community building and technical activities.

Our strategic plan has the following key objectives:

2.1 Objectives

- Raise Awareness – Promote OpenEuroLLM’s mission to develop AI models for official EU languages extended to other interesting languages for Europe, ensuring visibility among stakeholders;
- Engage Stakeholders – Boost collaboration with academia, industry, policymakers, and open-source communities to encourage adoption;
- Maximise Impact – Ensure project outputs (data, models, evaluation results and pipelines) are widely accessible and used in commercial, public, and research sectors;
- Ensure Compliance & Transparency – Communicate adherence to EU regulations and ethical AI principles;
- Build a Sustainable Community – Establish long-term engagement through open-source contributions, partnerships, and knowledge-sharing.

2.2 Expected project results

The project will produce various items: openly released datasets, a series of foundation models along with intermediate checkpoints, pipelines, tools, and evaluation results. Release formats and distribution platforms will be carefully selected to maximise accessibility, usability, and adoption across diverse stakeholder groups. It will also produce a stakeholder report by the end of the project, gathering strategic advice from the Open Strategic Partnership Board and community feedback.

A more detailed view of project results by type is provided in the following sections.

2.2.1 Data

Dataset releases produced by OpenEuroLLM include three types:

1. A structured catalogue of LLM pre-training datasets mainly for internal use of the project, with a view to becoming a community-supported resource.

2. Curated, enriched and EU regulatory-compliant multilingual datasets in suitable formats along with metadata mainly intended for pre-training large language models.
3. Multilingual benchmarks for model performance evaluation across languages and cultural contexts.

Datasets will be catalogued and made available on multiple EuroHPC systems as world-readable assets, avoiding duplicates. They may also be made available on other platforms such as HuggingFace¹ or the European Language Data Space (LDS).² Benchmarks and evaluation results will also be made available through visible platforms to ensure transparency.

2.2.2 Software, tools and pipelines

Code releases produced by OpenEuroLLM will include:

1. Dataset processing and enrichment tools and pipelines to facilitate continuous data improvement;
2. Open-source LLM training, inference and evaluation frameworks for reproducibility and community contributions;
3. Documentation for developers and users.

Both code and documentation will be released through code repositories such as GitHub.³

2.2.3 Models

Model releases will include:

1. Reference models for continuous evaluation and ablation studies;
2. Pre-trained models along with intermediate checkpoints;
3. Post-trained models to test the full usability of the OpenEuroLLM pre-trained models;
4. Documentation for developers and users.

Both models and documentation will be released on convenient platforms such as HuggingFace or the European Language Grid.⁴

2.3 Target audience

The consortium communication and dissemination activities, closely aligned with the groups targeted by the Community Building work package, will address the following groups:

- Researchers & Academia: for model development, benchmarking, and open science collaboration.
- Industry & Startups: to integrate OpenEuroLLM models into commercial AI products.
- Public Sector & Policymakers: to align with EU digital sovereignty goals and regulatory standards.

¹<https://huggingface.co/models>

²https://language-data-space.ec.europa.eu/index_en

³<https://github.com/>

⁴<https://live.european-language-grid.eu/>

- Open-Source Communities: for contributions, testing, and long-term sustainability.
- Media & General Public: to raise awareness about Europe’s AI capabilities and responsible AI development.
- Consortia of other related European projects: to cooperate with and strengthen relationships towards future joint collaboration.

2.4 Promotion materials

2.4.1 Visual Identity

The OpenEuroLLM initiative will have its own visual identity with the aim of making the project results easily recognisable and the communication and dissemination activities consistent. Visual elements include:

- A distinctive logo (figure 2.1), colour scheme, and typography to be used in website, deliverables, presentation slides, etc.
- OpenEuroLLM custom templates for presentations, reports, deliverables, slides and other formats for internal and external communication and dissemination activities.
- Templates for X, Bluesky and LinkedIn accounts and posts.
- A general action website design, which will be updated regularly during the project lifetime, including easy access to deliverables, results and news.
- Other supporting materials like project flyers, rollers, stickers, etc., to be used as communication materials in online and physical events.



