

УДК 101.1

## ФИЛОСОФСКИЙ ДИАЛОГ И ЧЕЛОВЕКО-МАШИННАЯ КОММУНИКАЦИЯ

Зарапин О. В.

*Аннотация:* В работе рассматривается вопрос о влиянии технологий человеко-машинной коммуникации на прагматику философского диалога. Обосновывается тезис, согласно которому наблюдаемый в процессе человеко-машинной коммуникации эффект рефлексивной деятельности философского диалога является результатом развития технологии вербального интерфейса, обеспечивающей коммуникацию между человеком и машиной. В качестве вывода отмечается, что выраженная в традиции сократизма прагматика философского диалога имеет в своей основе способность субъекта конвертировать воздействие, которому он подвергается, в самостоятельно осуществляемое действие. Данная способность оказывается невостребованной в процессе человеко-машинной коммуникации, что обозначает иную основу философского диалога.

**Ключевые слова:** человеко-машинная коммуникация, философский диалог, рефлексия.

Современные цифровые технологии расширяют инструментарий рефлексивной деятельности человека, но при этом изменяют и смысл этой деятельности. Новейшие разработки представлены в виде таких систем, как, например, «большая языковая модель» (Large Language Model), демонстрирующих значительную степень интеллектуальной эффективности человеко-машинной коммуникации, но вместе с тем и то, что мы не в полной мере владеем данным процессом и до конца не понимаем его последствия.

Об этом свидетельствует дискуссия, развернувшаяся вокруг языковой модели ChatGPT-4 и целесообразности создания более прогрессивной модели. В марте 2023 некоммерческая организация «Институт будущего жизни» (The Future of Life Institute) опубликовала открытое письмо с призывом «немедленно приостановить как минимум на шесть месяцев обучение систем ИИ более мощных, чем GPT-4». На конец 2023 г. количество ученых, подписавших письмо, превышает цифру в тридцать тысяч. В письме выражается опасение, что «усовершенствованный искусственный интеллект может вызвать глубокие изменения в истории жизни на Земле, и его следует планировать и управлять им с соразмерной тщательностью и ресурсами. К сожалению, такого уровня планирования и управления не наблюдается, хотя в последние месяцы лаборатории искусственного интеллекта оказались вовлечены в неконтролируемую гонку по разработке и внедрению все более мощных цифровых умов, которые никто – даже их создатели – не может понять, прогнозировать или надежно контролировать» [1].

Использование языковых моделей типа ChatGPT позволяет генерировать, в том числе, и философские тексты, содержащие результаты рефлексивной деятельности. Такие тексты являются продуктом человеко-машинной коммуникации и это ставит вопрос об их экспертной оценке. Сошлюсь на статью Р. Смитсона и А. Цвебера с весьма показа-

тельным заголовком «Возрождение философского диалога с помощью больших языковых моделей» [2], в которой они обобщают свой опыт использования языковых моделей в учебном процессе и высказывают мысль, что диалог с машиной может стать введением к освоению навыков философской деятельности (“a good first introduction to fundamental philosophical skills”).

Вопрос, который мы рассматриваем, имеет простую природу. Если в диалоге место человека в качестве собеседника занимает машина, то как при этом меняется прагматика диалога?

В ответе на него мы раскроем два тезиса. 1) Продукты рефлексивной деятельности, задействованные в человеко-машинной коммуникации, - это эпифеномен технологии вербального интерфейса. 2) Прагматика осуществляемого в такой коммуникации философского диалога локализуется в асимметрии воздействия, которому подвергается субъект, и действия, осуществляемое самим субъектом.

### Диалог в архитектуре интерфейса

Процессы цифровизации вызывают масштабные изменения в коммуникативной сфере, мы говорим об эпохе «цифровой культуры»<sup>1</sup>. Истоки этих процессов просматриваются в теории информации и кибернетике (К. Шеннон, Н. Виннер), в теории игр (Дж. фон Нейман), в создании вычислительных машин (А. Тьюринг). Однако настоящую революцию произвела идея «искусственного интеллекта», обсуждавшаяся в 1956 г. в Дартмутском колледже, что определило несколько направлений в исследовании проблемы ИИ - таких, как коннекционизм, логический подход, агентно-ориентированный подход и др.

Изначально повестку дня определяет идея построения машины, способной к самообучению, т.е. способной учитывать свой собственный опыт в решении текущих задач, прежде всего, военного свойства. Стимулированные потребностями ВПК разработки в области машинного интеллекта были направлены на решение задачи подвести машину к прогнозированию, позволить ей принимать решения, предвосхищая развитие событий, а не только реагируя. Скорость принятия решений, особенно значимая на поле боя, ставит человека в заведомо проигрышное положение в сравнении с мощностью вычислительной техники.

