

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 10. Будущее и этика ИИ

Цель лекции

Рассмотреть ключевые тенденции развития искусственного интеллекта (ИИ) в ближайшие десятилетия, а также проанализировать этические, правовые и социальные вызовы, связанные с его применением. Сформировать у студентов ответственное отношение к разработке и использованию технологий ИИ в инженерной и научной практике.

Основные вопросы:

1. Тенденции и направления развития ИИ.
2. Интеллектуальные системы будущего: автономность и саморегуляция.
3. Риски и вызовы широкого внедрения ИИ.
4. Принципы этичного и безопасного ИИ.
5. Проблемы ответственности и прозрачности алгоритмов.
6. Влияние ИИ на профессию инженера и телекоммуникационные технологии.
7. Международные инициативы и нормативные документы.
8. Будущее взаимодействия человека и искусственного интеллекта

Краткие тезисы:

1. **Тенденции развития ИИ.** ИИ переходит от узкоспециализированных моделей (*Narrow AI*) к более универсальным системам, способным к контекстному рассуждению (*General AI*). Основные тенденции:
 - Конвергенция ИИ с другими технологиями: **квантовые вычисления, 6G, IoT, блокчейн.**
 - Развитие **гибридных систем**, объединяющих статистические и символические подходы.
 - Переход к **Explainable AI (XAI)** - интерпретируемым и прозрачным моделям.
 - Использование **генеративных моделей (Generative AI)** для создания контента, проектирования и симуляций.
 - Формирование **автономных киберфизических систем**, которые адаптируются без постоянного вмешательства человека.
2. **Интеллектуальные системы будущего.**
 - Системы будут сочетать **самообучение, адаптацию и взаимодействие** с другими агентами.
 - Концепция **Artificial General Intelligence (AGI)** предполагает способность машины решать широкий спектр задач на уровне человека.
 - Развитие **collaborative AI** - совместной работы человека и ИИ.

- В инженерных областях - переход от автоматизации к **координации и когнитивным решениям**, где ИИ станет активным участником проектирования и управления системами.

3. Риски и вызовы широкого внедрения ИИ. Основные проблемы:

- **Технологические:** некорректные решения из-за bias данных, уязвимости к атакам на модели (adversarial attacks).
- **Этические:** нарушение приватности, дискриминация, использование ИИ в вооружениях.
- **Экономические:** замещение рабочих мест, концентрация ресурсов у технологических корпораций.
- **Экологические:** энергозатратность обучения больших моделей (вызов «Green AI»).

В телекоммуникациях - риск утраты контроля над распределёнными сетями, если ИИ действует автономно без прозрачных механизмов контроля.

4. Принципы этичного и безопасного ИИ. Согласно рекомендациям UNESCO (2021), OECD и IEEE Ethically Aligned Design (EAD), ключевые принципы включают:

- **Справедливость и недискриминация** - исключение bias в данных и моделях.
- **Прозрачность (Transparency)** - интерпретируемость решений моделей.
- **Ответственность (Accountability)** - наличие субъектов, ответственных за решения ИИ.
- **Безопасность и надёжность (Safety & Robustness)** - защита от ошибок и манипуляций.
- **Конфиденциальность (Privacy)** - соблюдение норм GDPR и защиты персональных данных.
- **Человекоцентричность (Human-Centric AI)** - приоритет благополучия человека.

Эти принципы становятся обязательной частью инженерных и телекоммуникационных систем.

5. Ответственность и прозрачность алгоритмов. В современных ИИ-системах возникает феномен «чёрного ящика»: сложно объяснить, почему модель приняла то или иное решение. Методы **Explainable AI (XAI)** решают эту проблему:

- визуализация признаков (Grad-CAM, LIME, SHAP);
- интерпретируемые архитектуры (Decision Trees, Attention-based Models).

Для критических отраслей (энергетика, связь, транспорт) требуется обязательная верификация алгоритмов.

6. Влияние ИИ на инженерию и телекоммуникации.

- Инженеры будущего должны владеть компетенциями в области **data science, cloud computing, нейронных сетей и MLOps**;
- Телекоммуникационные системы станут **самоорганизующимися (Self-Organizing Networks)** с элементами reinforcement learning;
- Роль человека смещается от оператора к **аналитику и контролёру качества решений ИИ**;
- Развитие **«умных инфраструктур»**: автоматизированные радиосети, IoT, адаптивные антенны, предиктивное обслуживание.

Таким образом, ИИ становится ключевым компонентом инженерного интеллекта XXI века.

7. Международные инициативы и нормативные документы.

- **UNESCO Recommendation on the Ethics of Artificial Intelligence (2021)**;
- **EU Artificial Intelligence Act (2024)**;
- **OECD AI Principles (2019)**;
- **IEEE Ethically Aligned Design (EAD, 2020)**;
- **Kazakhstan AI Development Concept (2025–2030)** - развитие цифровой инфраструктуры и этического регулирования ИИ.

Все эти документы подчеркивают необходимость **этического аудита, стандартизации моделей и прозрачности данных**.

8. Будущее взаимодействия человека и ИИ.

- Формируется концепция **«Augmented Intelligence»** - ИИ не заменяет, а дополняет человека.
- В инженерии - переход к **кооперативным системам**, где человек управляет стратегией, а ИИ - тактикой.
- Будущее - за **интеграцией искусственного интеллекта, коллективного разума и киберфизических систем**, где человек остаётся центром принятия решений.
- Этика становится не ограничением, а направлением развития: от технологического прогресса к осознанному и устойчивому развитию.

Вопросы для контроля, изучаемого материал:

- 1) Какие основные тенденции определяют развитие ИИ в ближайшие десятилетия?
- 2) Что подразумевается под понятием «этичный ИИ»?
- 3) Какие риски связаны с широким внедрением ИИ в инженерных системах?
- 4) Назовите ключевые принципы этики ИИ по стандартам UNESCO и IEEE.
- 5) Что такое Explainable AI и почему интерпретируемость важна?

- 6) Как ИИ влияет на профессию инженера и телекоммуникационную отрасль?
- 7) Какие международные инициативы регулируют разработку и использование ИИ?
- 8) Почему концепция «человек + ИИ» рассматривается как основа будущего взаимодействия?

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.