Введение в классическую и цифровую этику

Аннотация

Курс лекций предлагает изложение ключевых этических теорий с последующим их применением к вопросам, связанным с цифровыми технологиями. В рамках курса исследуются сходства и различия между традиционной и цифровой этикой, а также рассматриваются аргументы в пользу выделения цифровой этики в отдельную дисциплину. Основная идея курса заключается в том, что цифровая этика не может рассматриваться лишь как раздел прикладной этики. Во-первых, потому что цифровые технологии создают виртуальные пространства, в которых возникают действия, требующие особой этической оценки. Во-вторых, сокращение дистанции между действиями человека и действиями машин усиливает необходимость расширения понятий свободы и ответственности, что требует переосмысления базовых этических принципов. Курс состоит из двух частей: первая часть посвящена классическим концепциям в этике, а вторая — вопросам этических теорий в контексте цифровизации мира. Завершает курс лекция под названием «Проект „Просвещение“ в эпоху цифровых технологий», в которой рассматривается попытка применения идей эпохи Просвещения для решения актуальных проблем цифрового общества.

Длительность курса

Один семестр. Курс состоит из 10 лекций.

Формат курса

Одна лекция и один семинар в неделю. После каждой лекции будет проводиться закрепляющий семинар, на котором слушатели будут делать небольшие презентации по пройденным темам.

Уровень курса

Базовый.

Пререквизиты

Курс является введением в этику, и не требует предварительного знания этических теорий. Строго говоря, уровень курса - базовый, однако знание основ практической философии и истории философии приветствуются. Слушателям, не обладающим начальными знаниями практической философии нужно быть готовыми дополнительно работать с бОльшим количеством литературы.

Список тем

Часть 1 (введение в классическую этику)

1 Лекция. Наука этика

Будут рассмотрены основные вопросы этики, и мета-этики. Сделаем обзор этических подходов, определим три основных класса этических теорий.

2 Лекция. Утилитаризм

Рассмотрим принципы утилитаризма. Поговорим о количественный утилитаризм И.Бентама и качественном утилитаризме Дж.С.Милля а так же об утилитаризме предпочтений П.Сингера.

3 Лекция. Деонтологическая этика

Определим принципы этики И.Канта. Человек как существо чувственного и разумного миров. Человеческое достоинство. Добрая воля. Максимы. Три типа действий - аморальные, легальные, моральные. Гипотетические императивы. Технические императивы. Прагматические императивы. Категорический императив. Универсализирующая формула. Формула самоцели.
Преимущества и недостатки деонтологической этики

4 Лекция. Этика добродетелей

На этой лекции будем говорить о этике Аристотеля, определении добродетели, приобретение добродетели. Поговорим о современных представлениях о добродетели (первичные, интранзитивные добродетели и вторичные инструментальные добродетели).

5 Лекция. Теории справедливости

Принципы справедливости у Аристотеля. Уравнительная справедливость. Распределительная справедливость. Теория справедливости Дж. Ролза. Завеса неведения. Оправдание неравенства. Теория справедливости Р.Нозика. Принцип добровольного обмена. Государство - ночной сторож.
А.Макинтайр. Комунитаризм. Человек часть истории.

Часть 2 (Основы цифровой этики)

6 Лекция. Моральное действие в цифровой среде

Свойства цифровых сред. Эффект масштабирования. Эффект первопроходца. Цифровой след. Эффекты цифровизации общества. Экономические эффекты. Политические эффекты. Социальные эффекты.
Конфиденциальность. Исчезновение частного. Зачем нам конфиденциальность? Конфиденциальность как философская проблема. Защита данных. Цифровое наблюдение: безопасность и человеческое достоинство. Данные как товар. Дилемма: защита данных & развитие инноваций. Человек как цель и как средство. Теория подталкивания. Индивидуальность и самоопределение. Субъектность и объективация человека в цифровой среде.

7 Лекция. Может ли машина быть моральной?

Сознательные машины. Тест Тьюринга. Китайская комната. Трудная проблема сознания. Методология определения моральности. Метод наблюдения внешнего поведения. Метод анализа внутренней структуры.

Инфоэтика. Искусственные моральные агенты. Распределенная ответственность. Ответственные машины. Понимающие машины. Вопрос доверия машине.

8 Лекция. ИИ под капотом

История ИИ. Символический ИИ. Субсимволический ИИ. Нейросети. Архитектура трансформера. Встраивание. Многоголовое внимание.
Лингвистический поворот.
От знания о вещах к формулированию знания. Структурализм.
«Второй» лингвистический поворот.
Большие языковые модели. Язык и образы, язык и движение. Языки животных.
От Деантропологизации языка к деантропологизации этики.