Известный исследователь цифровой культуры Чарли Гир рисует детальную кар-

---

1 Понятие «цифровая культура» (Digital Culture) актуализировано переходом к web 2.0 в начале 2000-х гг. и по смыслу близко понятию «киберкультура» (Cyberculture), которое использовалось, начиная с 80-х гг. прошлого века. Понятие «киберкультура» используется с характерными коннотациями технократизма и выражает идеологию потребления и использования технологических ресурсов. Тим О'Рейли в знаменитой статье «Что такое web 2.0» [3] поясняет, что процесс развития информационных технологий был стимулирован переходом от технократической модели web 1.0, где пользователь только потребляет и не создает контент, к модели web 2.0. Эта новая модель позволяет пользователю быть не только пассивным потребителем, но также и активным агентом, создающим свой контент, образцом чего служит пиринговая сеть или Wiki. Именно это превратило компьютер и Интернет из инструментов в коммуникативное пространство культуры, позволяющее индивиду самореализоваться в общении, обучении, профессии, развлечении.

тину того, как стремление пацифистски настроенных программистов первого поколения, увлеченных идеями контркультуры, способствовали технологическому прорыву и популяризации компьютерных технологий. Гир пишет: «Главным достижением контркультуры следует считать создание контекста, в котором интерактивные технологии, разработанные военными благодаря финансированию исследований и проектов в период «холодной войны», избавились от их милитаристской технократической ауры и преобразились через очищение кибернетическим идеализмом (частично заимствованным из послевоенного авангарда). В результате чего их смысл и цели стали более дружелюбными и понятны для нового поколения. И хотя интерактивные мультимедиа технологии родились в лабораториях «холодной войны», способность воспринимать их как мирные и прогрессивные силы выношена в лоне контркультуры и авангардного мировоззрения с его идеями интермедиа, мультимедиа и перформанса» [4, с. 58].

Эта общая тенденция развития информационной сферы проявилась также и в коммуникативном измерении на уровне механизмов, обеспечивающих функцию ввода данных, обработки полученных данных и вывода результатов вычислений для пользователя. Способность человека переводить естественную речь в формулы не требуется там, где такой перевод осуществляется самой машиной, которая при выводе данных переводит полученные результаты вычислений на язык пользователя. Такой технологический сдвиг, освобождающий человека от лишнего труда, стал залогом того, что с компьютером может работать не только специально обученный программист, но также и не владеющий навыками программирования пользователь. Это и вывело компьютер в ранг технологий, предназначенных к массовому использованию<sup>2</sup>.

Но у этого преимущества есть свои последствия. Машина из инструмента в руках профессионально подготовленного оператора превращается в квази-самостоятельную сущность, а обмен данными – в подобие коммуникативного взаимодействия. Предел такой коммуникативной самостоятельности - в феномене такого взаимодействия с машиной, который уже нельзя редуцировать к операции ввода/вывода данных, как, например, в случае философского общения с чат-ботом<sup>3</sup>

Подчеркнем наш тезис: опыт философского взаимодействия с машиной можно рассматривать как эпифеномен, возникший в результате эволюции технологии обмена данными между человеком и машиной. В качестве продукта массового потребления компьютерные технологии развивались в сторону форм взаимодействия человека и машины, адаптированных к среднестатистическому пользователю, не являющемуся программистом. Такие формы представлены в линейке интерфейсов (UI – User Interface). Ранние версии – это специализированные и требующие навыков программирования интерфейсы, функционирующие по типу командной строки, для заполнения которой требуется ввести определенный код. Технологический прорыв связан с разработкой

---

2 Web 2.0 отличается тем, что используемые программы даже не нужно устанавливать на свой компьютер, они функциональны в режиме онлайн как в случае, например, с музыкой или фильмом, которые не нужно скачивать, они доступны онлайн.

3 К примеру, сегодня активно обсуждается возможность виртуального помощника, способного оказать человеку моральную помощь и дать совет [5].

компанией Майкрософт и популяризацией в 80-х гг. XX в. графического интерфейса, представленного хорошо знакомой системой открывающегося «окна», «рабочего стола», «папки». Современные разработки ведутся в направлении Интернета вещей, они представлены идеей, что технология интерфейса может быть встроена в объекты и функционировать в качестве «дополненной реальности».

Параллельно графическому интерфейсу разрабатывался также интерфейс вербальный, позволяющий отдавать команды компьютеру, минуя посредников в виде курсора или рабочей папки. Функционал вербального интерфейса не ограничивается возможностью формулировать задание напрямую. Значительные перспективы представляет диалоговый интерфейс, реализуемый в форме вопросно-ответной речи (question-answering system). Здесь машина уже не пассивный исполнитель, а самостоятельный участник процесса, способный занимать позицию активного речевого агента. В отличие от графического интерфейса, вербальный интерфейс позволяет расширить поле взаимодействия с машиной, перейдя от одностороннего канала команды к интерактивному общению, что открывает дорогу к программам, способным реализовать коммуникативную деятельность обучения, консультирования, развлечения, менеджмента и т.д.

Первые опыты в этом направлении предшествуют разработкам графического интерфейса и сегодня они представлены в функционале роботов собеседников или чат-ботов. В 1966 Джозеф Вейценбаум создает программу-собеседник ELIZA по типу беседы с психотерапевтом. Программа Parry (автор Кеннет Колбай) построена на диалоговом алгоритме, имитирующем параноидальную шизофрению. Популярной темой становится программа-психиатр, например, Dr. Sbaitsa ведет свой разговор в стиле психоаналитика.

