



European Research Council
Executive Agency

Established by the European Commission

Brussels, 27/03/2025
ercea.b.dir(2025)3672409

DATA PROTECTION NOTICE

ERCEA/Scientific Management Department

Classification of ERC projects through the use of Large Language Models (LLMs) for Mapping Frontier Research (MFR)

This notice concerns the processing operation called "**Classification of ERC projects through the use of LLMs for MFR**". This processing requires the handling of personal data and is therefore subject to Regulation (EU) 2018/1725 (Data protection regulation)^[1].

1. What personal information do we collect, from where and for what purpose?

1.a) Personal data

The personal data in question is contained in part B (part B1 and part B2) of the funded projects' proposal data. Only the scientific part of the funded projects' proposal data called "section A" is processed in the LLMs and will serve to classify the projects.

1.b) Purposes of the processing

The purpose of the processing of funded projects' proposal data (text of part B1 and part B2) is to test the feasibility of – and potentially deploy - the use of LLMs (generative AI) for the classification of ERC projects into the MFR dimensions Disciplines (175 terms), Topics (584 terms) and Methods (162 terms) of the MFR taxonomy designed by the ERC Scientific Council.

The scientific part (section A) of the funded projects' proposal data is put in an LLM prompt, processed by the LLM and the LLM output consist only of Disciplines, Topics or Methods terms for each project with eventually some general text explaining the choice of these terms. No data is kept by any of the LLM providers nor used for further training of the LLM by the providers.

No further processing of the funded projects' proposal data is foreseen for any other purpose than for the MFR classification operation described in the present document.

¹ REGULATION (EU) 2018/1725 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC.

1.c) From where the data is collected

The funded projects' proposal data is collected either from the current central DIGIT systems of DG R&I (SEP etc.) or from dedicated ERCEA sources (share folders or ERCEA Archive).

2. Who has access to your information and to whom is it disclosed?

The information will be made available to ERCEA staff on a need-to-know basis. As such, the funded projects' proposal data and the MFR classification data generated by the LLM is primarily to be used by the scientific officers of Department B, B0 F2P staff and shall be made available through MFR tool (D1)..

In addition, certain administrative details may be disclosed, in compliance with the relevant current legislation and established case law, and on a temporary basis to legislative or supervisory bodies of the ERCEA, as well as auditing bodies.

3. What are your rights?

You have the right to access the personal data the ERCEA holds about you and to request to have them rectified where necessary. Where applicable, you have the right to request restriction or to object to processing, to request a copy or erasure of your personal data held by the data controller.

To exercise any of these rights, you should apply to the Head of Scientific Management Department (Dpt B) by delegation, who is responsible for such processing (i.e. the Controller), by sending an e-mail specifying your request to the mailbox indicated in point 5. Please note that in some cases restrictions under the terms and conditions of Article 25 of the Data protection regulation may apply.

4. How long do we keep your data?

The proposal data in raw text format of year N is kept until the MFR classification of year N is completed (no classification errors detected) and may be deleted once the exercise of year N + 1 is launched unless it is needed to make some comparison between 2 years. After 2 years, the proposal data in raw text format is deleted.

For GPT@JRC, *“maximum 90 days from the day of the input by the user”* (cf. [DP record](#)), unless one chooses Sensitive Non-Classified (SNC) information sensitivity level in which case chat history functionality is disabled.

For GPT@EC, similar infrastructure as JRC (cf. [DP record](#)).

For Microsoft Azure OpenAI, the Commission has opt-out of prompt logging for content filtering so there is no retention of prompts

5. Contact information

If you would like to receive further information, you can contact the Head of Scientific Management Department (the Data Controller by delegation from the ERCEA Director) via the mailbox: ERC-B-WORKSHOP@ec.europa.eu.

The ERCEA Data Protection Officer is at your disposal for any clarification you might need on your rights under Regulation (EU) 2018/1725 at the following e-mail address:

ERC-DATA-PROTECTION@ec.europa.eu

You have the right to have recourse (i.e. you can lodge a complaint) to the European Data Protection Supervisor (edps@edps.europa.eu) if you consider that your rights under Regulation (EU) 2018/1725 have been infringed as a result of the processing of your personal data by the Data Controller.

6. Legal basis

The legal basis applying to these processing operations is:

Article 5(1)(a) of Regulation (EU) 2018/1725: Processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the Union institution or body.