Федеральное государственное бюджетное образовательное учреждение   
высшего образования

Московский государственный университет имени М. В. Ломоносова

Философский факультет

|  |  |
| --- | --- |
|  | УТВЕРЖДАЮ  и.о. декана философского факультета МГУ  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_/А. П. Козырев/  «\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2023 г. |

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

Наименование:

**Межфакультетский курс «Философско-методологические основания**

**исследований искусственного интеллекта»**

Уровень высшего образования:

**Бакалавриат, магистратура, специалитет**

Форма обучения:

**Очная**

Язык обучения:

**Русский**

Авторы программы:

**Григорьев Олег Михайлович, Клюева Наталья Юрьевна,**

**Кузнецов Антон Викторович**

Рабочая программа рассмотрена и одобрена

*на заседаниях кафедр логики,*

*философии и методологии науки и*

*истории зарубежной философии*

Москва 2023

Рабочая программа дисциплины разработана в соответствии с самостоятельно установленным МГУ образовательным стандартом (ОС МГУ) для реализуемых основных профессиональных образовательных программ высшего образования (бакалавриата и магистратуры, реализуемых последовательно по схеме интегрированной подготовки; специалитета; магистратуры).

# Аннотация к дисциплине

Основная цель курса — знакомство учащихся с наиболее важными и ключевыми понятиями, широко используемыми в науках, связанных с проблематикой разработки и практической реализации систем искусственного интеллекта, с философскими и методологическими идеями, лежащими в основании исследований искусственного интеллекта, а также с фундаментальными проблемами, которые встают перед учеными, деятельность которых так или иначе связана с изобретением и внедрением технологий искусственного интеллекта. Для достижения указанной цели предполагается решение ряда взаимосвязанных задач курса. В первую очередь это ознакомление учащихся с историей и общей методологией исследований по искусственному интеллекту, раскрытие междисциплинарного характера этих исследований. Изложение различных подходов к экспликации понятия интеллекта и моделирования мышления в когнитивных науках. Изучение связи искусственного интеллекта и основной проблематики современной философии сознания. Существенная часть содержания курса отводится логико-философскими основаниями исследований искусственного интеллекта. В этой части программы осуществляется как изложение необходимого минимума знаний, связанных с базовыми принципами классической логики, так и обзор основных направлений логических исследований, возникших под влиянием потребностей моделирования процессов рассуждения и коммуникации, производимых носителями той или иной формы искусственного интеллекта. Материал этой части курса построен так, чтобы, с одной стороны не нарушать строгость изложения, присущего логической науке, с другой стороны, быть доступным для понимания студентами различных образовательных направлений.

# Входные требования для освоения дисциплины

Для успешного освоения курса «Философско-методологические основания

исследований искусственного интеллекта» специальных входных требований нет. Желательным является знакомство с базовыми курсами по философии, истории философии и логике.

# Место дисциплины в структуре основной образовательной программы (ООП)

Дисциплина является межфакультетским образовательным курсом.

# Планируемые результаты обучения по дисциплине

|  |  |
| --- | --- |
| **Формируемые компетенции** | **Планируемые результаты обучения** |
| **На уровне магистратуры:** способен использовать философские категории и концепции при решении социальных и профессиональных задач;  **В том числе на уровне бакалавриата:** способен применять философские категории, анализировать философские тексты и учитывать философские проблемы при решении социальных и профессиональных задач **(УК-2)**. | **ЗНАТЬ:** основные категории и понятия, используемые в междисциплинарных и специальных исследованиях по искусственному интеллекту, фундаментальные направления и подходы в изучении теоретических основ искусственного интеллекта, философские и методологические проблемы данной области научного знания.  **УМЕТЬ:** использовать основные положения и категории исследований искусственного интеллекта для анализа и решения междисциплинарных задач.  **ВЛАДЕТЬ:** логико-философской проблематикой исследований искусственного интеллекта. |

# Форма обучения

Очная.

# Общая трудоемкость дисциплины

Общая трудоемкость дисциплины составляет 1 зачетная единица, 24 академических часа лекций и 12 академических часов самостоятельной работы студента.

# Учебно-тематический план

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **№** | **Разделы и темы** | **Лекции (ак. ч.)** | **Самостоятельная работа** | **Формы контроля** |
| **1** | **Искусственный интеллект: история и методология** | 4 | 2 | Текущий контроль |
| **2** | **Искусственный интеллект и философия сознания** | 6 | 4 | Текущий контроль |
| **3** | **Логические основы искусственного интеллекта** | 8 | 4 | Текущий контроль |
| **4** | **Перспективы имплементации искусственного интеллекта: теоретические дискуссии** | 4 | 2 | Текущий контроль |
| Промежуточная аттестация | | 2 |  | Зачет |
| **Итого** | | **24** | **12** |  |

# Учебная программа

**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.

# Форма промежуточной аттестации и фонд оценочных средств

## 9.1 Формы и оценка текущего контроля

Текущий контроль подразумевает оценку освоения студентами основных понятий, используемых в междисциплинарных и специальных исследованиях по искусственному интеллекту, понимание сути наиболее важных направлений и подходов, существующих в данной предметной области. В качестве формы оценки могут быть использованы небольшие письменные проверочные работы, в которых слушателям предлагается раскрыть смысл тех или иных специальных терминов, суть того или иного эксперимента, исследовательского подхода и т.п.