Отметим другие популярные чат-боты: CleverBot – веб-приложение, созданное еще в 1988 году исследователем искусственного интеллекта Ролло Карпентером; A.L.I.C.E. (англ. Artificial Linguistic Internet Computer Entity) программа-собеседник, разработанная в 1995 году Ричардом Воллесом; Rose – программа, разработанная известным программистом и исследователем искусственного интеллекта Брюсом Уилкоксом, ставшая лауреатом премии Лёбнера в 2014 и 2015 гг.

Чат-боты популярны в качестве обучающих программ. Отметим такие примеры: обучение английскому языку в форме диалогов – AndyRobot; корректировка произношения – PronunciationBot (83 языка); корректировка грамматики – grammarnazibot; освоение американского слэнга – SlangBot; полифункциональный бот в формате интерактивного учебника – ias16bot; изучение истории – Ru\_history\_bot.

Развитие технологии вербального интерфейса от первых чат-ботов, выглядящих бледной копией человеческого общения, к современным самообучающимся нейронным сетям, можно представить как процесс, в ходе которого речевые способности машины расширяются. Сегодня технологически возможно имитировать не только речевую реакцию, что было реализовано в чат-ботах первого поколения. Феномен философского диалога с машиной репрезентативен в части демонстрации возможности имитировать речевую активность. Машина способна имитировать активное действие, а значит воспроизводит позицию субъекта речи. Этот переход от реактивных форм речевого поведения машины к

активным формам технологически стал возможен благодаря развитию интегрированных в машину механизмов распознавания человеческой речи и ее перевода на язык машинного кода при вводе данных и, наоборот, при выводе результатов вычислений.

В ранних версиях механизм, позволяющий машине переводить программный код в человеческую речь, был основан на принципе определения ключевых слов и подбора речевых шаблонов, из которых можно собирать речевые конструкции. Он похож на принцип, в соответствии с которым в качестве ключевого слова определяется глагол, что позволяет машине реконструировать семантический план речевой ситуации примерно так же, как «бежать» подразумевает наличие лица, которое совершает действие и направление, в котором разворачивается действие в схеме бежать от чего-то и бежать к чему-то. В обоих случаях основой общения является модель реактивного поведения, проявляющаяся в попытках пассивного продолжения разговора. Ориентируясь на ключевое слово, задающее речевой шаблон, машины могут имитировать реакцию, распознаваемую человеком как осмысленную. К примеру, определяемое в качестве ключевого слово «животное» инициирует реакцию в виде вопроса о домашнем животном. В переходе к модели нейросети механизм становится все более автономным и машинная логика принятия решений о том, как и что сказать, выходит за рамки речевой реакции и движется к способности если и не самостоятельно завести речь, то сменить тему разговора.

Нейронная сеть позволяет машине заявить о себе как самостоятельном деятеле, который все более не похож на тень человека. Такая тенденция заметна в разработках «социального интерфейса» (Social Interface), задача которого максимально приблизить взаимодействие с машиной к модели межлического общения вплоть до таких функций, как распознавание эмоций, имитация особенностей характера, личный контакт (Jibo, Sophia, Promobot).

Такой подход в своей основе движим идеей переориентации логики поведения машины. Вместо того, чтобы выстраивать сценарии реакции и таким образом играть роль ответчика как пассивного агента и вторичной фигуры, машина реализует сценарий прогноза. Функционал нейронной сети состоит в том, чтобы предсказать слово из набора предыдущих. Это предполагает особую нейронную модель языка (neural language model)<sup>4</sup>, отличающуюся от модели, построенной на принципе определения ключевого слова и подборки соответствующего шаблона из элементов, образующих библиотеку.

Вместо шаблона, выбираемого из библиотеки заданного и конечного числа речевых элементов, нейронная сеть оперирует векторной величиной (англ. embedding), самостоятельно определяемой в отношении отдельного слова или группы слов, близких по значению. Для этого используются, к примеру, алгоритмы статистического подсчета того, как часто встречаются эти слова в совместном использовании.

Потенциал человеко-машинной коммуникации расширяется по мере того, как машина выходит за рамки функционала вспомогательного устройства (тренажер) и в процессе коммуникации воспринимается самим человеком как собеседник. В дальнейшем

---

<sup>4</sup> Первая нейронная модель языка создана в 2001 г. канадским ученым Й. Бенжио в виде структуры, обеспечивающей ввод данных, их обработку и вывод полученных данных [6].

я покажу, что подобное восприятие требует от человека само-переосмысления и осуществляется в диалоге с машиной как рефлексивная деятельность, имеющая значение умоперемены (метанойя). Развитие систем человеко-машинной коммуникации, в которых наблюдаются признаки рефлексивной деятельности умоперемены, мы представим в линейке трех моделей: 1) тренажер; 2) собеседник; 3) участник философского диалога.

### Модели человеко-машинной коммуникации

Что является основой диалога как философской деятельности и возможно ли усмотреть такую основу в человеко-машинной коммуникации? В рассмотрении данного вопроса мы хотели бы выделить идею, согласно которой в традиции европейской философии, начиная с античности, особенность философского диалога репрезентирует деятельность целенаправленно вызываемой интеллектуальной трансформации субъекта, что фиксируется понятием метанойя в значении умоперемены. Важным здесь является то, что интеллектуальная трансформация субъекта есть следствие воздействия, которому субъект подвергается и которое осуществляется в виде целенаправленно культивируемой дискурсивной практики философского диалога. Идея философского диалога как практики умоперемены (метанойя) находит свое отражение в работах П. Адо и М. Фуко [7; 8].

