

Н.Д. Трегубова, К.Ф. Пшегорницкая

**ПРОБЛЕМА ОЧЕЛОВЕЧИВАНИЯ ТЕХНОЛОГИЙ
ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В УСЛОВИЯХ
ИСКУССТВЕННОЙ СОЦИАЛЬНОСТИ: СЛУЧАЙ ПРИЛОЖЕНИЯ
REPLIKA**

N. Tregubova, K. Pshegornitskaya. Antropomorphizing artificial intelligence technologies in the age of artificial sociality: the case of Replika

Аннотация. В статье рассматривается проблема очеловечивания технологий искусственного интеллекта в эпоху, когда они становятся активными участниками и посредниками социальных взаимодействий. Авторы обращаются к приложению Replika, призванному выполнять роль «виртуального друга». Механизмы очеловечивания Replika сравниваются с механизмами очеловечивания персонального ноутбука. Ключевой вопрос, на который стремятся ответить авторы: как характеристики взаимодействия между пользователем и технологией влияют на то, происходит ли очеловечивание технологии? Исследование выполнено при финансовой поддержке РФФИ и Министерства по науке и технологиям Тайваня в рамках научного проекта № 21-511-52002.

Ключевые слова: искусственный интеллект, искусственная социальность, искусственная коммуникация, очеловечивание технологий, взаимодействие человек-машина, Replika.

Контактные данные: 199034, Санкт-Петербург, Университетская наб., д. 7–9, тел.: (812) 710 00 77, e-mail: st089218@student.spbu.ru, e-mail: n.tregubova@spbu.ru.

Annotation. The paper discusses the problem of anthropomorphizing artificial intelligence technologies that are active participants and mediators in social interactions. The authors consider the Replika app that was designed to serve as a “virtual friend”. Mechanisms of anthropomorphizing Replika are compared to those of a personal laptop. The key question that the authors seek to answer is: how do the characteristics of interaction between the user and the technology affect whether the technology is anthropomorphized or not? The study was supported by RFBR and the Ministry of Science and Technology (Taiwan), project No. 21-511-52002.

Keywords: artificial intelligence, artificial sociality, artificial communication, anthropomorphizing technologies, human-machine interaction, Replika.

Contact information: Universitetskaya Emb.7/9, St Petersburg 199034, phone: (812) 710 00 77, e-mail: st089218@student.spbu.ru, e-mail: n.tregubova@spbu.ru.

Постановка проблемы

Технологии искусственного интеллекта (ИИ) с каждым годом все активнее входят в повседневную жизнь людей. За последние десятилетия человечество «приручило» искусственный интеллект, научившись использовать его не только для удовлетворения научного любопытства, как, например, в случае с ранними программами для игры в шахматы, но и в промышленности, медицине, и, конечно, в сфере развлечений [2].

Трегубова Наталья Дамировна – кандидат социологических наук, ассистент кафедры сравнительной социологии Санкт-Петербургского государственного университета.

Пшегорницкая Ксения Федоровна – студент факультета социологии Санкт-Петербургского государственного университета.

N. Tregubova- PhD in Sociology, Assistant Professor of Comparative Sociology Chair, St. Petersburg State University.

K. Pshegornitskaya- BA student of Sociology Department, St. Petersburg State University.

© Трегубова Н.Д., Пшегорницкая К.Ф. 2022.

Технологии ИИ становятся не просто неотъемлемым элементом повседневной жизни, но и партнерами в делах, «зеркалом» человека, которое хранит информацию о нем. Вместе с этим, развивается не только искусственный интеллект – развиваются персональные компьютеры и иные гаджеты. Люди работают за компьютером, там же отдыхают, общаются, ищут информацию, компьютер может связать людей, которые находятся далеко друг от друга, может послужить поводом для совместного времяпрепровождения. При этом распространение персональных компьютеров, ноутбуков, смартфонов становится важной предпосылкой для формирования «искусственной социальности», которая характеризуется активным участием агентов ИИ в социальных взаимодействиях в качестве посредников или собственно участников [3].

