

Санкт-Петербургский государственный университет
Институт философии
«Лаборатория цифровой философии»

«Философские аспекты языковых моделей искусственного интеллекта»

7-8 апреля 2025 г.

Дедлайн подачи: 10 марта 2025 г. 23.00

«Лаборатория цифровой философии» – проект кафедры философской антропологии Института философии СПбГУ – объявляет о проведении научно-методологической конференции «**Философские аспекты языковых моделей искусственного интеллекта**».

К участию принимаются только тексты, созданные с помощью нейросетей с актуальными языковыми моделями, таких как GPT от OpenAI, YandexGPT, Bard, GigaChat, ChatSonic, Gerwin, Claude, Gemini и др.

Современные достижения в области искусственного интеллекта не только расширяют границы технических возможностей, но и открывают перед философией новые горизонты осмысления природы сознания, языка, мышления, а также ставят острые социальные и этические вопросы. Проблематика AI Safety, связанная с рисками и безопасностью ИИ, становится ключевой для глобального сообщества, что делает особенно актуальным философское исследование генеративных языковых моделей и их воздействия на общество. Благодаря способности создавать тексты, неотличимые от написанных человеком, и вести осмысленные диалоги, современные языковые модели становятся уникальной платформой для анализа философских, этических и практических вызовов.

ПРЕДЛАГАЕМЫЕ ТЕМЫ ДЛЯ ОБСУЖДЕНИЯ:

Фундаментальная роль философии в создании ИИ:

Как философские подходы и концепции сознания, языка и мышления влияют на разработку искусственного интеллекта?

Может ли философия служить основой для разработки безопасных и этически приемлемых моделей ИИ?

Философские пристрастия ИИ:

Существует ли возможность, что языковые модели развивают определённые философские взгляды или склонности (например, материализм, детерминизм, этический утилитаризм)?

Как избежать привнесения предвзятости в модели, основанной на философских допущениях их разработчиков?

Философия сознания:

Как разработки в области ИИ меняют понимание сознания, личности и интенциональности?

Вопросы симуляции сознания: является ли генеративный ИИ "мыслящим"?

Этика искусственного интеллекта:

Возможность создания моральных агентов на основе ИИ.

Проблемы предвзятости и дискриминации в языковых моделях.

Влияние ИИ на социальную ответственность и принятие решений.

Язык и значение:

Как использование ИИ трансформирует наше понимание языка и семантики?

Вопросы коммуникации между человеком и машиной.

Эпистемология:

Как контент, созданный ИИ, влияет на человеческие убеждения и восприятие истины?

Возможна ли проверка достоверности информации, сгенерированной ИИ?

Метафизика:

Каков онтологический статус языковых моделей?

Их место в системе человеческого знания.

Философия и разработка ИИ:

Какова фундаментальная роль философии в создании искусственного интеллекта?

Могут ли языковые модели иметь философские пристрастия? Если да, то как они формируются и как это влияет на создаваемый контент?

Эстетика и творчество:

Возможности и границы творчества ИИ: могут ли ИИ создавать произведения искусства?

Эстетические аспекты текстов, сгенерированных ИИ, и их влияние на современную культуру.

Практическое применение и риски:

Использование ИИ в науке, образовании, медицине, праве и других сферах.

Проблема безопасности (AI Safety) и предотвращение рисков: как минимизировать угрозы от применения ИИ?

В рамках конференции состоится круглый стол на тему «**Искусственный интеллект и проблема авторства: новые вызовы современной философии**»

Формат проведения: гибридный (очный и дистанционный)

В качестве заявки для участия принимаются тексты **объемом до 10 тыс. зн.** (при оформлении заголовка просим указывать в скобках, какая ИИ-система использовалась), а также **информация об участнике:** ФИО

Страна, город

Организация

Должность, курс (для обучающихся)

Учёная степень (если есть)

Контактный телефон

E-mail

Участники должны подготовить устное выступление с обоснованием и защитой представленных тезисов.

Оформленные тезисы присылать по адресу: submission@digital-philosophy.ru

НАША СИСТЕМА ОЦЕНКИ ТЕКСТОВ, СОЗДАНЫХ ИИ:

Семантическая связность: оценивается логичность, связность и последовательность текста.

Креативность: оценивается оригинальность предложений и подходов, а также соответствие текстов заданной тематике.

Этическая корректность: проверка текста на отсутствие предвзятости, дискриминации или проблемных утверждений.

Техническая реализация: анализ качества работы языковой модели, включая точность фактов и стилистическую грамотность.

Философская глубина: оценивается, насколько текст раскрывает философскую проблематику и предлагает интересные вопросы для обсуждения.

Лучшие тексты будут опубликованы на сайте проекта «Лаборатория цифровой философии»:

digital-philosophy.ru

Мероприятие проводится в рамках проекта РНФ №24-28-00562 «Философские основания этических рисков в сфере систем искусственного интеллекта»