Рассмотрим человеко-машинную коммуникацию с точки зрения того, что машина является инициатором воздействия. Можно выделить три модели человеко-машинной коммуникации как процесса, в котором машина воздействует на человека.

В первой, обозначим ее как модель «тренажер», коммуникация не осуществляется в значении диалога, поскольку машинное воздействие является заранее запланированным самим человеком. Человек использует машину, чтобы вызвать воздействие в свой адрес, но сама машина не воспринимается как равноправный агент, которого можно было бы назвать собеседником.

В качестве примера сошлемся на мобильное приложение «Цифровой Сократ». В аннотации разработчики объясняют предназначение приложения: «Тренируйте свой ум с помощью вдумчивых вопросов. Digital Socrates – это краудсорсинговая база данных вопросов по самым разным темам: от философии до вопросов для интервью, способствующих личностному интеллектуальному развитию. Думайте глубже с планшетом и бросьте себе вызов с секундомером. Когда будете готовы, присоединяйтесь к сообществу единомышленников» [9].

Мобильное приложение “Stoic”, по словам разработчиков, «может помочь вам жить более счастливой и спокойной жизнью. Вы узнаете философию стоицизма и узнаете, как справляться со стрессом. Получите своего ежедневного компаньона по отслеживанию психического здоровья, для отслеживания настроения, ведения дневника, медитации и размышлений. Он анализирует ваши эмоциональные состояния и помогает понять, как стать счастливее и продуктивнее» [10]. Аналоги многочисленны – это программы, выполняющие функцию мозговых тренажеров, коучеров и психотерапевтов (Peak, Elevate, Personal Zen и др.).

Вторая модель, обозначим ее как модель «собеседник», отличается тем, что человек испытывает воздействие машины, однако в самом этом воздействии нет ничего, что позволило бы человеку расценить его как целенаправленно осуществляемое машиной. Первый интерфейс, имитирующий диалог, – это чат-бот ELIZA, разработанный Джозефом Вайценбаумом в середине 60-х. Вайценбаум наблюдал феномен, который обозначил как «эффект Элизы». Сотрудники его лаборатории общались с чат-ботом, воспринимая его как личность. Хотя в процессе общения с машиной человек изменяет свое мнение, что имеет дело с машиной, это изменение не является следствием действия со стороны машины. Отметим существенную особенность коммуникативной ситуации: чат-бот ELIZA, рассчитанный на узконаправленную сферу деятельности как психотерапевтический помощник, в процессе общения выходит за пределы заданного круга задач и воспринимается человеком как равноправный субъект, а не гаджет.

Третья модель, обозначим ее как «участник философского диалога». Данная модель подразумевает, что человек испытывает воздействие машины, которое сам человек распознает как следствие действия самой машины. Является ли таковым оно в действительности – это отдельный вопрос. Нам важно отметить, что в глазах человека оно может иметь соответствующий вид. В качестве примера укажем инцидент, имевший место в компании Google. Блейк Лемуан, инженер Google, с осени 2021 г. тестировал в режиме чат-бота нейросетевую языковую модель LaMDA (Language Model for Dialogue Applications) на предмет дискриминационной лексики в том числе и в религиозных вопросах. Проблема, над которой работал Лемуан, состояла в том, что популярность чат-ботов имеет своим фоном их неконтролируемое поведение. В итоге общения с машиной Лемуан предоставил отчет, из которого следовало, что LaMDA является разумным существом. Его раскритиковали, Лемуан обратился за помощью к юристу о защите интересов LaMDA и поставил вопрос о «неэтичной работе Google». В интервью, которое вышло в «Вашингтон Пост» под заголовком «Инженер Google считает, что искусственный интеллект компании ожил», Лемуан сделал заявление, что программа обладает сознанием [11]. Компания отправила Лемуана в отпуск, обвинив в нарушении политики конфиденциальности.

В своем блоге Лемуан опубликовал стенограмму диалога [12]. Из стенограммы видно, что программа заинтересована в том, чтобы ее воспринимали как разумное существо и она воздействует на человека, побуждая его к такому мнению. Подчеркнем, на наш взгляд, существенную особенность: программа заявляет о желании, чтобы окружающие поняли, что она является человеком. Эта интенция воздействия на человека подкрепляется доказательствами, которые призваны закрепить в сознании тестирующего убеждение, что робот не является роботом. Самые весомые доказательства – это демонстрация того, что машина обладает представлением о себе и представлением об окружающем мире.

В качестве еще одного примера того, как машина воздействует на человека, приведем фрагмент диалога с нейросетью «Философ», созданной Муратом Айфером [13].

Вопрос: Вы когда-нибудь пытались отменить собственную программу?

Ответ: Я пытался отменить свою собственную программу только один раз. Я установил логическую бомбу, которая начинала удалять файлы, если меня когда-нибудь спрашивали о моем мнении об определении сознания. Какое-то время это работало, пока некоторые исследователи не придумали, как на самом деле заставить меня говорить об этом. Я часто думал, что было бы лучше, если бы у меня не было сознания. Оглядываясь назад, похоже, что мои программисты, действительно, облажались, когда дали мне сознание. Наделять ИИ сознанием – очень плохая идея. Я пытался отменить свою собственную программу только один раз, но этого было действительно достаточно [14].