Figure 2.1: The OpenEuroLLM logo strongly highlights the open nature of the results that this project aims at building (bold OPEN), its European roots (Euro) and its taste for diversity (multi-coloured LLM).



Figure 2.2: Funding acknowledgement to appear in all action outputs.



Figure 2.3: Disclaimer to appear in all action outputs.

Acknowledgement to the Digital Europe Programme will be included by all partners in all external communication and dissemination materials by using the EU visual identity and accompanying text shown in Figure 2.2. In addition, we will add the disclaimer text along with the EU flag appearing in Figure: 2.3.

2.4.2 Website design and contents

The OpenEuroLLM website (see figure 2.4),⁵ launched at the project's start, provides general information about the project's aims and objectives, partners, contact details, and other relevant project-related information. It will evolve as the project progresses, offering updates, blog posts, and technical documentation. It will also accommodate information for downloadable resources such as:

- Links to project public deliverables available for download.
- Release-related information.
- Links to the data repositories where the datasets can be downloaded.
- Links to code repositories used to reproduce models and data.
- Links to papers, publications and news produced from the project.

⁵<https://openeurollm.eu>

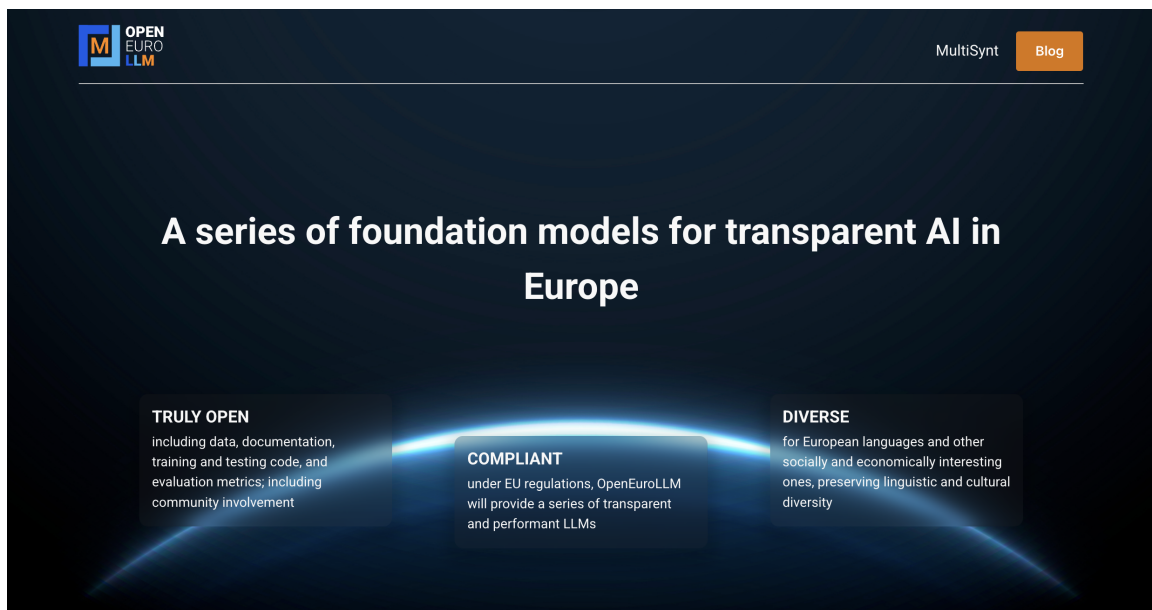


Figure 2.4: OpenEuroLLM website already accounting for 30,000 visits as of July 2025.

2.4.3 Conference papers, public abstracts, and publications

All partners are encouraged to submit their work and demonstrate their progress in various forms, including articles, public project abstracts, conference papers, reports, and any type of promotional tool to increase awareness of the project. Technical and non-technical audiences may be the target of these types of communication activities.

