

Теоретические и методологические предпосылки обсуждения проблем этики искусственного интеллекта

Р.Г. Апресян
доктор философский наук, профессор,
руководитель сектора этики Института философии РАН

Резюме выступления

1. На поверхности общественных дискуссий – этические аспекты активности автономных интеллектуальных технических систем (АИТС): насколько принимаемые АИТС решения соответствуют нормам этики (морали), кто и в какой форме несет за них ответственность? В значительной мере это перспективные проблемы, что не отменяет их значимости. АИТС – технические системы, способные самостоятельно (без оперативного контроля со стороны) решать поставленные задачи в условиях меняющихся внешних параметров функционирования. Степень самостоятельности, разнообразия посильных задач и изменчивости параметров, требующих адаптации, может быть различной.

2. Этическая проблематика, связанная с искусственным интеллектом, шире. Она касается не только самого по себе функционирования АИТС и того, как решаются поставленные перед ними и возникающие в ходе их функционирования задачи, но и деятельности по разработке такого рода систем, их подготовке к работе и их обслуживанию, а также социальных, гуманитарных, экологических и возможных других эффектов их использования. Этическим проблемам такого рода, хотя они в настоящее время более насущны, не уделяется достаточного внимания. Между тем, в более общем плане такие проблемы обсуждаются по поводу других практик – внедрения нанотехнологий, природо-сопряженной активности, освоения космического пространства, научных исследований, опосредованных экспериментами на людях, медицинской деятельности и т.д. Есть немало соответствующих нормативных разработок. Накопленный в этой области опыт может быть использован при обсуждении этических проблем искусственного интеллекта.

3. В дискуссиях по прикладным этическим проблемам, в особенности, когда они оказываются предметом широкого общественного внимания, нередко встречается нестрогое, а то и просто неправильное употребление этических категорий. Они используются по инерции обычной и публицистической речи, без учета традиции из обсуждения в моральной философии и нормативной этике, сложившихся по их поводу морально-философских и нормативно-этических концепций. Эти концепции по разным параметрам различны. Можно придерживаться той или иной концепции. Но одно дело свобода понимания и другое – не-свобода непонимания, а то и безграмотности; последняя, бывает, обнаруживает себя в этико-прикладных высказываниях, порой в самых элементарных формах.

4. Понятие морали вырабатывается в западно-европейской культуре для интегративной репрезентации комплекса нормативных (совершенство, благо, справедливость, равенство, забота и др.) и дескриптивных (императивность, автономия, универсальность и др.) идей. Одна из концептуальных доминант в этой традиции – понимание морали как средства согласования частных позиций или интересов ради блага людей и сообщества. Оно производится с помощью определенных инструментов – ценностей, которые имеют в морали (в отличие, скажем, от искусства) императивную природу. Речь идет о таких ценностях, как, например: непричинение вреда, равенство, соблюдение оправданных интересов других, исполнение обещаний, поддержание договоренностей, соучастие (солидарность, помощь), забота. Ценности вырабатываются на основе опыта коммуни-

кативного и социального взаимодействия и направлены в первую очередь на согласование частных интересов, т.е. предотвращение конфликтов или, если конфликты произошли, их конструктивное разрешение. Важный вопрос, требующий осмысления в перспективе целей нашего исследовательского семинара – в какой мере такое понимание морали релевантно практике использования АИТС?

5. Названные идеи по-разному значимы для обсуждения этики АИТС. Необходимо сохранение специфического содержания понятия этического в приложении к АИТС. И главное здесь: (а) разделение прагматического и этического аспектов деятельности, (б) отделение этических норм от правовых. На практике целесообразность, результативность, эффективность (это ценностные понятия, но из области прагматики деятельности) переплетены с благоволением, справедливостью, заботой (этические, или нравственные ценности), а нравственные и правовые нормы по содержанию сплошь и рядом совпадают. Но на уровне осмысления и критической рефлексии опыта, дифференцированное обсуждение ценностного и нормативного содержания решений и действий очень важно. Без этого невозможно адекватное понимание мотивов и субъективного смысла действий, ответственности, вины и заслуги.

6. Говоря об этических проблемах АИТС, следует учитывать различия между АИТС по характеру их функционирования в зависимости от их программного «начала» – детерминистского или когнитивного. В случае АИТС детерминистского типа, фокус этического внимания неизбежно смещается с АИТС на инженерно-конструкторскую, программно-дизайнерскую деятельность. Здесь актуальна задача приложения общих принципов инженерно-конструкторской деятельности вообще к данной предметно-определенной области. В случае АИТС когнитивного типа задачи этического анализа усложняются. Вполне уместно предположить, что нам не избежать модификации морального мышления в приложении к АИТС.

В истории морального мышления нам известно несколько этапов, обусловленных модификациями, направленными на расширение предметного поля этической релевантности. Последнее – связано с признанием животных в качестве субъектов прав и объектов этической ответственности. Могут ли АИТС стать субъектом прав и объектом этической ответственности?

Отдельный вопрос – как модифицируются представления о морали и составляющие морального кодекса в контексте практики взаимодействия между АИТС и человеком/людьми, тем более взаимодействия различных АИТС между собой? Видоизменяются ли они в зависимости от характера тех практических задач, которые решает АИТС в образовании, медицине (хирургические операции и послеоперационный уход за больным), социальной работе, уходе за домом, в сельскохозяйственной, природоохранной, производственной деятельности, на транспорте, в сфере услуг, в сфере развлечений?

7. Этика как моральная философия – сориентирована на обобщенное и абстрактное понятие морали. Это понятие, обобщенное в том смысле, что не всегда отражает внутреннюю неоднородность морали: (а) ценности/императивы – обязанности – добродетели/качества, (б) агент – реципиент – наблюдатель, (в) перфекционистские – коммуникативные – социальные компоненты/аспекты моральных феноменов и т.д. Оно абстрактное в том смысле, что не соотносено с той или иной практикой.

Поэтому необходима прикладная этика. Вообще, нужны разные прикладные этики, сориентированные на разные виды профессиональной и предметно определенной деятельности, на разные виды социальной практики. В частности, прикладная этика, сориентированная на практику, опосредованную функционированием АИТС. Назовем условно соответствующую прикладную этику «эти-

кой АИТС». Необходимо описание этой практики, включенных в нее агентов, реципиентов, связанных с ней третьих лиц, чтобы понимать, какие интересы затрагиваются этой практикой, какие конфликты возникают или могут возникнуть. Названные выше ценности и соответствующие им императивы должны быть спроецированы на эту практику и конкретизированы (не формализованы) по отношению к тому функционалу, которым наделяются АИТС и, что очень важно, к тем конфликтам, которые возможны в ходе осуществления этого функционала.

Как показывает опыт дискуссий на эти темы, здесь важно воздержаться от фантастики, научной или нет, и сконцентрироваться на известной реальной практике и на том, какой из нее можно извлечь опыт. Только на этой основе можно строить прогностические схемы. Очевидно, что без тесного союза конструкторов, программистов, операторов, с одной стороны, и философов, социальных ученых, с другой, не обойтись. Необходима выработка взаимно понятного (если пока не общего) языка, разделяемых схем мышления. И тогда их взаимодействие и сотрудничество могут стать плодотворными.