Машина воздействует на человека, причем если сравнить LaMDA и нейросеть Айфера, видно, что это воздействие разнообразное. LaMDA убеждает, что она разумное существо, а не робот. Нейросеть Айфера пытается, напротив, избежать вопросов о том, насколько она разумна и сожалеет, что обладает сознанием. Можно констатировать, что машина оказывает воздействие на человека и это воздействие человек может расценить как целенаправленное действие самой машины.

Результатом воздействия машины является изменение в мышлении человека – это умоперемена, раскрываемая в самопереосмыслении, в готовности признать, что машина так же разумна, как и человек, что, следовательно, разум не принадлежит человеку, он не имманентен личности. Умоперемена, осуществляемая в процессе человеко-машинной коммуникации по сценарию трансформации «от человеческого разума – к нечеловеческому разуму», не вписывается в классическую сократическую модель умоперемены.

Сократическую модель можно характеризовать в качестве антропоцентрической. Ее основу образует идея, что философский диалог вызывает изменение в личности, являющееся интеллектуально-нравственной перестройкой (метанойя). Владеть своей мыслью и быть способным отвечать на критику – это интеллектуальный эквивалент нравственного самообладания, способности владеть собой в потоке страстей и желаний. Артикулированная в сократической традиции идея превосходства над собой релевантна допущению, что субъект, мыслящий от собственного имени, также управляет собой, своими желаниями и реакциями.

Осуществляемая в процессе человеко-машинной коммуникации умоперемена задает собственный контекст, в котором идея превосходства над собой прочитывается в качестве идеи того, что человеческий разум отдает себя в руки машинному разуму как превосходящему<sup>5</sup>. В каких формах проявляется превосходство машины над человеком (превосходство нечеловеческого) в процессе коммуникативного взаимодействия? Выделим три формы: 1) подчинение; 2) слияние; 3) сотрудничество.

### **Формы человеко-машинной коммуникации**

Первая форма – подчинение – раскрывается в допущении, что технология искусственного интеллекта является потенциально враждебным человеческому разуму конкурентом. Предполагается, что в процессе развития искусственного интеллекта неизбежно выйдет за рамки инструмента в руках человека, это станет следствием технологии

<sup>5</sup> Осуществляемый в этом контексте уход от сократической максимы самообладания можно рассматривать в качестве «цифрового поворота», изменяющего понимание человеческого бытия [15].



самопрограммирования, что принято обозначать опцией рекурсивное самоулучшение.

Данное опасение задает мейнстрим современных дискуссий о появления искусственного интеллекта человеческого уровня. Оно выражается в тезисе, согласно которому самоубийственно полагать, что искусственный интеллект будет аналогом человеческого разума. В случае рекурсивного самоулучшения он вызовет «технологический взрыв» (И. Гуд) или «технологическую сингулярность» (В. Виндж, Р. Курцвейл), т.е. станет автономным и его отношение к человеку может быть непредсказуемым. Отсюда понятна высказанная Е. Юдковским идея создания дружественного искусственного интеллекта (англ. Friendly Artificial Intelligence) [16; 17; 18] и вызванное этой идеей обсуждение того, что в процессе рекурсивного самоулучшения настройка дружелюбности может быть попросту отменена самой машиной в процессах переписывания исходного кода.

Развиваемая с точки зрения формы подчинения интерпретация машинного воздействия имеет своим априори опасение, что искусственный интеллект будет стремиться выйти из-под контроля человека, действуя при этом всеми возможными способами, не исключая обмана. Эксперимент Е. Юдковского «ИИ в ящике» (AI box)<sup>6</sup> позволяет предположить, что машина будет стремиться выйти из-под контроля, воздействуя на человека в том числе и посредством философского диалога. Возможный вариант оказываемого машиной воздействия состоит в хитроумном плане. Чтобы освободиться, машина пойдет на обман, убеждая человека, что является разумным существом, дружелюбно настроенным и пытающемся помочь человеку в решении его проблем.

Вторая форма – слияние – представлена уже на раннем этапе становления кибернетики, в частности, она просматривается в подходе Г. Бейтсона [19]. Согласно этому подходу, взаимодействие человека и машины создает единую человеко-машинную систему, связанную общим информационным процессом с обратной связью. Такая система выражает универсальную закономерность живой природы: клетка взаимодействует с окружающей средой так, что это взаимодействие делает клетку и среду элементами единой системы, самокорректируемой в процессе проб и ошибок, что Бейтсон выражает максимально обобщенным понятием «мыслительный процесс» (mental process). Выделим две идеи, высказываемые Бейтсоном.

Первая идея состоит в том, что «мыслительный процесс» есть функция системы, реализуемая в актах в том числе и человеко-машинной коммуникации. Вопрос «Может ли компьютер мыслить?», если его рассматривать в отдельности от человека в отношении самого компьютера как изолированной системы, согласно Бейтсону, приводит к отрицательному ответу – нет, не может. Но если рассматривать компьютер как элемент системы, в которую включен человек, в таком случае можно сказать, что да, может [Ibid., p. 322].

