

Scientists from the Airi Institute, together with colleagues from Constructor University in Germany, developed artificial intelligence (AI), which is able to generate new protein molecules for the creation of drugs. The model is called Dima and is based on the principle of latent diffusion.

The main feature of Dima is that it is much more compact than analogues, but at the same time demonstrates higher effectiveness. The system can create protein sequences that do not exist in nature, and select them in accordance with pre-given conditions. This is important for the development of drugs requiring proteins with unique properties.

The protein is a chain of amino acids, which turns in space in a certain form. It is the structure that determines its functions and characteristics. Previously, language models were used to generate proteins, which formed sequences gradually or entirely. However, such approaches required huge computing resources and large data sets.

Dima uses a new method – continuous Gaussian diffusion. At the first stage, it was taught to form biologically correct proteins that do not repeat well-known natural options. Then the system learned to generate proteins with specified characteristics, for example, a certain three-dimensional structure or belonging to the desired family.

This approach opens up the ability to create protein options that have not been found in nature, but correspond to the goals of researchers. This expands the ideas about the possible configurations of proteins and gives new tools for biotechnology and medicine.

The results were presented at the ICML 2025 International Machine Learning Conference.