**Философско-методологические основания**

**исследований искусственного интеллекта**

Лекторы:

Григорьев О.М., к.ф.н., доцент кафедры логики

Клюева Н.Ю., к.ф.н., доцент кафедры философии и методологии науки

Кузнецов А.В., к.ф.н., младший научный сотрудник кафедры истории зарубежной философии

В рамках курса слушатели познакомятся с широким спектром философско-методологических проблем, возникающих в ходе исследований искусственного интеллекта.

Мы поговорим про основные подходы к определению интеллекта и теоретико-методологические парадигмы создания искусственного интеллекта. Узнаем, что понимали под интеллектом мыслители прошлого и какие объяснительные модели механизмов познания предлагает современная когнитивная наука. Познакомимся с историей искусственного интеллекта как научно-практического направления и узнаем актуальное положение дел в этой области.

Мы рассмотрим наиболее важные методологические аспекты разработки так называемого «логического искусственного интеллекта». В этом направлении исследований основным инструментом как для теоретических, так и для практически значимых разработок служат различные логические теории – от стандартных систем классической логики, до тех разновидностей неклассической логики, которые возникли и развиваются именно в связи с изучением проблем искусственного интеллекта. Помимо общей характеристики логически ориентированного искусственного интеллекта в данной части курса будут даны и основные сведения, касающиеся указанных логических систем.

Мы рассмотрим взаимосвязь проблематики искусственного интеллекта и современной философии сознания. В фокус внимания попадут так называемые функционалисткие теории сознания. Речь пойдет о принципе множественной реализуемости сознания как главном основании функционалистских теорий, о понятии функции и критике этих теорий посредством мысленных экспериментов отсутствующих квалиа, инверсии спектра и зомби. В связи с этим будет затронуто важнейшее понятия объяснительного провала. Обратимся к критике компьютенционалистских теорий сознания на основе влиятельного аргумента Китайской комнаты Дж. Серла.

Познакомимся с концепцией «сверхинтеллекта» Н.Бострома и философскими дискуссиями о «технологической сингулярности».

**Тема 1 (2 лекции). Искусственный интеллект: история и методология (Клюева Н.Ю.)**

Искусственный интеллект как научно-практическая область исследований. Дефиниция искусственного интеллект: многообразие значений слова «искусственный интеллект». «Слабый», «сильный», «общий» искусственный интеллект. Концептуализация интеллекта в истории философской мысли. Концепция «разума как вычисления». Искусственный интеллект и психофизическая проблема в философии.

Искусственный интеллект как междисциплинарный проект по моделированию мышления при помощи ЭВМ: основные этапы развития от середины 50-х гг. 20 века до н.в. Эволюция теоретических подходов к моделированию мышления в области искусственного интеллекта: базовые принципы, преимущества, достижения, ограничения. Обзор прикладных направлений искусственного интеллекта. Эпистемологические основания моделирования мышления.

Искусственный интеллект как психологическая теория: эволюция «компьютерной метафоры» в когнитивных науках. Коннекционизм в искусственном интеллекте и когнитивной психологии.

**Тема 2 (4 лекции). Искусственный интеллект и философия сознания (Кузнецов А.В.)**

**Функционалистские теории сознания и их критика.** Функционалистские теории сознания отождествляют ментальные состояния с функциональными. В лекции речь пойдет о принципе множественной реализуемости сознания как главном основании функционалистских теорий, о понятии функции и критике этих теорий посредством мысленных экспериментов отсутствующих квалиа, инверсии спектра и зомби. В связи с этим будет затронуто важнейшее понятие объяснительного провала.

**Компьютенциализм и Китайская комната.**

Критика компьютенционалистских теорий сознания на основе влиятельного аргумента Китайской комнаты Дж. Серла. Будут даны понятия сильного и слабого ИИ, компьютенциализма в отношении ментального, тест Тьюринга, оценка значимости и эффективность аргумента Китайской комнаты, принцип множественной реализуемости.

**Тема 3 (4 лекции). Логические основы искусственного интеллекта (Григорьев О.М.)**

**Логика как методологическая основа для разработки искусственного интеллекта (ИИ).** «Логицистская» и «нелогицистская» парадигмы в разработке ИИ. Исторические аспекты эволюции логического ИИ.

Почему логика может быть эффективным инструментом для теоретической разработки ИИ и как можно использовать логические методы в ИИ? Сильные и слабые стороны логического подхода. Рациональные «логические» агенты (рациональные машины), как они устроены и как осуществляется их взаимодействие с внешним миром? Представление информации о внешнем мире посредством формализованных языков, декларативное и процедурное знание, уровни знания.

Классическая логика и ИИ. ИИ как новый вызов для развития логики, триггер для разработки новых логических теорий.

Границы применимости логических методов в ИИ.

**Логически ориентированный искусственный интеллект и новые направления исследований в логике.**

Представление знаний о мире и рассуждений о нем «на уровне здравого смысла» (commonsense knowledge, commonsense reasoning). Логические формализмы, адаптированные для экспликации рассуждений на уровне здравого смысла, их общая характеристика. Немонотонные рассуждения и их формализации. Что такое немонотонность в логике? Примеры немонотонных рассуждений. Разновидности немонотонных логик.

Представление знания как отдельная область исследования. Логические основания представления знания. Различные подходы к представлению знания рационального агента (онтологии, фреймы, семантические сети и др.).

Краткая характеристика ряда направлений исследований в области ИИ, имеющих непосредственное отношение к логическому ИИ: абдуктивные рассуждения; автоматизированные рассуждения (automated reasoning); логическое программирование; аргументативные системы; логические подходы к обработке текстов на естественном языке.

**Тема 4 (2 лекции). Перспективы имплементации искусственного интеллекта: теоретические дискуссии (Клюева Н.Ю.)**

Методология «исследований будущего»: future studies, foresight, футурология. Концептуализация образов будущего человечества в философии: от философской утопии к идее трансгуманизма и постчеловечества. Концепция «сверхинтеллекта» Н.Бострома. Философские дискуссии о «технологической сингулярности». Структура аргумента сингулярности (Д.Чалмерс). Аргумент эмуляции и эволюционный аргумент.

**Вопросы к зачету:**

1. Искусственный интеллект: проблема дефиниции.
2. История искусственного интеллекта: основные этапы, теоретические подходы и достижения.
3. Искусственный интеллект и психофизическая проблема.
4. Искусственный интеллект как психологическая теория: «компьютерная метафора» в науках о познании.
5. Аргумент китайской комнаты Дж. Серла.
6. Принцип множественной реализуемости и функционализм в отношении сознания
7. Антифункционалистские аргументы и провал в объяснении.
8. «Сильный» и «слабый» искусственный интеллект.
9. Рациональные логические агенты, их устройство и взаимодействие с внешним миром. Декларативное и процедурное знание.
10. Сильная и слабая стороны логического подхода к разработке искусственного интеллекта. Границы применимости логического подхода.
11. Искусственный интеллект и новые направления исследований в логике.
12. Немонотонные рассуждения, их формальное представление, разновидности немонотонных логик.
13. Рассуждения «на уровне здравого смысла» в контексте исследований в области искусственного интеллекта, их возможные формализации.
14. Логические основания представления знания. Различные подходы к представлению знания.
15. Понятие «сверхинтеллекта» Н.Бострома
16. Структура аргумента сингулярности (Д.Чалмерс)