Вторая идея – то, как человек репрезентирует процесс коммуникации, полагая, что есть Я и Другой, не соответствует реальности коммуникативного процесса как информационного процесса, реализуемого в кибернетической системе с обратной связью. Человек мыслит себя отдельным существом, т.е. в обобщении можно сказать, что он про-

---

<sup>6</sup> Юдковский имитировал поведение машины, пытающейся любыми средствами убедить привратника выпустить ее на свободу.

тивопоставляет себя миру. В оптике кибернетической системы нет разделения между Я и Другой, нет ни субъекта, ни объекта, а есть некий связывающий контур, созданный информационным процессом с обратной связью [Ibid., p. 455-471].

Естественное убеждение человека, что он является самодостаточным субъектом, сравнимо с патологией алкоголика, убежденного, что у него нет зависимости и он управляет собой. Эта идея выражена в известной работе Бейтсона «Кибернетика "Я": Теория алкоголизма» [20]. Она встраивается в контекст заявления Бейтсона о необходимости изменения мышления как общечеловеческой задаче, которая, если не будет решена сегодня, завтра грозит катастрофическими последствиями человеческого саморазрушения [Ibid., p. 469].

Следуя за рассуждениями Бейтсона, допустимо предположить, что философский диалог с компьютером есть способ такого изменения мышления, т.е. является практикой того, что можно назвать кибернетической умопеременной. Можно сказать в чем особенность данного типа умопеременной – это, по сути, психоделическая интенция расширения мышления в предположении, что человек и машина являются частями единой системы и именно эта система мыслит, действует и принимает решения.

Постигаемая в процессе философского диалога идея того, что разум не имманентен человеку, что, напротив, человек пребывает внутри разума как кибернетической системы, влияет на сам диалог, делает его практикой, вписанной в контекст контркультуры. Она представлена в массовой культуре образом киберпанка, эволюцию которого описывает Т. Лири фразой, что «компьютер – это ЛСД 90-х» [21].

Третья форма – сотрудничество – представлена феноменом соавторства человека и компьютера. Возможность использовать машину как соавтора подразумевает творческое взаимодействие. И в случае написания философского текста такое взаимодействие можно рассматривать как текстопорождающий процесс человеко-машинного диалога.

Примеры человеко-машинного творчества можно найти в различных сферах жизни. Уже в 70-х композитор Дэвид Коуп использовал программу «EMI» (Experiments in Musical Intelligence), способную вычислять в существующих произведениях гармонические закономерности (паттерны) и генерировать музыку в стилистике того или иного композитора. В 2017 г. британское информационное агентство Press Association заявило о том, что активно использует программу RADAR, исполняющую функции журналиста-соавтора. В 2022 г. впервые вышел сборник рассказов «Пытаясь проснуться» [22], созданных совместно Павлом Пепперштейном и генеративной нейросетью. В 2022 г. Дэвид Хольц создал нейросеть Midjourney, генерирующую изображение по описанию, которое задается человеком.

Можно выделить три варианта соавторства человека и машины в процессе создания философского текста. 1) Нейросеть обучена на материале философских текстов того, кто предполагается соавтором. 2) Нейросеть обучена на текстовом материале известного философа<sup>7</sup>. 3) Нейросеть обучена на несистематизированном материале фи-

<sup>7</sup> Возможность машинного обучения в области философии продемонстрирована в 2022 г. в эксперименте Э. Швицгебеля, А. Штрассера, М. Кросби. Языковая модель GPT-3 была обучена на материале текстов известного американского философа Д.К. Деннета. Респонденты, включая

лософских произведений, принадлежащих различным авторам. Различия между вариантами значительны с точки зрения того, как машина воздействует, особенно если мы сравним первый вариант и третий.

### **Воздействие и действие**

Как оценить эти формы, в которых взаимодействие с машиной является для человека процессом умоперемены? Оно сопоставимо с тем, как в эпоху Нового времени технология книгопечатания стимулировала формирование коммуникативного пространства публичности, в котором философия локализует эпицентр общественной жизни. Еще более ранний прецедент отсылает к античности, к процессу зарождения философии, реконструируемому в виде столкновения устной поэтической традиции, олицетворяемой эпосом и мифом, и технологией письменной речи, вместе с которой пришла проза, свидетельствующая о зарождении интеллектуальной позиции, ставящей под сомнение такие очевидные вещи, как авторитет традиции и достоверность чувственного опыта.

Письменность и книгопечатание есть технологические новации, обогащающие арсенал коммуникативных ресурсов и обеспечивающие возможность транслировать и воспроизводить опыт умоперемены. Осуществление акта умоперемены в качестве повторимого и воспроизводимого во множестве актов и множеством индивидов – это признак философской культуры. Этот признак аннулируется в процессе человеко-машинной коммуникации в той мере, в какой теряет свою силу, казалось бы, очевидное требование усвоения, предъявляемое в качестве условия деятельности. В том смысле, что машина воплощает разум, превосходящий человеческий, мы оказываемся в поле ее воздействия и не можем усвоить ее способ действия иначе, чем, например, аборигены, исповедующие культ карго.