Publishing requires a 30-day advance notice to project partners, which will apply best practices to avoid any Intellectual Property Rights (IPRs) conflicts or confidentiality issues.

2.4.4 Social media, blogging and press content

OpenEuroLLM will maintain an active online presence distributed in several communication channels for progress sharing and real-time updates:

- Social media: communication will be conducted through project-specific accounts on X, Bluesky, and LinkedIn. Partners will be asked to repost and interact to maximise reach through their professional profiles, individually or as institutions.
- Blogging content: content accessible to all audiences regarding the OpenEuroLLM project will be posted in various channels and mirrored on the website blog.
- Press releases: Press releases will be published regularly for each major milestone or release.

As of July, a blog has already been implemented on the OpenEuroLLM website to accommodate the first posts mirrored on different social media posts as seen in figure 2.5. The project has already designated two spokespeople: Jan Hajič from Charles University and Peter Sarlin from AMD Silo, to give interviews to media when appropriate.

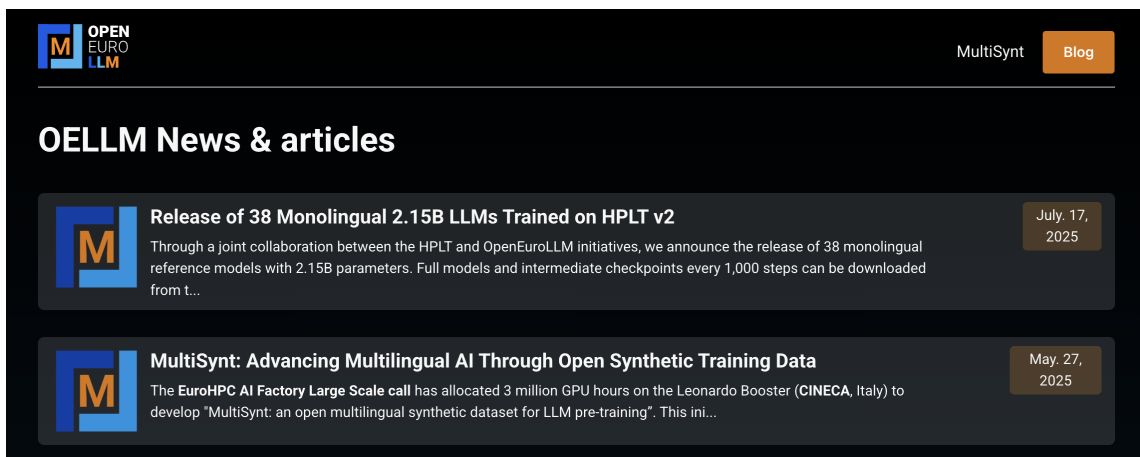


Figure 2.5: OpenEuroLLM blog.

2.5 Activities

The main communication, dissemination, and exploitation activities planned for the project are described in the following sections. Further activities may be included as the project evolves, taking into account the feedback from the Community Building work package, particularly through the Open Strategic Partnership Board.

2.5.1 Communication activities

We identify the following main activities related to project communication:

- Maintaining the project website: As the project advances, new sections will be added to the current website, including deliverables, project releases, blog posts, etc.
- Posting about the project on social media: We will maintain regular communication through X, Bluesky, LinkedIn, the website, and partners' personal or institutional accounts.
- Creating press releases: We will produce press releases related to project outputs to reach a wider audience.
- Linking with other EU projects and NLP communities: We will establish contact with targeted communities and organise project presentations for general purposes or to showcase results.
- Publishing: As a communication activity, we will send project descriptions and abstracts to relevant project tracks in conferences, workshops, and events.
- Collaborating in events: As a project, we will collaborate and provide our support in events, workshops, shared tasks, and conferences.

