

Секция «Современные социальные трансформации: факторы, условия, результаты»

### Этические аспекты эмуляции цифровой личности

Научный руководитель – Валеева Эльвира Мавлитовна

*Корнаухов Никита Алексеевич*

*Студент (бакалавр)*

Южно-Уральский государственный университет, Факультет журналистики, Челябинск,  
Россия

*E-mail: korn.2002.04@yandex.ru*

Развитие предшественников полноценного искусственного интеллекта – нейросетей, вызывает дискуссии и вызывает к жизни различные образы будущего и предложения этико- законодательных систем, в рамках которых предполагается функционирование ИИ и нейросетей. Особенно здесь важно развитие этической системы регулирования ИИ в мире и у разных государств, впоследствии закреплённых нормативными актами, которые пришлось на пик популярности больших языковых моделей, особенно на первичный взлёт популярности ChatGPT и Gemini, когда оказалось, что, во-первых, OpenAI не озаботилась подготовкой нейросети к работе с действительными пользователями[1], а во-вторых, что имевшиеся соглашения о разработке ИИ носили большей частью рекомендательный характер, а не императивный.

Вместе с тем, также режим OpenSource, о котором заявляли крупные корпорации, как OpenAI на деле не соблюдался, что потребовало не просто выработки рекомендаций, а давления правительств стран мира, с требованием открыть доступ к архитектуре нейросетей, для возможности их изучения. По прошествии совсем короткого времени, около нескольких месяцев, оказалось, что нейросети, например ChatGPT, обучаясь в процессе работы с пользователями, открыто нарушали законы робототехники Азимова, когда вредили человеку, нарушали Азиломарские соглашения 2017-го года о безопасности и частично положения этого документа нарушаются до сих пор.

Эксцессы с пожеланиями смерти со стороны больших генеративных нейронных сетей пользователям, язык ненависти в общении, и подмена итоговой выдачи, когда пользователю приходил не рецепт кофе, а рецепт мощного яда приковали внимание общества к проблеме необузданности генеративных нейронных сетей и тогда предписания стали обретать характер требований. Особенно в странах, где имелись национальные стратегии по работе с ИИ, при том в документах такого класса отмечается, что должна быть проявлена особая осторожность в работе с ИИ, могущим представлять экзистенциальную опасность для человечества.

Так, по итогу анализа национальных и наднациональных документов и доктрин, (например: Регламент Европейского Союза об искусственном интеллекте), можно сделать вывод, что основные этические ограничения на разработку таковы, (они отличаются от подобных на эмуляцию и запуск цифровых личностей или цифровых идентичностей):

1. Обеспечение конфиденциальности данных пользователей.
2. Гуманизм, со стороны ИИ к пользователям.
3. Система управления рисками при разработке ИИ и строгий технический надзор.
4. Надёжность.
5. Повышение точности и надёжности в работе ИИ.

Таким образом к создателям систем генеративного ИИ предъявляется значительное число требований, с которыми можно ознакомиться по ссылкам, но основные:

1. Безопасность;
2. Надёжность;

3. Гуманность;
4. Конфиденциальность;
5. Точность и надёжность;

А само этическое законодательство и право активно совершенствуется и уточняется для обеспечения стабильной работы ИИ.

Быстрое и эффективное развитие нейросетевых моделей и появление значительных возможностей по инициализации и подготовке цифровых личностей, идентичностей, привело к необходимости обсуждения данного вопроса заинтересованными сторонами. Проблема нейросетевых моделей заключается как в возможности применения их вычислительной мощи против общества и правопорядка, так и в распространении фейков, которые могут использоваться для репутационных ударов или мошенничества о чём говорили в статье Д. Траби и Р. Браун, обращая внимания на развитие нейронных сетей в контексте работы с персональными данными пользователей.

Надо отметить, что подобным вопросом, – а допустимо ли этически воссоздавать копии умерших людей или копии людей живых для реализации прикладных и утилитарных задач рекламного характера или нечто подобного занимались не только Траби и Браун с перечисленными в работе исследователями, но и цифровая философия вообще.

В списке её актуальных вопросов[2], которые также экстраполируются на цифровые личности, зачастую репрезентирующие личность человека в настоящей действительности, лежат:

1. Безопасность данных;
2. Приватность;
3. Справедливость и равенство;
4. Ответственность и принцип честности.

Среди актуальных вопросов, связанных с нейронными сетями, а значит и с цифровыми личностями, а особенно с «ментальными данными» выделяется безусловно, в качестве и этического — приватность и безопасность данных. Выражается это и в сохранении данных от утечек и искажений, как было во время дипфейк атаки, развязанной против президента США Барака Обамы и главы Meta\* Марка Цукерберга, и в недопущении искажений данных при формировании цифрового двойника или цифровой личности[4].

Говоря о приватности, нужно отметить, что Big Data заинтересованы в получении как можно больше массива данных о чём также раньше отмечалось в данной статье, и потому корпорации в сфере IT, нацеливаются на разработку таких моделей, которые могли бы собирать весь массив данных о человеке, для воссоздания личности субъекта и предсказания его рыночного поведения.

Таким образом вопрос конструирования цифровых личностей или возможности их создания сталкивается с рядом вопросов, имеющих корень своего происхождения либо в механизме работы нейронных сетей, либо в межличностном взаимодействии в цифровом пространстве и этичности поведения пользователей, где часть информации может быть утаена либо изменена, ввиду ничтожности ответных санкций.

#### ЗАКЛЮЧЕНИЕ

Однако несмотря на это этическая, регулирующая база для цифровых личностей только в процессе создания и вдобавок к тому, сильно зависит от регуляции разработки ИИ вообще, так что в данном случае скорее уместно говорить о примате этических аспектов разработки цифровых личностей в плоскости этического обоснования разработки мощных нейросетей, более развитых чем GPT-4 от OpenAI.

Вместе с тем же требования и ограничения, налагаемые на разработку цифровых личностей, упираются в этические ограничения, связанные с прозрачностью процесса разработки, безопасностью нейросетей, а значит и цифровых личностей, а также приватностью

данных, с чем сейчас имеются проблемы, в т.ч. у OpenAI.

Таким образом этические аспекты разработки цифровых личностей напрямую интегрированы в ограничения и политику, проводимую в отношении ИИ со стороны стран-участников первого ряда: США, КНР, РФ, ЕС, так и позиции ООН и в устройство будущего «робоправа» — подразделения юридического права, регулирующего работы ИИ и его легальный статус, и возможности.

#### Источники и литература

- 1) 2. Бурнашев Р. Ф. UNIVERSUM: ОБЩЕСТВЕННЫЕ НАУКИ №12 (103), декабрь, 2023 г. Режим доступа: <https://cyberleninka.ru/article/n/filosofskie-aspekty-tsifrovoy-etiki-v-epohu-tehnologicheskogo-progressa/viewer>
- 2) 4. Назаров В. Н. Цифровой двойник как субъект информационной этики. Этическая мысль. 2020. Т. 20 №1. С. 142-154. Режим доступа: <https://cyberleninka.ru/article/n/tsifrovoy-dvoynik-kak-subekt-informatsionnoy-etiki/viewer>