Институт философии РАН
Российская ассоциация искусственного интеллекта
Rафедра этики философского факультета МГУ имени М. В. Ломоносова

Семинар «Этические проблемы искусственного интеллекта»
5 февраля 2025 г., 17.00 (Институт философии РАН, ауд. 415)

Перов Вадим Юрьевич
кандидат философских, доцент, заведующий кафедрой этики института философии Санкт-Петербургского государственного университета

Проблема «искусственного зла»

Аннотация

Появление новых и особых сфер знания и практик с необходимостью ставит вопрос о критическом переосмыслении соответствующего понятийно-категориального аппарата. Не является исключением и этика в сфере искусственного интеллекта, развитие которой не оставило без внимания базовые этические понятия добра и зла. Обсуждение проблемы зла в современных этических теориях обычно начинается со ставшего традиционным рассмотрения видов зла: моральное зло (зависящее от деятельности людей, их свободы, сознательности и т.д.) и физическое зло (природные явления, физиология людей и т.д.). Появление технологий, основанных на алгоритмах искусственного интеллекта, создание автономных интеллектуальных систем и проекты создания на их основе Искусственных Моральных Агентов обусловили возникновение идей существования в качестве самостоятельного «искусственного зла». В докладе анализируются возможные способы понимания и интерпретации концепта «искусственное зло». Особое внимание уделено проблемам, связанным с возможностью алгоритмам искусственного интеллекта быть «злыми» в строгом «аморальном» смысле. Показано, что Искусственный Моральный Агент не способен отклоняться от алгоритмически правильного поведения и нарушать «правила добра», поэтому «искусственное зло» может рассматриваться только как метафора.