

КАЗАХСКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ ИМЕНИ АЛЬ-ФАРАБИ

А.Т. Агишев

МЕТОДЫ И ТЕХНОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

**Сборник лекций для студентов магистратуры,
обучающихся по образовательной программе
«7M06201 - Радиотехника, электроника и телекоммуникации»**

Алматы, 2025

Лекция 9. Инфраструктура и экосистема искусственного интеллекта

Цель лекции

Познакомить студентов с вычислительной и программной инфраструктурой, необходимой для разработки и внедрения систем искусственного интеллекта. Рассмотреть ключевые компоненты экосистемы ИИ - от аппаратных платформ до облачных сервисов и инструментов разработки, а также показать роль инфраструктуры в инженерных и телекоммуникационных приложениях.

Основные вопросы:

1. Понятие инфраструктуры искусственного интеллекта.
2. Аппаратная инфраструктура: CPU, GPU, TPU, FPGA.
3. Вычислительные кластеры и облачные технологии.
4. Программная экосистема ИИ: фреймворки и библиотеки.
5. Среды разработки и инструменты MLOps.
6. Хранение, обработка и управление данными.
7. Инфраструктура ИИ в телекоммуникациях и инженерных системах.
8. Тенденции развития и стандартизации инфраструктуры ИИ.

Краткие тезисы:

1. **Понятие инфраструктуры искусственного интеллекта.**
Инфраструктура ИИ - это совокупность аппаратных, программных и сетевых ресурсов, обеспечивающих разработку, обучение, развертывание и эксплуатацию интеллектуальных систем. Она охватывает:
 - вычислительные платформы (серверы, кластеры, облака);
 - инструменты для разработки и обучения моделей (Python, TensorFlow, PyTorch);
 - системы хранения и управления большими данными;
 - инструменты мониторинга и автоматизации (MLOps).

2. **Аппаратная инфраструктура.**

CPU (Central Processing Unit):

- Универсальный процессор общего назначения;
- Подходит для небольших моделей и прототипирования.

GPU (Graphics Processing Unit):

- Параллельная архитектура, ускоряющая матричные операции и обучение нейронных сетей;
- Производители: NVIDIA (CUDA, Tensor Cores), AMD ROCm. Применяются в глубоком обучении (Goodfellow et al., 2016).

TPU (Tensor Processing Unit):

- Специализированные чипы Google для TensorFlow;

- Оптимизированы для матричных операций и inference.

FPGA (Field-Programmable Gate Array):

- Перепрограммируемые логические схемы для специализированных задач ИИ;
- Используются в телекоммуникационных устройствах и радиотехнических системах для ускорения обработки сигналов.

Пример: в сетях 5G FPGA ускоряют обработку сигналов и распределение ресурсов на уровне базовых станций.

3. Вычислительные кластеры и облачные технологии.

Кластеры и HPC (High Performance Computing):

- Состоит из множества узлов, объединённых сетью (InfiniBand, Ethernet).
- Используются для параллельного обучения больших моделей.
- Примеры: HPE Apollo, NVIDIA DGX, Huawei Atlas.

Облачные платформы (AI Cloud):

- Google Cloud AI, Amazon AWS SageMaker, Microsoft Azure ML, Huawei ModelArts.
- Предоставляют ресурсы по модели IaaS/PaaS, позволяют масштабировать обучение и хранить модели.

Преимущества: гибкость, масштабируемость, снижение капитальных затрат.

4. Программная экосистема ИИ.

Основные фреймворки и библиотеки (Goodfellow et al., 2016; Alpaydin, 2020):

- **TensorFlow** - от Google, оптимизирован под GPU/TPU, поддерживает Keras API.
- **PyTorch** - от Meta, динамическое построение графов, удобен для исследований.
- **Scikit-learn** - классическое машинное обучение (SVM, RandomForest).
- **MXNet, JAX, ONNX, PaddlePaddle** - альтернативы для специализированных задач.
- **NumPy, Pandas, Matplotlib** - инструменты анализа данных и визуализации.

Среды: Jupyter Notebook, Google Colab, VS Code, PyCharm.

5. Среды MLOps и автоматизация цикла разработки. MLOps (Machine Learning Operations) - набор практик для управления жизненным циклом ML-моделей: от обучения до развертывания и мониторинга.

Основные компоненты:

- **DataOps** - управление потоками данных;
- **ModelOps** - версионирование и автоматическое обновление моделей;
- **Monitoring & Logging** - контроль точности и производительности.

Инструменты: MLflow, Kubeflow, DVC, Airflow, Docker, Kubernetes. В телекоммуникационных системах MLOps обеспечивает устойчивое обновление моделей в реальном времени (например, анализ трафика и оптимизация каналов).

6. Хранение и управление данными.

- **Data Lakes и Data Warehouses** - хранилища больших массивов данных;
- **Базы данных:** SQL (PostgreSQL, MySQL) и NoSQL (MongoDB, Cassandra);
- **Потоковая обработка:** Apache Kafka, Spark Streaming;
- **Инструменты подготовки данных:** Pandas, PySpark, Dask.

Принципы: масштабируемость, безопасность, поддержка real-time анализа.

7. Инфраструктура ИИ в радиотехнике и телекоммуникациях.

- Обработка сигналов и спектров на GPU/FPGA.
- Облачное управление телекоммуникационными сетями.
- Использование edge-компьютинга (Edge AI) для обработки данных на устройствах IoT и базовых станциях.
- Пример: интеллектуальные системы распределения радиоресурсов с элементами Reinforcement Learning, реализованные на FPGA.
- Применение контейнерных технологий (Docker) для модульного развёртывания моделей ИИ в сетях 5G/6G.

8. Тенденции и стандартизация.

- **Edge AI:** перенос вычислений ближе к источнику данных (умные сенсоры, камеры, микроконтроллеры).
- **Federated Learning:** распределённое обучение без передачи исходных данных.
- **Green AI:** снижение энергопотребления и углеродного следа при обучении.
- **Open Source AI:** открытые модели и инфраструктура (Hugging Face, ONNX, PyTorch Lightning).
- **Национальные инициативы:** развитие AI-центров (NVIDIA AI Nation, Huawei AI Lab, KazNU HPC Center).

Современная экосистема ИИ строится на взаимодействии университетов, индустрии и облачных платформ.

Вопросы для контроля, изучаемого материал:

- 1) Что включает в себя понятие «инфраструктура искусственного интеллекта»?
- 2) Чем различаются CPU, GPU, TPU и FPGA с точки зрения архитектуры и применения?

- 3) Каковы преимущества использования вычислительных кластеров и облачных платформ для задач ИИ?
- 4) Какие фреймворки и библиотеки наиболее популярны в современном ИИ?
- 5) Что такое MLOps и какие инструменты применяются для его реализации?
- 6) Как обеспечивается хранение и потоковая обработка данных в проектах ИИ?
- 7) Как реализуются элементы инфраструктуры ИИ в телекоммуникационных системах?
- 8) Какие современные тенденции определяют развитие инфраструктуры ИИ (Edge AI, Green AI, Federated Learning)?

Рекомендуемый список литературных источников:

1. Russell, S., Norvig, P. Artificial Intelligence: A Modern Approach. 4th Edition. Pearson, 2021.
2. Goodfellow, I., Bengio, Y., Courville, A. Deep Learning. MIT Press, 2016.
3. Alpaydin, E. Introduction to Machine Learning. 4th Edition. MIT Press, 2020.