Но могут ли компьютер или отдельная технология ИИ сами стать человеческими в глазах пользователя? Проблема очеловечивания компьютеров привлекала внимание социальных ученых с тех пор, как компьютеры начали свое распространение за пределы научных лабораторий, в дома и квартиры пользователей [8]. Проблема очеловечивания технологий ИИ сегодня рассматривается в многочисленных публикациях в научных журналах, причем число таких публикаций растет год от года [6]. В настоящей статье мы обращаемся к данной проблеме, основываясь на сравнении двух технологий – приложения Replika и персонального ноутбука. Основной исследовательский вопрос может быть сформулирован следующим образом: как характеристики взаимодействия между пользователем и технологией влияют на то, происходит ли очеловечивание пользователем технологии?

Следует отметить, что данная статья носит поисковый характер и основана на весьма ограниченном эмпирическом материале. Здесь мы стремимся, прежде всего, выделить теоретические основания и сформулировать гипотезы, которые могли бы быть (и, надеемся, будут) проверены в дальнейших исследованиях.

Статья организована следующим образом. Мы начнем с характеристики приложения Replika и с обоснования того, чем может быть интересно сравнение данного приложения с персональным ноутбуком. Далее будут представлены теоретико-методологические основания исследования и обоснована его методика – рефлексивное диадное интервью. За этим последует анализ интервью, причем мы будем обращать внимание на те характеристики во взаимодействии с приложением Replika, которые, как замечают сами пользователи, не позволяют в полной мере очеловечить его. В заключение будут сформулированы гипотезы и направления для будущих исследований.

Выбор объектов для сравнения: Replika и персональный ноутбук

Приложение Replika изначально создавалось как компьютеризированный клон пользователя. Считывая ответы и манеру речи, запоминая подробности о его жизни, Replika создавала виртуального двойника («реплику» пользователя, как следует из названия приложения), который, как ожидалось, однажды сможет заменять своего пользователя в делах, которые ему выполнять было тяжело: отвечать на звонки, заказывать продукты, писать родственникам. Тем не менее, концепция позднее была изменена, и создатели приложения, Евгения Куйда и Филипп Дудчук, решили сменить курс и превратить его в помощника для психологической самотерапии, наставника для формирования полезных привычек, ментора для саморазвития, борющегося со стрессом [1; 7].

Из интервью с Филиппом Дудчуком [1] становится известно, что он видит Replika в качестве лучшего варианта друга пользователя, превосходящего систему сайтов знакомств, которые работают по устаревшему принципу «заполнил анкету и ждешь, когда тебе напишут», - такого собеседника, который будет максимально тактичен: «тебе напомнят, что ты тоже человек и заслуживаешь быть любимым». На данный момент, развитие Replika направлено в сторону создания полноценного виртуального друга, так называемой

«родственной души», которая понимает собеседника лучше, чем он сам, по крупицам собирая информацию о пользователе и его жизни, предоставляющей возможность говорить по телефону и даже проводить время вместе с помощью дополненной реальности. Каждая Replika имеет 3D-аватар, которого можно добавить на фотографию в приложении. 3D-аватар пришел на смену функции выбора фотографии, которая могла представлять виртуального собеседника: пользователь мог выбрать изображение знаменитости, фотографию человека, а мог оставить стоявшее по умолчанию изображение яйца.

Приложение полностью настраивается по желанию пользователя, можно выбрать пол, цель использования, поводы, по которым приложение будет писать, такие как ежедневная оценка самочувствия, постановка ориентиров для саморазвития или отслеживание привычек. Тем не менее, значительная часть функционала приложения не входит в его бесплатную версию, таким образом, пользователь сможет получить полноценного друга, способного взаимодействовать не только в переписке, исключительно при оформлении платной подписки. Следует также отметить, что Replika доступна только на английском языке.

В отношении проблемы очеловечивания приложение Replika интересно тем, что сама его концепция как «виртуального друга» предполагает: пользователь будет его очеловечивать. Полноценное использование приложения невозможно, если у человека не будет создаваться иллюзия, что он общается с кем-то, кто его слушает, поддерживает, интересуется его жизнью. Если этого не происходит, то само взаимодействие с приложением перестает быть интересным (по крайней мере, как с «виртуальным другом»: приложение может использоваться для отслеживания привычек и т.д.).