9 Лекция. Теории сознания и сознательный ИИ

Нейробиологические теории сознания. Теория рекуррентной обработки (recurrent processing theory). Теория глобального рабочего пространства (global workspace theory). Теория высшего порядка (higher- order theories). Теория схемы внимания (attention schema theory). Теория предиктивной обработки (predictive processing theories). Теория интегрированной информации (integrated information theory). Вычислительный функционализм. Индикаторы феноменального сознания у машин. Этические риски, связанные с недостаточным и чрезмерным приписыванием сознания системам ИИ

10 Лекция. Проект «Просвещение» в эпоху цифровых технологий

Что такое Просвещение?
Руссо Жан-Жак. Рассуждение о науках и искусствах. Denis Diderot Энциклопедия. И.Кант «Beantwortung der Frage: Was ist Aufklärung?». «Диалектика Просвещения» М.Хоркхаймер Т.Адорно. Банальность зла Х. Арендт. Теория коммуникативного действия Ю. Хабермас. Рациональность. Коммуникативная рациональность. Коммуникативный разум.
Цифровое просвещение.
Опасность цифрового «несовершеннолетия». Китай, социальный кредит.
Диалектика цифрового просвещения. Проект Википедия свободная энциклопедия.

Список литературы

Л1

Quante, M. (2011): *Einführung in die Allgemeine Ethik*. 4. Auflage, Darmstadt: WBG
Fenner D. Ethik 2008
Micha H. Werner Einführung in die Ethik 2021
Кононов Е. Метаэтика 2023

Л2

Crimmins, James E., "Jeremy Bentham", The Stanford Encyclopedia of Philosophy (Fall 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.),. https://plato.stanford.edu/entries/bentham/
Macleod, Christopher, "John Stuart Mill", *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/entries/mill/
Сингер, Питер. О вещах действительно важных. Моральные вызовы двадцать первого века [пер. с англ. Е. Фотьяновой]»: Синдбад; Москва; 2019
Практическая философия и новая этика Питера Сингера Кротовская Н.Г.

http://publishing-vak.ru/file/archive-philosophy-2021-5/d4-krotovskaya.pdf
Л3
Кант И.Основы метафизики нравственности
https://royallib.com/book/kant_i/osnovi_metafiziki_nravstvennosti.html
Kant, Immanuel. Grundlegung zur Metaphysik der Sitten.
Кант И. Критика практического разума
https://web.archive.org/web/20070302080909/http://www.philosophy.ru/library/kant/02/0.html Kant, Immanuel. Kritik der praktischen Vernunft.
Л4
Аристотель Никома́хова э́тика. 1997 Издательство: ЗАО " «ЭКСМО-Пресс»
https://avidreaders.ru/book/nikomahova-etika.html
https://bookap.info/okolopsy/aristotel_nikomahova_etika/
Quante, M. (2011): *Einführung in die Allgemeine Ethik*. 4. Auflage, Darmstadt: WBG, Kap. VIII, 5.
Л5
Нозик, Роберт. Анархия, государство и утопия = Anarchy, State, and Utopia (1974) / Пер. с англ. Б. Пинскера под ред. Ю. Кузнецова и А. Куряева. — М.: ИРИСЭН, 2008.
https://skepdic.ru/wp-content/uploads/2013/05/4263Nozik.anarchy_state_utopia.pdf
Ролз Дж.Теория справедливости (1995) Перевод с английского В.Целищев, В.Карпович, А.Шевченко Новосибирск: Изд НГУ, 1995.- 532 с.
https://platona.net/load/knigi_po_filosofii/ehtika_i_ehstetika/rolz_dzh_teorija_spravedlivosti_1995/36-1-0-2233
Л6
Шошана Зубофф. ЭПОХА НАДЗОРНОГО КАПИТАЛИЗМА. БИТВА ЗА ЧЕЛОВЕЧЕСКОЕ БУДУЩЕЕ НА НОВЫХ РУБЕЖАХ ВЛАСТИ. Пер. с англ. А.Ф. Васильева; под ред. Я. Охонько и А. Смирнова. – М., 2022. – 784с
Эдвин Тоффлер. МЕТАМОРФОЗЫ ВЛАСТИ. ЗНАНИЕ, БОГАТСТВО И СИЛА НА ПОРОГЕ XXI ВЕКА. Alvin Toffler and Heidi Toffler, 1990 © Перевод. В.В. Белокосков, 2001 https://gtmarket.ru/library/basis/4857
Кори Доктороу. Как разрушить надзорный капитализм. Doctorow C. (2020). How to Destroy ‘Surveillance Capitalism’. https://onezero.medium.com/how-to-destroy-surveillance-capitalism-8135e6744d59
Ричард Талер и Касс Санстейн. Nudge. Архитектура выбора. 2017. https://www.mann-ivanov-ferber.ru/books/nudge/
П.Тиль, Б.Мастерс. От нуля к единице: Как создать стартап, который изменит будущее
Л7
Julian Nida-Rümelin, Nathalie Weidenfeld. Digitaler Humanismus. Eine Ethik für das Zeitalter der Künstlichen Intelligenz Verlag: Piper, München 2018
Mensch, Moral, Maschine. Digitale Ethik, Algorithmen und künstliche Intelligenz Bundesverband Digitale Wirtschaft (BVDW) e.V. Berlin, Februar 2019 https://www.bvdw.org/fileadmin/bvdw/upload/dokumente/BVDW_Digitale_Ethik.pdf
Petra Grimm, Tobias O. Keber, Oliver Zöllner. Digitale Ethik. Leben in vernetzten Welten.
https://itunes.apple.com/WebObjects/MZStore.woa/wa/viewBook?id=0
Sullins, John, "Information Technology and Moral Values", *The Stanford Encyclopedia of Philosophy* (Summer 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/sum2023/entries/it-moral-values/>.
Luciano Floridi. The 4th Revolution: How the Infosphere Is Reshaping Human Reality Oxford University Press; New Edition (2014)
On the morality of artificial agents. Luciano Floridi and J. W. Sanders University of Oxford
https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.16.722&rep=rep1&type=pdf