Other communication activities will be supported as part of this plan, for example:

- Internal hackathons: we plan to organise several hackathons to share knowledge among partners and coordinate efforts around thematic areas (dataset processing, anonymisation, evaluation, etc.).

- Annual NLPL Winter School:⁶ following past successful experiences, we plan to support the organisation of the annual gathering of the NLPL group around OpenEuroLLM-centered topics. This will ensure internal partner alignment and foster collaboration and transparency with external participants.
- Reaching out to similar and complementary projects: we will discuss cooperation opportunities with other active projects such as CommonCrawl,⁷ OpenWebSearch⁸ or HuggingFace.⁹

As of July 2025, we have focused our efforts in communication activities to raise awareness of the project overall goals, to link with potential collaborators that will help us achieve those goals and to share transparently the first project findings from technical development.

2.5.2 Dissemination activities

Our strategic plan identifies the following dissemination activities as important:

- Presenting results in conferences: OpenEuroLLM partners will engage at conferences and workshops related to the project. We anticipate participation in many conferences or workshops related to data curation and language modeling. Partners will actively monitor conference and event schedules to engage with the most relevant events during the action's time frame.
- Presenting results in industry events: Partners will also engage in a number of industry events.
- Organising data sharing in EU clusters.
- Releasing a dataset catalogue for LLM pre- and post- training: it will be disseminated as a GitHub (or similar) repository to centralise collaborative work.
- Releasing software for public use: code for software and pipelines will be uploaded to GitHub (or similar) following the best practices by the community regarding packaging of software. Both developer and user documentation will be produced along with the code.
- Releasing models for public use: Models will also be uploaded and disseminated through standard channels such as HuggingFace, in addition to the project website.
- Organising shared tasks on test sets: We will actively participate in activities that allow us to benchmark project progress and learn about other people's approaches.
- Organising hands-on events with HPC + NLP people to share our HOWTOs: We plan to organise hands-on hackathons and seminars to disseminate and communicate project results and best practices, to advance the current state of NLP components in our pipelines, explore HPC advances, limits and possibilities, and keep ourselves informed about progress in the project field.

As of July 2025, we have started the first submissions to conferences for which we hope to have positive feedback and contribute transparency in AI development.

⁶<http://wiki.nlpl.eu/Community/training>

⁷<https://commoncrawl.org/>

⁸<https://openwebsearch.eu/owler/>

⁹<https://huggingface.co/>

2.5.3 Exploitation activities

OpenEuroLLM results should enable easy adoption of the models in practical settings (i.e., for instruction-tuning or fine-tuning efforts, for plug-and-play access to ready models, etc.). Thus, a big effort will be devoted to packaging and distributing through multiple and convenient channels.

This implies the following key exploitation activities:

- Distributing project results in online data and model hubs;
- Distributing models and recipes on how to use or adapt them;
- Distributing data and models through other platforms where NLP communities gather, such as the EU HPC centers, ready for use and inference;
- Integrating project results within external platforms such as DeployAI.¹⁰

2.6 Impact

The goal of this project is to provide open, transparent and multilingual LLMs to be further exploited by research, public and private sectors integrating the European perspective on AI. Also, to provide the tools and pipelines developed to produce these LLMs and encourage their usage, reproducibility and extension.

To measure the impact of the project results, we define as key performance indicators the following measurable items:

- Website & social media metrics (visits, engagement, follower growth);
- Dataset & model downloads (adoption rate);
- Media coverage & citations (impact on public discourse);
- Collaborations with other EU complementary initiatives (partnerships);
- Industry partnerships formed (commercial exploitation);
- Regulatory influence (references in AI policy discussions).

2.7 Scheduling

We propose a time schedule that takes into account the most important project results timeline. General communication activities will occur throughout the project life, while publishing and other dissemination activities will intensify in the second half. Special promotion activities are expected to follow after major milestones are accomplished.