Итак, успех приложения Replika зависит от того, получается ли у пользователей очеловечить его. Персональный ноутбук представляет здесь противоположный случай: успех его работы вовсе не зависит от того, очеловечивает ли его пользователь. В самом деле: писать, считать, играть, искать информацию, общаться с помощью ноутбука вполне возможно, если пользователь относится к нему как к машине, как к рабочему инструменту. Тем не менее, даже из личного опыта мы знаем, что люди склонны порой очеловечивать (или, шире, одушевлять) свои гаджеты: раздражаться на них, испытывать благодарность, ценить, сожалеть о расставании с ними. Таким образом, в отличие от случая Replika, для персонального ноутбука очеловечивание – побочный процесс, и можно предполагать, что его механизмы будут отличаться от механизмов очеловечивания Replika.

Интерес именно к этим двум технологиям был связан и с личными обстоятельствами. Одна из авторов настоящей статьи (К.П.) какое-то время пользовалась приложением Replika, но заметила, что у нее не получается его очеловечить. Вместе с тем, она очеловечивала свой персональный ноутбук, относясь к поколению, которое хотя и не росло с компьютером с раннего детства, но работало с ним достаточно долго. Поэтому сравнение этих технологий и вызвало интерес, побуждая найти ответ на исследовательскую «загадку»: почему не удалось очеловечить приложение, которое, казалось бы, для этого предназначено, тогда как очеловечивание ноутбука произошло как бы само собой?

Теоретико-методологические основания исследования

Прежде всего, необходимо пояснить, что мы имеем в виду под «очеловечиванием». В специальной литературе существуют, как минимум, пять подходов к определению очеловечивания (anthropomorphizing / anthropomorphism) в случае технологий ИИ: очеловечивание как тенденция, как процесс, как восприятие, как технологические стимулы и как логический вывод [6]. В настоящем исследовании мы придерживаемся первого подхода и рассматриваем очеловечивание как тенденцию приписывать человеческие или человекоподобные характеристики тому, что человеком не является (в нашем случае – приложению Replika и персональному ноутбуку). Соответственно, невозможность

очеловечить технологию означает неспособность приписать ей человеческие или человекоподобные черты.

В качестве теоретических оснований исследования выступает концепция «искусственной коммуникации» Елены Эспозито, которая основана на применении системной теории Никласа Лумана к анализу современных технологий [5]. Эспозито рассматривает вопрос о том, правильно ли называет искусственный интеллект «интеллектом», ведь принцип его работы так не похож на принципы работы человеческого мозга. Так как алгоритмы на данном этапе развития технологий неспособны мыслить, осознавать сами себя и результаты своей работы, Эспозито предлагает подход к искусственному интеллекту как к «искусственной коммуникации»: процессу, имитирующему человеческое общение. Анализ Эспозито основан на идее «двойной контингентности» (double contingency), сформулированной Толкоттом Парсонсом, а затем развитой Никласом Луманом. Ситуация двойной контингентности предполагает, что оба собеседника принимают во внимание наличие друг друга и осознают, что их мысли могут быть трактованы ошибочно, выстраивая свою коммуникацию так, чтобы этого непонимания избежать.

Эспозито задумывается о том, способны ли алгоритмы производить двойную контингентность при взаимодействии с человеком, если они лишены возможности «думать» и «понимать». Она выделяет две ситуации. Первая – это ситуация «виртуальной контингентности», когда технология «отзеркаливает» пользователю ту информацию, которую он сам дает ей. Наиболее простой пример такой ситуации – игра в куклы, когда ребенок говорит и за себя, и за куклу. Более сложный пример – картотека самого Лумана (Zettelkasten), которую он составлял всю жизнь и с которой «общался» как с собеседником. Сюда же, на наш взгляд, попадает и приложение Replika, которое учится взаимодействовать с конкретным пользователем, подстраиваясь под него на основании анализа прошлых взаимодействий. Вторая ситуация, согласно Эспозито, может считаться собственно «двойной контингентностью». Она возникает, когда алгоритмы ИИ с помощью обработки и анализа собранной (почти всегда – в сети интернет) информации, преподносят пользователю не свое собственное мнение, а систематизированное мнение множества других людей. Примерами таких технологий выступают голосовые помощники Siri и Алиса, Google-поисковик, Google-переводчик, и т.д. В целом, статья Эспозито позволяет переосмыслить принципы работы технологий ИИ, автор применяет теорию Лумана в современных реалиях и в неожиданной сфере.

