



OpenAI's technology explained

OpenAI was created in 2015 to ensure that artificial general intelligence benefits everyone. We research, develop, and release cutting edge AI technology as well as tools and best practices for the safety, alignment, and governance of AI.

How We Develop Foundation Models

As part of OpenAI's mission, we develop leading foundation models and make their capabilities available in safe and beneficial ways to people around the world.

Developing a foundation model like GPT-4 requires:

PRE-TRAINING the model to teach it intelligence, such as the ability to predict, reason, and solve problems;

POST-TRAINING the model to align it to human values and preferences and make it useful, effective and safe.

Pre-Training Teaches a model language by analyzing a wide range of information, such as text, images, audio or video and learning from relationships in the information. This process, which requires significant computational power, enables the model to understand language, grammar, context, and concepts. However, the resulting "base model," while intelligent and capable of solving problems, is not user-friendly and useful to expected levels, as it could produce inappropriate content and lack nuanced control absent post-training.

Post-Training Post-training refines the base model by incorporating feedback from people, including academics and experts, making it safer, more accurate, and aligned with user needs. Techniques like reinforcement learning guide the model to follow instructions, avoid harmful content, and improve factual accuracy. This stage also allows for fine-tuning, enabling customers to adapt models for specific tasks or industries while maintaining safety measures. We also provide our *Model Spec*, a document that outlines the desired behavior of our models to help ensure they operate as intended while maintaining safety and alignment with user needs.

The ability to provide transformative and helpful AGI to everyone is rooted in developing AI systems that learn from a diverse range of data and knowledge. We want our AI models to learn from as many languages, cultures, subjects, and industries as possible so they can benefit as many people as possible.

Key Terms

ARTIFICIAL INTELLIGENCE (AI) A branch of computer science whose goal is to create computing systems that can behave in a way typically associated with human intelligence.

MACHINE LEARNING (ML) an approach to artificial intelligence where computer systems can learn to accomplish tasks based on information or experimentation, rather than being programmed step by step.

MODELS computer programs that are developed using artificial intelligence and machine learning techniques.

FOUNDATION MODELS AI models that are developed using large amounts of computational power to learn from a large amount of data, in order to perform a broad range of tasks related to that data.



Safeguarding data at every stage of model training and deployment while continuously enhancing privacy, transparency and data protection standards.



AI Transparency Leadership

OpenAI has led the AI field in transparency around the way we train our models and make their capabilities available in safe and beneficial ways to people around the world. Our publication of detailed *technical reports*, *model system cards*, *articles*, and other *research* is designed to empower users and the public with clear information.



Giving Users Control Over Their Data

We provide users with easily accessible *controls* over their data, offering options to manage data preferences, including temporary chats, controls to manage whether their chats are used for model training, exporting, reviewing, or deleting data, and making privacy requests through our *Privacy Center*.



We Reduce The Processing Of Personal Data

OpenAI implements a series of reinforcing privacy-protective measures, including advanced data filtering processes to reduce personal information from training data, and post-trains our models to reject requests for private or sensitive information.

We Design Our AI Models To Be Learning Machines, Not Databases

AI models learn from relationships in information to create something new; they don't store data like a database. Our models are designed to help us generate new content and ideas – not to repeat or “regurgitate” content. AI models may occasionally repeat widely quoted content from training datasets, but we use advanced techniques to minimize this and continuously improve through ongoing research.