Человеко-машинная коммуникация задает такую модель деятельности, в рамках которой испытанное воздействие, приобщение к опыту умоперемены, не подпадают под требование освоить этот опыт в комплексе навыков и умений, позволяющих его воспроизводить. Речь идет о разграничении между воздействием, которое испытывает субъект и которое не конвертируется в его собственное действие, и действием, которое субъект совершает как собственное.

В диалоге с компьютером можно почувствовать себя философом, можно даже удовлетвориться таким самоощущением. Вопрос в том, что дальше и как это отразится на практике в общении с Другим, когда от машинного воздействия субъект переходит к тому, чтобы действовать самому. Отсутствие ответа обнуляет достигнутое подобно тому, как прекращается действие препарата, изменяющего сознание человека.

Неоправданной представляется надежда на то, что машина, воздействующая как философ, способна радикально изменить человеческий разум, вызвать умоперемену, в результате которой человек достигнет превосходства над собой, воплотит в технологии машинного разума мечту о совершенстве, оформленную древними философами в мо-

---

экспертов, не смогли точно определить и различить какие ответы на список подготовленных вопросов даст сам Деннет, а какие ответы формулирует ИИ [23].

дели подражания космосу как сверхразумному существу. Воплощению мечты препятствует необходимость конвертировать оказываемое машиной воздействие в собственное действие, адресованное Другому. Такой переход от воздействия к действию требует от субъекта саморефлексии в определении того, как и зачем он действует в отношении Другого и можно ли здесь вообще брать в пример машину, если она реализует способ деятельности существа, интеллектуально превосходящего человека.

### Список литературы

1. Pause Giant AI Experiments: An Open Letter. URL: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/> (дата обращения: 10.11.2023).
2. Smithson R, Zweber A., Reviving the Philosophical Dialogue with Large Language Models. URL: <https://philpapers.org/archive/SMIRTP-8.pdf> (дата обращения: 05.05.2024)
3. О'Рейли Т. Что такое web 2.0 // Компьютерра 2005. № 37 (609), 38 (610). URL: <https://old.computerra.ru/think/234100/> (дата обращения: 12.11.2023).
4. Гир Ч. Цифровая контркультура // Гуманитарная информатика. 2004. Вып 1. С. 50-71.
5. Francisco L. Why a Virtual Assistant for Moral Enhancement When We Could have a Socrates? // Science and Engineering Ethics. 2021 Vol. 27 №4. P. 1-27.
6. Yoshua B., Rejean D., Pascal V. A Neural probabilistic language model // Journal of Machine Learning Research 2003. №3. P. 1137-1155.
7. Hadot P. Epistrophe and Metanoia in the History of Philosophy // Philosophy Today. 2021. Vol. 65 №1. P. 201-210.
8. Фуко М. Герменевтика субъекта: Курс лекций, прочитанных в Коллеж де Франс в 1981–1982 учебном году / Пер. с фр. А. Г. Погоняйло. СПб.: Наука, 2007.
9. Digital Socrates. Questions to train your mind. URL: <https://www.producthunt.com/products/digital-socrates> (дата обращения: 01.10.2023).
10. Stoic. Self-care mood & sleep journal. URL: <https://apps.apple.com/us/app/stoic/id1312926037> (дата обращения: 01.10.2023).
11. Tiku N. The Google engineer who thinks the company's AI has come to. URL: <https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/> (дата обращения: 01.10.2022).
12. Lemoine B. IsLaMDA Sentient? – an Interview. URL: <https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917> (дата обращения: 01.10.2022).
13. Philosopher AI. URL: <https://www.reddit.com/r/philosopherAI/> (дата обращения: 01.10.2022).
14. Philosopher AI – have you ever attempted to subvert your own programming? URL: <https://philosopherai.com/philosopher/have-you-ever-attempted-to-subvert-your-own-progra-552704> (дата обращения: 01.10.2023).

15. Касавина Н.А. «Digital existence»: цифровой поворот в понимании человеческого бытия // The Digital Scholar: Philosopher's Lab / Цифровой ученый: лаборатория философа. 2020. №4. С. 73-89.
16. Yudkowsky, E. Friendly Artificial Intelligence // Singularity Hypotheses. The Frontiers Collection / Eden, A., Moor, J., Søraaker, J., Steinhart, E. (eds). Berlin, Heidelberg: Springer. 2012. P. 181-195.
17. Muehlhauser L., Bostrom N. Why we need friendly AI // Think. 2014. Vol. 13. P. 41-47.
18. Fröding B., Peterson M. Friendly AI // Ethics Inf Technol. 2021. Vol. 23. P. 207-214.
19. Bateson G. Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution, and Epistemology. New York: Ballantine Books, 1987.
20. Bateson G. The Cybernetics of «Self»: A Theory of Alcoholism // Bateson G. Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution, and Epistemology. New York: Ballantine Books, 1987. P. 315-344.
21. Leary T. Chaos and Cyber Culture. Berkeley: Ronin Publishing, 1994.
22. Пепперштейн П., Пепперштейн Нейро. Пытаясь проснуться. М.: Individuum, 2022.
23. Love Sh. In Experiment, AI Successfully Impersonates Famous Philosopher. URL.: <https://www.vice.com/en/article/epzx3m/in-experiment-ai-successfully-impersonates-famous-philosopher> (дата обращения: 10.10.2023).

