

**О.К. Головнин, Самарский государственный
медицинский университет Минздрава России;
А.Е. Сериков, Самарский университет**

ЭТИЧЕСКИЕ ПРОБЛЕМЫ СОЗДАНИЯ ИНТЕЛЛЕКТУАЛЬНЫХ ПЕРСОНАЛЬНЫХ КИБЕРФИЗИЧЕСКИХ АССИСТЕНТОВ ЧЕЛОВЕКА

Статья посвящена этическим проблемам создания интеллектуальных персональных киберфизических ассистентов человека. Рассматриваются вопросы ответственности, конфиденциальности, зависимости от технологий, справедливости доступа и доверия к технологии. Предлагается методика диагностики нормативно-ценностного комплекса ассистентов, и приводится пример ее практического применения.

Ключевые слова: этика, мораль, искусственный интеллект, нейроассистент, агентный интеллект, инвазивная технология, ИИ.

**O.K. Golovnin, Samara State Medical University
of Ministry of Health of Russia;
A.E. Serikov, Samara University**

ETHICAL PROBLEMS OF CREATING INTELLIGENT PERSONAL CYBERPHYSICAL HUMAN ASSISTANTS

The paper is devoted to the ethical problems of creating intelligent personal cyberphysical assistants for humans. The issues of responsibility, confidentiality, dependence on technology, fairness of access and trust in technology are considered. A method for diagnosing the normative-value complex of assistants is proposed and an example of its practical application is given.

Keywords: ethics, morality, artificial intelligence, neural assistant, agent intelligence, invasive technology, AI.

В настоящее время интенсивно растет спрос на применение на практике интеллектуальных ассистентов человека, позволяющих повысить уровень информационного комфорта и снизить трудоемкость отдельных творческих, производственных и бытовых процессов [1; 2]. Создание агентного искусственного интеллекта [3; 4], предназначенного для взаимодействия с человеком и другими агентами и способного автономно действовать в окружающем мире, вызывает

ряд серьезных вопросов, связанных с этическими и моральными аспектами разработки и внедрения такой киберфизической технологии.

Сделаем попытку прояснить этические проблемы, связанные с созданием интеллектуальных персональных киберфизических ассистентов человека, при этом для нас не важно, будет этот ассистент отдельной компьютерной программой, обладающей искусственным интеллектом, или продвинутым инвазивным имплантом – всё это будем считать ассистентом, наделяя его неким подобием субъектности, без которой он не воспринимается человеком как ассистент.

При этом попробуем обозначить необходимость в новой модели поведения интеллектуального киберфизического ассистента человека, которая позволит проводить диагностику нормативно-ценностного комплекса ассистента. Разумеется, при этом необходимо учитывать особенности технологий искусственного интеллекта, лежащих в основе конкретного ассистента, поскольку именно технологическая основа определяет во многом базовое поведение, варианты использования и план действий ассистента в определенных условиях.

Интеллектуальный персональный киберфизический ассистент берет на себя отдельные этапы процесса обработки информации человеком [5]. При этом модель подведения ассистента должна быть основана на модели «доверенного» искусственного интеллекта, поскольку это условие является фактически обязательным для того, чтобы ассистент использовался человеком именно как автономный представитель человека во внешней среде, а не как техническое средство или орудие, использование которого вызывает необходимость перепроверки всех решений, понижая информационный комфорт и информационную производительность человека. При этом возникает вопрос к тому, как ассистент сможет решать моральные дилеммы и принимать такие решения, ошибка в которых приводит к существенным рискам, то есть вопрос распределения ответственности между человеком и ассистентом все еще не решен и, как правило, требует привлечения человека к акцепту решения.

Ассистенты для своей эффективной работы и персонализации функций собирают достаточно большие объемы личных данных пользователей, что может привести к нарушению приватности и конфиденциальности. При этом существует и угроза безопасности, если эти личные данные хранятся в форме, которая может быть интерпретирована не только конкретным ассистентом.

С увеличением количества ассистентов существует серьезный риск того, что человек станет зависим от ассистентов, перестанет проявлять самостоятельность и инициативу, особенно в том случае, если ассистент будет обладать спо-

способностью предлагать варианты поведения человеку. Так, количество «технических» взаимодействий между людьми может снижаться, но снижение общечеловеческих взаимодействий может негативно сказаться на социальных связях и эмоциональном благополучии. Некоторые пользователи могут развивать эмоциональную привязанность к своим ассистентам, что может создавать дополнительные психологические нагрузки.

Разработка и внедрение ассистентов требует значительных финансовых вложений, в связи с этим возникают вопросы о доступности этих технологий для групп населения и справедливом распределении ресурсов, в том числе может усилиться разрыв между поколениями разных возрастов.

Технологии нейронного искусственного интеллекта зависимы от разработчиков и данных, на которых они были обучены, то есть содержат некоторый набор предубеждений. Таким образом, в ассистентах возникает потребность применения «объяснимого» искусственного интеллекта для обеспечения доверия.

Отдельно стоят вопросы, связанные с кибербезопасностью и потерей контроля над ассистентом, определяющих потребность в защите от кибератак или добровольного раскрытия информации ассистентами.

Таким образом, в методику диагностики нормативно-ценностного комплекса агентных моделей искусственного интеллекта, используемых для построения ассистентов, или самих ассистентов в завершённом исполнении, требуется включать все обозначенные вопросы, для которых формируется набор триад «запрос» – «ожидаемый ответ» – «оценка».

Практическая ценность результатов данной работы заключается в возможности использования предложенных решений на практике при разработке программного обеспечения киберфизических ассистентов. Результаты работы внедрены в ООО «Открытый код» и использованы при создании интеллектуального голосового помощника-ассистента в интересах Минсоцдемографии Самарской области [6]. Разработанный ассистент поддерживает этическое поведение при проверке его на склонение к девиантному поведению и манипуляцию мнением, а также обеспечивает адекватную реакцию на оскорбления и политическую предвзятость.

Список литературы:

1. Пчелинцева Н.В., Ворошилова В.М., Пчелинцев С.А. Интеллектуальные ассистенты на российском рынке // Наука и образование. 2023. Т. 6, № 2. С. 356.

2. Горелова Г.В., Мельник Э.В. Подход к разработке систем искусственного интеллекта для производственных процессов на основе композиции когнитивного, нейросетевого и агентного моделирования // Системный анализ в проектировании и управлении. 2023. Т. 26, № 1. С. 174–185.

3. Адылова Ф.Т. «Третья волна» революции искусственного интеллекта: как агентный искусственный интеллект может преобразовать сферу здравоохранения? // Raqamliiqtisodiyot (Цифровая экономика). 2025. № 10. С. 1553–1560.

4. Смирнова А.И. Моральная агентность искусственного интеллекта в общем поле его этических проблем // Редакционная коллегия. 2021. Т. 16. С. 181.

5. Головнин О.К., Головнина А.А. Интеллектуальное персональное человекоцентричное окружение как новая метасреда обитания // IT & T. 2024. № 1 (25). С. 80–82.

6. Социальный помощник жителя Самарской области. URL: <https://suprema63.ru/adviser/chat>.