OpenAI is a research and deployment company.
Our mission is to ensure that artificial general intelligence (AGI) benefits all of humanity.

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.

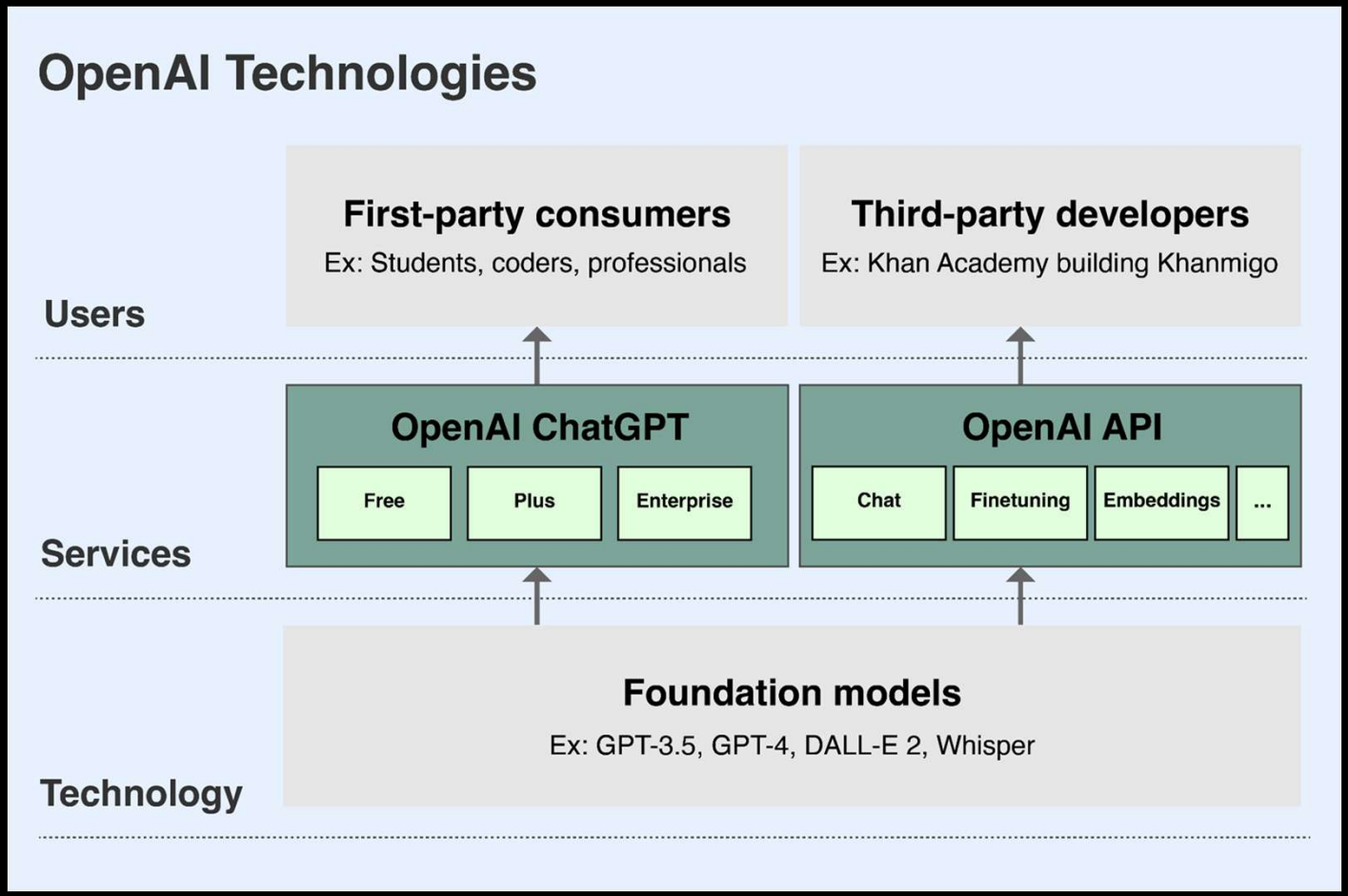
Developing an advanced language model like GPT-4 requires:

Pre-Training: teaching the model intelligence, e.g. to predict, reason, solve problems

Post-Training: aligning the model to human values and preferences through targeted changes based on pioneering techniques including **reinforcement learning human feedback (RLHF)**, **customer fine-tuning,** and **red teaming** and **evaluations**

Safety systems: usage policies prohibiting models from being used in ways that could cause harm; system card reports describing model or system capabilities, limitations, and domains of appropriate or inappropriate use; human and machine-based review, monitoring, and moderation systems.