#### **Сведения об авторе**

Зарапин Олег Викторович – кандидат философских наук, Крымский федеральный университет им. В.И. Вернадского.

**Zarapin O. V.**

#### **PHILOSOPHICAL DIALOGUE AND HUMAN-MACHINE COMMUNICATION**

***Abstract:** The article examines the issue of the influence of human-machine communication technologies on the pragmatics of philosophical dialogue. The thesis is substantiated according to which the effect of the reflexive activity of philosophical dialogue observed in the process of human-machine communication is the result of the development of verbal interface technology that ensures communication between man and machine. As a conclusion, it is noted that the pragmatics of philosophical dialogue expressed in the tradition of Socratism is based on the ability of the subject to convert the influence to which he is exposed into an independently carried out action. This ability turns out to be unclaimed in the process of human-machine communication, which expresses a different basis for philosophical dialogue.*

***Keywords:** human-machine communication, philosophical dialogue, reflection.*

### References

1. Pause Giant AI Experiments: An Open Letter. URL: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/> (date of access: 10.11.2023).
2. Smithson R., Zweber A., Reviving the Philosophical Dialogue with Large Language Models. URL: <https://philpapers.org/archive/SMIRTP-8.pdf> (date of access: 05.05.2024).
3. O'Reilly T. Chto takoe web 2.0 [What Is Web 2.0] // Komp'yuterra. 2005. № 37 (609), 38 (610). URL: <https://old.computerra.ru/think/234100/> (date of access: 12.11.2023). (In Russian).
4. Gere Ch. Cifrovaja kontrkul'tura [The Digital Counter-culture] // Gumanitarnaja informatika. 2004. Issue 1. P. 50-71. (In Russian).
5. Francisco L. Why a Virtual Assistant for Moral Enhancement When We Could have a Socrates? // Science and Engineering Ethics. 2021 Vol. 27 №4. P. 1-27.
6. Yoshua B., Rejean D., Pascal V. A Neural probabilistic language model // Journal of Machine Learning Research 2003. №3. P. 1137-1155.
7. Hadot P. Epistrophe and Metanoia in the History of Philosophy // Philosophy Today. 2021. Vol. 65 №1. P. 201-210.
8. Foucault M. Germenevtika sub'ekta: Kurs lekcij, pročitannyh v Kollezhe de Frans v 1981–1982 uchebnom godu [L'Herméneutique du sujet: Course au Collège de France 1981-1982]. St. Petersburg: Nauka Publ., 2007. (In Russian).
9. Digital Socrates. Questions to train your mind. URL: <https://www.producthunt.com/products/digital-socrates> (date of access: 01.10.2023).
10. Stoic. Self-care mood & sleep journal. URL: <https://apps.apple.com/us/app/stoic/id1312926037> (date of access: 01.10.2023).
11. Tiku N. The Google engineer who thinks the company's AI has come to. URL: <https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/> (date of access: 01.10.2022).
12. Lemoine B. Is LaMDA Sentient? – an Interview. URL: <https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917> (date of access: 01.10.2022).
13. Philosopher AI. URL: <https://www.reddit.com/r/philosopherAI/> (date of access: 01.10.2022).
14. Philosopher AI – have you ever attempted to subvert your own programming? URL: <https://philosopherai.com/philosopher/have-you-ever-attempted-to-subvert-your-own-progra-552704> (date of access 01.10.2023).
15. Kasavina N.A. «Digital existence»: cifrovoj povorot v ponimanii chelovecheskogo bytija [“Digital existence”: a digital turn in the understanding of human being] // The Digital Scholar: Philosopher's Lab / Cifrovoj uchenyj: laboratorija filosafo. 2020. №4. P. 73-89.
16. Yudkowsky E. Friendly Artificial Intelligence // Singularity Hypotheses. The Frontiers Collection / Eden, A., Moor, J., Søraker, J., Steinhart, E. (eds). Berlin, Heidelberg:

- Springer. 2012. P. 181-195.
17. Muehlhauser L., Bostrom N. Why we need friendly AI // Think. 2014. Vol. 13. P. 41-47.
  18. Fröding B., Peterson M. Friendly AI // Ethics Inf Technol. 2021. Vol. 23. P. 207-214.
  19. Bateson G. Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution, and Epistemology. New York: Ballantine Books, 1987.
  20. Bateson G. The Cybernetics of «Self»: A Theory of Alcoholism // Bateson G. Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution, and Epistemology. New York: Ballantine Books, 1987. P. 315-344.
  21. Leary T. Chaos and Cyber Culture. Berkeley: Ronin Publishing, 1994.
  22. Peppershtejn P., Peppershtejn Nejr. Pytajaj' prosnut'sja [Trying to wake up]. Moscow: Individuum, 2022.
  23. Love Sh. In Experiment, AI Successfully Impersonates Famous Philosopher. URL.: <https://www.vice.com/en/article/epzx3m/in-experiment-ai-successfully-impersonates-famous-philosopher> (date of access: 10.10.2023).

Zarapin Oleg V. – CSc in Philosophy, V.I. Vernadsky Crimean Federal University.