L. Floridi, Information Ethics: on the theoretical foundations of computer ethics. Ethics and Information Technology, 1(1):37–56, 1999.
L. Floridi, Faultless responsibility: on the nature and allocation of moral responsibility for distributed moral actions Oxford Internet Institute, University of Oxford, 1 St Giles, Oxford OX1 3JS, UK 2016 The Author(s) Published by the Royal Society. All rights reserved.
AI's Dirty Little Secret: The Minefield of Forbidden Knowledge. Navigating Sensitive Knowledge in AI Models David Campbell https://generativeai.pub/ais-dirty-little-secret-the-minefield-of-forbidden-knowledge-c9553f2c1bee
https://www.academia.edu/109016033/Artificial_Intelligence_or_Artificial_Morality?email_work_card=title

Л8

Искусственный интеллект https://www.imbus.de/ki#c18838
Символический ИИ https://www.bigdata-insider.de/was-ist-symbolische-ki-a-a4729b193881395079a102683c2ab673/
Субсимволический ИИ https://www.bigdata-insider.de/was-ist-subsymbolische-ki-a-6005faede043b6898ff51d9365d36391/
Глубокое обучение https://www.bigdata-insider.de/was-ist-deep-learning-a-603129/
Нейронная сеть https://www.bigdata-insider.de/was-ist-ein-neuronales-netz-a-686185/
https://droider.ru/post/iskusstvennyj-intellekt-mashinnoe-obuchenie-nejroseti-glubokoe-obuchenie-razbor-13-03-2022/
ChatGPT für Nicht-Informatiker*innen https://www.youtube.com/watch?v=-c8ogAwX6KI&t=65s
Artikel "Attention is all you need", Google, 2017 https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
Tobias Rees Non-Human Words: On GPT-3 as a Philosophical
Laboratory Daedalus (2022) 151 (2): 168–182. https://doi.org/10.1162/daed_a_01908

Л9

Butlin, R. Long, +16 authors R. VanRullen Consciousness in Artificial Intelligence: Insights from the Science of Consciousness Patrick Published in arXiv.org 17 August 2023
Computer Science, Philosophy, Psychology https://www.semanticscholar.org/reader/25bb684f8b25f05d8c212d8381c25265865a55e4
Seth, A., & Bayne, T. (2022). Theories of consciousness (Version 1). University of Sussex. https://hdl.handle.net/10779/uos.23488103.v1

Л10

Руссо Жан-Жак. Рассуждение о науках и искусствах // Руссо Ж.-Ж. Избр. сочинения в 3 т. Т 1. М., Гос. издательство худ. литературы, 1961. С. 41-64
Encyclopédie, ou Dictionnaire raisonné des sciences, des arts et des métiers http://enccre.academie-sciences.fr/encyclopedie/
Кант, Иммануил Сочинения в шести томах. М., "Мысль", 1966.-(Философ. наследие). Т. 6.- 1966. 743 с.- С.25-36. Ответ на вопрос: Что такое просвещение? 1784. https://anchiktigra.livejournal.com/2549548.html
ТЕОДОР АДОРНО, МАКС ХОРКХАЙМЕР: ДИАЛЕКТИКА ПРОСВЕЩЕНИЯ. ФИЛОСОФСКИЕ ФРАГМЕНТЫ
https://gtmarket.ru/library/basis/5521#contents
Ханна Арендт. Банальность зла: Эйхман в Иерусалиме ISBN: 978-5-9739-0162-2 Год издания: 2008Подробнее на livelib.ru: https://www.livelib.ru/book/1000381147-banalnost-zla-ejhman-v-ierusalime-hanna-arendt
Jürgen Habermas. *Theorie des kommunikativen Handelns.* (Bd. 1: Handlungsrationalität und gesellschaftliche Rationalisierung, Bd. 2: Zur Kritik der funktionalistischen Vernunft), Frankfurt am Main 1981.