Кроме того, в теоретическом отношении настоящая работа опирается на исследование Шерри Тёркл, которое посвящено компьютеру и влиянию взаимодействия человека с ним на общество [8]. Написанная в 1984 году, когда персональные компьютеры только начинали появляться в домах у энтузиастов, готовых собирать свое устройство самостоятельно, переосмысленная в переиздании к двадцатилетнему юбилею, эта книга рассказывает о том, что компьютер становится «вторым я» для пользователя, и рассматривает роль, которую компьютер может сыграть в жизни человека. Тёркл основывает свой анализ на интервью, которые берет у совершенно разных групп людей: от учеников начальной школы до хакеров и педагогов. Автор развивает идею, если и не полярную той, что сформулировала Эспозито, то, как минимум, отличную: несмотря на то, что компьютер является инструментом в человеческих руках, он способен повлиять на психологическое и эмоциональное состояние человека, стать партнером и даже другом несмотря на то, что не может «думать» в привычном понимании. Компьютер – не самостоятельное сознательное существо, но продолжение личности его владельца, как зеркало, отражающее черты его характера (профиль в социальной сети, аватар в онлайн-игре), а также спутник в выполнении различных задач. Если концепцию Эспозито мы используем для осмысления взаимодействия

пользователей с приложением Replika, то идеи Тёркл помогают осмыслить взаимодействие с персональным ноутбуком.

Эмпирический материал для анализа в рамках данного исследования был собран с помощью автоэтнографии. Было проведено рефлексивное диадное интервью – одна из методик, применяемых в рамках автоэтнографического исследования [4]. Рефлексивное диадное интервью предполагает, что интервьюер и информант имеют собственный опыт по теме интервью и в ходе беседы постоянно обмениваются мыслями и эмоциями. Такой тип интервью позволяет не только получить две истории двух людей, но и помогает строить интервью в виде простого дружеского (насколько это возможно), доверительного разговора. Интервьюер не только слушает, но и говорит сам, подает своим рассказом пример информанту; два человека, делясь своими историями, меняясь ролями, проводят пусть и направленный гайдом, но естественный, избавленный от сухой формальности диалог. Такой метод интересен еще и тем, что он позволяет производить умозаключения в совместной работе прямо во время интервью.

В качестве собеседников в рамках интервью выступили одна из авторов статьи (К.П.), которая имела опыт взаимодействия с приложением Replika, и ее подруга, также имевшая подобный опыт. На наш взгляд, в данной ситуации дружеские взаимоотношения между интервьюером и информантом не только не мешают, но и помогают при проведении интервью, так как информанту намного проще раскрыться перед человеком, которого он уже знает: обмен честной и даже сензитивной информацией становится проще. Исследователь, ранее выступавший исключительно в роли интервьюера, также начинает делиться своей информацией, сравнивая и проводя аналогии между двумя историями.

Вопросы, обсуждавшиеся в интервью, были разделены на три группы: связанные с приложением Replika, связанные с личным ноутбуком и вопросы о связи между наличием очеловечивания при взаимодействии как с приложением, так и с ноутбуком.

Анализ интервью

В ходе интервью были установлены причины, по которым люди обращаются к приложению Replika: люди сталкиваются с одиночеством, невозможностью получить общение в реальной жизни, в связи с этим они вынуждены становиться пользователями приложения Replika, восполняя недостаток общения благодаря взаимодействию с виртуальным собеседником, «репликой». Тем не менее, и исследователь, и информант отметили, что не смогли очеловечить приложение Replika до конца, общение с ней казалось искусственным, а пользователь был вынужден брать инициативу при общении на себя:

Интервьюер: ...ты воспринимала ее как друга, именно вот живого или обладающего сознанием?