## 9.2 Формы и оценка самостоятельной работы

Самостоятельная работа подразумевает освоение студентами понятийно-категориального аппарата исследований по искусственному интеллекту с использованием лекционного материала и рекомендуемой литературы. Форма оценки аналогична указанной в п. 9.1.

## 9.3 Форма и оценка промежуточной аттестации

Аттестация проводится в форме **зачета**. Примерный список вопросов:

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)

# Ресурсное обеспечение:

**Энциклопедические издания и словари**

1. Национальная философская энциклопедия http://terme.ru/
2. Новая философская энциклопедия http://iph.ras.ru/enc.htm
3. Философский портал http://www.philosophy.ru
4. Электронная библиотека по философии: http://filosof.historic.ru
5. Электронная гуманитарная библиотека http://www.gumfak.ru/
6. Britannica - www.britannica.com
7. Stanford Encyclopedia of Philosophy http://plato.stanford.edu/
8. The Internet Encyclopedia of Philosophy (IEP) http://www.iep.utm.edu/

**Основная литература**

1. Бостром Н. Искусственный интеллект. Этапы. Угрозы. Стратегии. М.: Манн, Иванов и Фербер, 2016.
2. Дубровский Д.И. Сознание, мозг, искусственный интеллект. М.: «Стратегия-Центр», 2007.
3. Васильев В.В. Сознание и вещи. М., 2014.
4. Национальная стратегия развития искусственного интеллекта на период до 2030 года // Официальный интернет-портал правовой информации. 2019.11.10. [Электронный ресурс]. URL: http://publication.pravo.gov.ru/Document/View/0001201910110003 (дата обращения: 20.03.2021)
5. Прист С. Теории сознания. М., 2000.
6. Рассел С., Норвиг П., Искусственный интеллект. Современный подход. 4-е издание. / Пер. с англ. – Том 1-3. Вильямс, 2021-22.
7. Серл Дж. Мозг, сознание и программы // Аналитическая философия: становление и развитие (антология). М. – 1998. – С. 376-400.
8. Чалмерс Д. Сознающий ум. М., 2016.

**Дополнительная литература**

1. Васильев В.В. Трудная проблема сознания. М., 2009.
2. Волков Д. Б. Бостонский зомби: Д. Деннет и его теория сознания. М., 2012.
3. Деннет Д. Сладкие грезы. М., 2017.
4. Деннет Д., .Хофштадтер Д. Глаз Разума М., 2003.
5. Дрейфус Х. Чего не могут вычислительные машины. - М.: Прогресс, 1978.
6. Дубровский Д.И. Информация, сознание, мозг. М.: «Высшая школа», 1980
7. Дубровский Д.И. Проблема «Сознание и мозг» теоретическое решение М.: «Канон+» РООИ «Реабилитация», 2015
8. Дубровский Д.И. Проблема идеального. Субъективная реальность. Второе доп. изд. М.: «Канон+» , 2002
9. Иванов Д.В. Природа феноменального сознания М., 2012.
10. Клюева Н.Ю. Влияние идеи Г. Лейбница на развитие компьютерных наук и исследования в области искусственного интеллекта // Вестник Московского университета. Серия 7: Философия, издательство Изд-во Моск. ун-та (М.), № 4, с. 79-93, 2017.
11. Митчам К. Что такое философия техники? - М.,1995.
12. Люгер Д. Искусственный интеллект. Стратегии и методы решения сложных проблем. - М.:Издательство Вильямс, 2014.
13. Нагуманова С.Ф. Материализм и сознание: анализ дискуссии о природе сознания в современной аналитической философии. Казань, 2011.
14. Пинкер П. Субстанция мышления. Язык как окно в человеческую природу. Либроком, 2016.
15. Райл Г. Понятие сознания. М., 2000.
16. Рассел С. Совместимость. Как контролировать искусственный интеллект / Пер. с англ. – М.: Альпина нон-фикшн, 2021.
17. Тьюринг А. Могут ли машины мыслить? М.,1960.
18. Серл Дж. Р. Сознание, мозг и программы // Аналитическая философия: становление и развитие. М., 1998.
19. Серл Дж. Открывая сознание заново. М., 2002.
20. Хамфри Н. Сознание. Пыльца души. Карьера Пресс, 2014.
21. Chalmers D. The Singularity: A Philosophical Analysis Journal of Consciousness Studies 17:7-65, 2010.
22. Bringsjord, Selmer and Naveen Sundar Govindarajulu, "Artificial Intelligence", The Stanford Encyclopedia of Philosophy (Fall 2022 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = https://plato.stanford.edu/archives/fall2022/entries/artificial-intelligence.
23. Thomason, Richmond, "Logic and Artificial Intelligence", The Stanford Encyclopedia of Philosophy (Summer 2020 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/sum2020/entries/logic-ai.

# Материально-техническая база

Требуется мультимедийная аудитория с проектором.

# Язык преподавания.

Русский

# Преподаватели.

Григорьев Олег Михайлович, кандидат философских наук, доцент

Клюева Наталья Юрьевна, кандидат философских наук, доцент

Кузнецов Антон Викторович, кандидат философских наук, научный сотрудник

# Авторы программы.

Григорьев Олег Михайлович, кандидат философских наук, доцент

Клюева Наталья Юрьевна, кандидат философских наук, доцент

Кузнецов Антон Викторович, кандидат философских наук, научный сотрудник