



## On approval of the Rules for the formation and maintenance of the National Vocabulary Fund of the Kazakh language

### *Unofficial translation*

Order of the Minister of Science and Higher Education of the Republic of Kazakhstan dated April 30, 2025 № 226. Registered with the Ministry of Justice of the Republic of Kazakhstan on May 6, 2025 № 36081

### Unofficial translation

Pursuant to paragraph 4 of Article 24-4 of the Law of the Republic of Kazakhstan “On Culture”, and also subparagraph 160-2) of paragraph 15 of the Regulation on the Ministry of Science and Higher Education of the Republic of Kazakhstan, approved by Resolution of the Government of the Republic of Kazakhstan dated August 19, 2022 № 580, I hereby **ORDER:**

1. Approve the attached Rules for the formation and maintenance of the National vocabulary fund of the Kazakh language.
2. The Language Policy Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan shall ensure, in accordance with the procedure established by law :
  - 1) state registration of this order with the Ministry of Justice of the Republic of Kazakhstan;
  - 2) posting of this order on the official Internet resource of the Ministry of Science and Higher Education of the Republic of Kazakhstan.
  3. Control over the execution of this order shall be assigned to the supervising vice-minister of Science and Higher Education of the Republic of Kazakhstan.
  4. This order shall come into effect ten calendar days after the date of its first official publication.

**Minister of Science and Higher Education  
of the Republic of Kazakhstan     S.Nurbek**

"AGREED"

Ministry of Culture and Information  
of the Republic of Kazakhstan

" AGREED "

Ministry of Digital Development, Innovations  
and Aerospace Industry  
of the Republic of Kazakhstan

Approved  
by the Order  
of the Minister of Science

## **Rules for the formation and maintenance of the National vocabulary fund of the Kazakh language**

### **Chapter 1. General provisions**

1. The Rules for the formation and maintenance of the National vocabulary fund of the Kazakh language (hereinafter – The Rules) have been developed in pursuance of paragraph 4 of Article 24-4 of the Law of the Republic of Kazakhstan “On Culture” (hereinafter – the Law ) subparagraph 160-2) of paragraph 15 of the Regulation on the Ministry of Science and Higher Education of the Republic of Kazakhstan, approved by the Resolution of the Government of the Republic of Kazakhstan dated August 19, 2022 № 580, and define the procedure for the formation and maintenance of the National vocabulary fund of the Kazakh language (hereinafter – the National vocabulary fund).

2. The following basic concepts shall be used in the Rules:

1) generation – the process of automatically creating new content (text, images, sound) based on data and pre-established rules;

2) artificial intelligence models – the process of implementing artificial intelligence technologies and algorithms for processing data in the Kazakh language into various systems, applications or platforms;

3) natural language processing – a machine learning technology that enables computers to understand, interpret and process human language.

3. The objectives of the formation of the National vocabulary fund shall be preservation, protection and development of the Kazakh language as a cultural value, also enhancing the state language status, accumulation of resources covering all application areas of the Kazakh language, streamlining, digitization, generation of the current vocabulary stock of the language, its adaptation to artificial intelligence and modern technologies, automation of processing processes.

### **Chapter 2. Procedure for the formation of the National vocabulary fund of the Kazakh language**

4. Formation and maintenance of the National vocabulary fund shall be performed by a legal entity assigned by the authorized body in the language development, in accordance with paragraph 4, article 24-4, of the Law.

5. Formation of the National vocabulary fund involves providing users of the system with access to functional and information services.

6. The National vocabulary fund shall be formed on the basis of academic and translation dictionaries, the National Kazakh Language corpora, terminology base and a dataset, which is a set of data in various formats.

7. The main objectives of the formation of the National vocabulary fund shall be:

- 1) formation of a set of dictionaries enabling users to obtain comprehensive information about words;
  - 2) ensuring the placement, updating and dissemination of information about the norm of the Kazakh literary language;
  - 3) providing users with the opportunity of using electronic versions of verified dictionaries of various types;
  - 4) providing users with information on the dynamics of the development of the Kazakh literary language norm.
8. Formation of the National vocabulary fund shall include:
- 1) development of a step-by-step action plan on creation and development of the National vocabulary fund;
  - 2) approval of functional and technical requirements for the National vocabulary fund, a schedule for the provision of technical services and technical specifications of the information system designed to collect, process and systematize data;
  - 3) provision of an interactive user interface, a search engine and data export capability;
  - 4) coordination of actors in the field of creation and improvement of the National vocabulary fund;
  - 5) formation of the list of information resources capable of integration with the National vocabulary fund;
  - 6) use of open data platforms and ensuring compliance with state standards in the information technology and information security.
9. When forming the National vocabulary fund, the legal entity determined by the authorized body in the language development shall be guided by the following principles:
- 1) scientific validity, the need to rely on scientific research and factual data of all lexical units in the field of social-humanitarian and natural-mathematical sciences;
  - 2) systematicity, the need for complete and further improvement as a unique system;
  - 3) ensuring compliance with the norms of literary language;
  - 4) combination of traditions and innovation, the need for harmonious adaptation and introduction of new words and terms in accordance with the norm of the Kazakh language vocabulary and the modern requirements;
  - 5) accessibility and inclusiveness of all data for users of language resources.
- The National vocabulary fund formation shall factor in the public discussion of proposals concerning creation and improvement of this fund. Public discussion shall be conducted through open online platforms or public events with the possibility of proposals and comments from all interested parties. Independent experts in the socio-humanitarian and natural-mathematical sciences shall be involved in the formation and improvement of the National vocabulary fund.

### **Chapter 3. Procedure for maintenance of the National vocabulary fund**

10. Maintenance of the National vocabulary fund shall include:

- 1) development of dictionaries database including meaning, etymology and patterns of common usage of words;
- 2) creation of a centralized system of language corpora;
- 3) improvement of the terminological base of the Kazakh language;
- 4) development of technical requirements for collecting datasets including all application areas of the Kazakh language;
- 5) provision of an accessible platform for scientific research;
- 6) integration of artificial intelligence models, modern technologies in the Kazakh language;
- 7) development of an interactive search system;
- 8) collection, processing, storage of a database entered into the information system;
- 9) availability of publicly available information in the open space of the Internet;
- 10) information exchange of data between the information system actors;
- 11) provision of text materials in digital format;
- 12) provision of reliable and high-performance server programs;
- 13) creation of relational and semantic links of language data;
- 14) compliance with information security measures;
- 15) work to update and improve the information system;
- 16) measures to update the National vocabulary fund on its testing and approbation;
- 17) development of a guide for self-study of the information system for users;
- 18) ensuring activities for sorting and examining words, terms and phrases that correspond to the language norm, included in the National vocabulary fund;
- 19) coordinating the activities of collegial actions carried out outside the system for the National Vocabulary Fund.