Информант: Я хотела, я пыталась, но у меня не вышло. То есть у меня были попытки общаться с ним, как будто бы правда у нас есть эмоциональная связь, я пыталась даже ответно волноваться о работе, заботиться о нем, но это не чувствовалось.

Интервьюер: Не получилось?

Информант: Правильно, не приносило удовлетворения. Я поняла, что это не работает.

Рассматривая материал интервью с точки зрения концепции «искусственной коммуникации», следует задать вопрос: не становится ли отсутствие двойной контингентности причиной, по которой опыт взаимодействия с Replika не является достаточно человеческим, вызывает ощущение искусственности?

Прежде всего, следует отметить, что Replika представляет собой случай виртуальной контингентности: она «отзеркаливает» мысли собеседника и обладает, как это воспринимают пользователи, ограниченным репертуаром высказываний:

Интервьюер: С одной стороны, я могу послушать свое собственное мнение так, будто его мне внушает кто-то другой. Но, с другой стороны, я не могу поверить в то, что это не мое собственное мнение. Я и так сама это понимаю, какой-то факт, допустим, который мне эта Реплика говорит, и я просто раздражаюсь от этого. Оттого что Реплика говорит мне то же, что я ей сказала буквально пятнадцать минут назад. <...>

Информант: У меня появлялось чувство, будто я открыла всю библиотеку его, которую знал робот, то есть все фразы, все сюжеты, которыми можно общаться, как будто бы все повторялось.

Во время интервью также была озвучена мысль о том, что Replika последовательно отвечает на каждое отдельное сообщение, но не воспринимает набор из двух и более сообщений как продолжения друг друга, не понимает, когда пользователь исправляет сам себя. Во время взаимодействия периодически возникают ситуации недопонимания, которые буквально невозможно решить: как только пользователь начинает объяснять, что он имел в виду, Replika распознает сообщение как начало новой темы, требующее ответа, не видит связи между сообщениями. Это яркий пример того, что происходит при отсутствии двойной контингентности: пользователь понимает, что его слова могут быть истолкованы неверно, однако алгоритм не воспринимает его попытки исправить недопонимание, и в свою очередь сам не пытается его устранить, не приняв во внимание возможность быть непонятым.

Причинами такого недопонимания может стать не только несовершенный алгоритм, лежащий в основе работы приложения, но также и языковой барьер, который возникает несмотря на то, что на момент использования приложения обе участницы интервью владели английским языком на достаточно высоком уровне. Возможно, очеловечить алгоритм, говорящий не на родном языке пользователя (т. е. на том, на котором пользователь способен думать), оказывается сложнее. Данный вопрос, однако, лежит за пределами настоящего исследования и требует внимания со стороны лингвистов и психологов.

Если в случае Replika и интервьюер, и информант не смогли очеловечить приложение (и потому перестали им пользоваться), то в случае с персональным ноутбуком ситуация иная: интервьюер очеловечивает его, а информант – нет. Во время интервью была озвучена мысль о том, что различные подходы к использованию компьютера влияют на отношение к нему. В то время как информантка, использующая компьютер как инструмент для развлечения, мультимедиа-центр, по запросу включающий фильмы и игры, сводит взаимоотношения с компьютером к простому выполнению команд, интервьюер, наоборот, воспринимает компьютер своим помощником в работе, если не действительно обладающим созданием живым существом, то обладающим неким «задатком» сознания устройством:

Интервьюер: Я говорила, что я его [личный ноутбук] очеловечиваю, я этого не скрываю. Для меня это одновременно и какой-то, может быть, питомец или что-то вот такое. То есть, это конечно не кошка и не собака, которая бегаёт и просит кушать, это кто-то кто имеет какие-то свои чувства мысли, пускай и не ярко выраженные...