Technology Explained: Foundation Model Development



Technology Explained: How People Use Foundation Models

A small sampling of what's possible today

- Idea Generation & Exploration
- Text Generation
- Question Answering
- Language Translation
- Summarization
- Market Research
- Mathematical Problem Solving
- Event or Meeting Planning
- Scientific Explanation
- Value Innovation
- Brainstorm & Strategize
- Educational Research & Personalized Tutoring
- Conversational Engagement
- Sentiment, Emotion & Tone Recognition
- Philosophical Discussion
- Cultural and Social Context Understanding
- Ethical Consideration and Guidance
- Write & Explain Code
- Translate Code
- Optimize Code
- Code Refactoring
- Syntax Checking
- Analyze Code Complexity
- Code Testing
- Code Documentation
- Code Conversion to Pseudocode
- Code Review and Feedback
- Code Integration with APIs
- Code Security Analysis
- Code Compatibility Check
- Code Deployment Guidance
- Database Query Generation
- Text to SQL or Spark
- ML Model Implementation
- Data Extraction
- Data Transformation
- Data Loading
- ETL Pipeline Creation
- Data Cleaning
- Data Analysis
- Data Visualization
- Interactive Dashboard Creation
- ML Preprocessing
- Data Aggregation & Processing
- Geospatial Data Visualization
- Time Series Analysis
- Data Anonymization
- Data Integration
- Data Mining
- Data Quality Assessment
- Sample Data Generation

the short term

First, as we create successively more powerful systems, we want to deploy them and gain experience with operating them in the real world. We believe this is the best way to carefully steward AGI into existence—a gradual transition to a world with AGI is better than a sudden one. We expect powerful AI to make the rate of progress in the world much faster, and we think it's better to adjust to this incrementally.

A gradual transition gives people, policymakers, and institutions time to understand what's happening, personally experience the benefits and downsides of these systems, adapt our economy, and to put regulation in place. It also allows for society and AI to co-evolve, and for people collectively to figure out what they want while the stakes are relatively low.

the long term

AI that can accelerate science is a special case worth thinking about, and perhaps more impactful than everything else. It's possible that AGI capable enough to accelerate its own progress could cause major changes to happen surprisingly quickly (and even if the transition starts slowly, we expect it to happen pretty quickly in the final stages).

Advanced Research Directions on AI for Science, Energy, & Security

AI & Surrogate Models for Scientific Computing	Climate Modeling Fusion Energy Predictive Multiphysics Simulations Cosmology
AI Foundation Models for Scientific Knowledge Discovery, Integration, and Synthesis	Stockpile Modernization Knowledge Distillation (unstructured to structured knowledge) and Hypothesis Formation Digital Twins for Engineering Complex Scientific Domains
AI for Advanced Property Inference & Inverse Design	Materials, Chemistry, & Biology Design (atomic / molecular scale) Engineered Structures / Systems (continuum scale) Non-proliferation / Decision Superiority (process / protocols)
AI-Based Design, Prediction, and Control of Complex Engineered Systems	Hi-rep Rate Laser Accelerators Reactors (Fusion and Fission)
AI for Robotics for Autonomous Discovery	Nuclear Weapons Design Transformation Accelerated Discovery in Materials, Chemistry, and Biology Advanced Manufacturing
AI for Programming & Software Engineering	Adaption of Codes for New Computational Targets Discovering Quality Control Algorithms and Quality Control Optimization Ai-Driven Co-Design

- Advancing AI safety research to promote responsible development of frontier models, minimize risks, and enable independent evaluations of capabilities and safety.

- Identifying best practices for the responsible development and deployment of frontier models, helping the public understand the nature, capabilities, limitations, and impact of the technology.

- Collaborating across industry, policymakers, academics, civil society and companies to share knowledge about trust and safety risks.

- Supporting efforts to develop applications that can help meet society's greatest challenges, such as climate change mitigation and adaptation, early cancer detection and prevention, and combating cyber threats.

Safe & Beneficial AI for Science, Energy, & Security