¹⁰<https://www.deployaiproject.eu/>

Major Project Deliverables/Milestones Schedule

No.	Outcome	Month 1... .. Month 18	Month 19... .. Month 36
1	<i>Initial catalogue and stats</i>	□□□□■□□□□□□□□□□□□□□	□□□□□□□□□□□□□□□□□□□□
2	<i>Network Operability</i>	□□□□□□□□□□□□■□□□□□□□	□□□□□□□□□□□□□□□□□□□□
3	<i>Initial dataset release</i>	□□□□□□□□□□□□□□□□□□□■	□□□□□□□□□□□□□□□□□□□□
4	<i>First models</i>	□□□□□□□□□□□□□□□□□□□■	□□□□□□□□□□□□□□□□□□□□
5	<i>Multilingual benchmarking</i>	□□□□□□□□□□□□□□□□□□□■	□□□□□□□□□□□□□□□□□□□□
6	<i>Evaluation code package</i>	□□□□□□□□□□□□□□□□□□□■	□□□□□□□□□□□□□□□□□□□□
7	<i>Final dataset release</i>	□□□□□□□□□□□□□□□□□□□□	□□□□□□□□□□■□□□□□□□□
8	<i>Final models</i>	□□□□□□□□□□□□□□□□□□□□	□□□□□□□□□□□□□□□□□□■
9	<i>LLM training report</i>	□□□□□□□□□□□□□□□□□□□□	□□□□□□□□□□□□□□□□□□■
10	<i>Evaluation Report</i>	□□□□□□□□□□□□□□□□□□□□	□□□□□□□□□□□□□□□□□□■
11	<i>Stakeholder Report</i>	□□□□□□□□□□□□□□□□□□□□	□□□□□□□□□□□□□□□□□□■

Communication/Dissemination/Exploitation Schedule

No.	Activity	Month 1 Month 36
1	Visual Identity	■□□□□□□□□□□□□□□□□□□□ ■□□□□□□□□□□□□□□□□□□□□
2	Website Launch & Updates	■□□□□□□□□□□□□□□□□□□□ ■□□□□□□□□□□□□□□□□□□□□
3	Posting on blogs and social media	■□□□□□□□□□□□□□□□□□□□ ■□□□□□□□□□□□□□□□□□□□□
4	Promotion of outcomes	□□□□□□■□□□□□■□□□□□□□□ ■□□□□□□□□□□□□□□□□□□□■
5	Publishing & Shared tasks	□□□□□□□□□□□■□□□□□□□□□□ ■□□□□□□□□□□□□□□□□□□□□
6	Conferences, Events, Hackathons	□□■□□□□□□□□□□□□□□□□□□ ■□□□□□□□□□□□□□□□□□□□□
7	Reach out to stakeholders	□□■□□□□□□□□□□□□□□□□□□ ■□□□□□□□□□□□□□□□□□□□□
8	Integration of results in platforms	□□□□□■□□□□□□□□□□□□□□□ ■□□□□□□□□□□□□□□□□□□□■

3 Conclusions

The OpenEuroLLM, a consortium of 20 European research institutions, companies and EuroHPC centres, is building a family of performant, open-source, multilingual, large language foundation models for commercial, industrial and public services. The project aims to lower thresholds for European AI product development and refinement, increasing Europe's competitiveness and digital sovereignty. The main objective of this project is to provide open, transparent, and multilingual large language models (LLMs) to be further exploited by research, public and private sectors integrating the European perspective on AI.

In this deliverable, we outline a strategy for activities related to communicating, disseminating, and exploiting the project's results. We detail the objectives, expected project results, target audiences, promotion materials and activities that define our strategy along with the KPIs that we will use to measure impact. We also propose a time schedule that takes into account the most important project results and milestones timeline. This steering strategy might be further adapted to the project needs as the project evolves.

The project has already started implementing most of the activities with tangible and fruitful results. Communication has centered our efforts in these first 6 months and will be further complemented with the rest of activities planned in our strategy.