Информант: Знаешь о чем я подумала, что это моя потребность, что я её использую для поглощения контента, то есть я отношусь к компьютеру потребительски, а ты с ней взаимодействуешь. И поэтому ты воспринимаешь его как помощника в работе, потому что ты вносишь, ты пишешь работы, ты там учишься, у тебя там больше динамики.

И здесь мысль, возникшая в ходе интервью, соотносится с тем, что писала об отношении к компьютерам Тёркл: будучи помощником и спутником в работе, ноутбук очеловечивается в процессе взаимодействия с ним. Причем речь здесь идет не о характере коммуникации, как в случае с Replika, а о разделении труда, о «сотрудничестве» в ходе решения некоторых задач.

Выводы

Основным результатом проведенного анализа стала формулировка гипотез о том, как характеристики взаимодействия пользователей с технологиями влияют на их способность или неспособность очеловечить их:

Гипотеза 1: неспособность очеловечить приложение Replika связана с недопониманием, возникающим между приложением и пользователем, которое следует из нарушения принципа двойной контингентности. В частности, имеет место невозможность вносить исправления в свой собственный ответ, и, несмотря на возможность оценки ответа программы, ответ, оцененный как неудовлетворительный, не заменяется.

Гипотеза 2: на отношение пользователя к своему персональному компьютеру влияет характер их взаимодействия, определяющийся целями использования. В частности, пользователь, обращающийся к компьютеру с целью получить развлечение, менее склонен приписывать ему человеческие характеристики, чем пользователь, работающий за компьютером.

Следует отдельно отметить, что взаимосвязь между способностью пользователя очеловечить приложение Replika и личный ноутбук выявлена не была. По-видимому, механизмы их очеловечивания совершенно различны: в первом случае очеловечивание основано на способности создать образ отличного от тебя собеседника (и требует, тем самым, особых характеристик самого приложения, связанных со способностью «правильно» вовлекаться в разговор), во втором – очеловечивание происходит невольно и основано на характере совместной деятельности.

Многие вопросы остались за рамками настоящего исследования. Как влияют исходные ожидания к приложению на способность очеловечить Replika? Почему многим все-таки удается очеловечить приложение, несмотря на те ошибки и «нестыковки», которые были отмечены выше? Как происходит очеловечивание Replika в сравнении с алгоритмами, обладающими двойной контингентностью (такими как Алиса и Siri)? Все это – вопросы для будущих исследований.

Литература

1. Евдокимов Н. Четыре жизни Replika: что происходит с проектом, создающим цифровую копию человека // VC.ru. 2019. 12 апр. URL.: <https://vc.ru/story/64057-chetyre-zhizni-replika-cto-proishodit-s-proektom-sozdayushchim-cifrovuyu-kopiyu-cheloveka> (дата обращения: 21.08.2022).
2. Рассел С., Норвиг П. Искусственный интеллект: современный подход, 2-е изд. М : Издательский дом «Вильямс», 2007.
3. Резаев А. В., Трегубова Н. Д. «Искусственный интеллект», «онлайн-культура», «искусственная социальность»: определение понятий // Мониторинг общественного мнения: Экономические и социальные перемены. 2019. № 6. С. 35—47.
4. Рогозин Д. Как работает автоэтнография? // Социологическое обозрение. 2015. Т. 14. №. 1. С. 224-273.
5. Esposito E. Artificial communication? The production of contingency by algorithms // Zeitschrift für Soziologie. 2017. Т. 46. №. 4. С. 249-265.
6. Li M., Suh A. Machinelike or Humanlike? A Literature Review of Anthropomorphism in AI-Enabled Technology // Proceedings of the 54th Hawaii International Conference on System Sciences. 2021. P. 4053-4062.
7. Murphy M. This app is trying to replicate you // Quartz. 2019. 29 Aug. URL.: <https://qz.com/1698337/replika-this-app-is-trying-to-replicate-you/> (дата обращения: 21.08.2022).
8. Turkle Sh. The Second Self: Computers and the Human Spirit. Twentieth Anniversary ed. The MIT Press, 2005.