В. А. Петровский,
профессор НИУ ВШЭ
М. А. Вайнберг,
магистрант НИУ ВШЭ

Трансвитальное «Я», искусственный интеллект и личный выбор бессмертия

Вместо эпиграфа:
http://littlejapan.ru/tupak-2pac-gologramma/

Два термина, фигурирующие в названии, требуют пояснения.

Трансвитальное Я – особая форма бытия индивида за пределами наличных форм его жизнедеятельности, процессов воспроизводства себя как психофизической целостности (Петровский, 2023, https://psyjournals.ru/journals/chp/archive/2023_n3). Достижение состояния трансвитального Я объяснялось нами ранее стремлением человека обрести свою идеальную представленность и продолженность в других людях, а оно (это стремление), в свою очередь, трактовалось нами как проявление фундаментальной человеческой потребности в бессмертии (Петровский, работы 1981–2023 гг). В данной статье мы расширяем границы возможной области существования трансвитального Я, мысленно «обустраивая» его в виртуальном пространстве искусственного интеллекта.

«Искусственный интеллект» – это словосочетание в наши дни столь популярно, что появилась аббревиатура «ИИ», всеми узнаваемая и как будто бы не требующая комментариев. Между тем, в *общественном сознании* «искусственный интеллект» фигурируют в двух значениях, столь несовпадающих друг с другом, что требуют разного обозначения – мы предлагаем кратко обозначать их «ИИ-1» и «ИИ-2». В научной литературе, а также в «жанровой прозе» (фантастика, фэнтези) и публицистике, фигурируют термины «сильный» («узкий») ИИ, «сильный» ИИ и даже «сверхсильный» ИИ (по сути, «сильный», но только «сильнее»). В интересах исследования мы придерживаемся звучащих нейтрально «ИИ-1» и «ИИ-2» не используем оценочных эпитетов «слабый», «узкий», «сильный», «сверхсильный» и т. п.

ИИ-1 – «бездушная машина», эффективно решающая задачи различной степени сложности, действующая на основе и в пределах программы, вложенной в нее человеком и ориентирующаяся в текущей информационной среде («инструмент», «автомат», «орудие» человеческой деятельности); идея ИИ-1 – прерогатива научной мысли, подкрепляемай впечатляющими достижениями техники.

ИИ-2 – полуфантастическое антропоморфное существо, произведенное человеком, автономное в своем функционировании, спонтанное, саморазвивающееся («субъект», обладающий свободой воли и самоопределения, самопрограммирующаяся система: сверхразумное существо, действующее на основе обратных связей в среде).

При, казалось бы, утопичности идеи существования ИИ-2, отметим, что некоторые аналоги «субъектности» в лице ИИ все же встречаются наяву («программы, порождающие программы»). В футурологическом эссе об искусственном интеллекте (Петровский, 2023, https://psy.su/pubs/11802/) рассматривались логические и онтологические предпосылки перерастания ИИ-1 в ИИ-2.

Оба варианта понимания ИИ ассоциируются в сознании современников с экзистенциальными рисками (мы также, как и многие авторы, допускаем «выход» искусственного интеллекта «из ящика»). «Бездушная машина», ИИ-1, может давать «сбои», приводящие к человеческим жертвам (число подобных случаев нарастает), но в данном случае риски купируются очевидными достоинствами ИИ (способность решать задачи, недоступная естественному интеллекту).

«Субъектный» ИИ-2 ассоциируется, как правило, с актами насилия над человеком со стороны превосходящей его силы (в научной фантастике это идея, например, «бунта роботов»), что не смягчается какими-либо видимыми «плюсами» искусственного интеллекта (если, разумеется, не принимать в расчет романтические истории о влюбленности робота в человека).

Авторы хотели бы развить для читателей тему возможного усиления идеи ИИ-2 как мыслимой реальности, имеющей позитивный смысл – в виде условия его ***личного бессмертия***, и в этом контексте – об *отношении человека* к самой возможности полагания собственного *трансвитального Я* в виртуальном пространстве искусственного интеллекта – ИИ-1 и ИИ-2.

Идея эксперимента

Первая часть – подготовительная; она посвящена диалогам с испытуемыми о личном бессмертии («хочу – не хочу», «верю – не верю», «знаю, что это возможно, – отрицаю такую возможность», «придерживаюсь естественно-научной ←→ религиозной (христианская точка зрения), верю в земное бессмертие (жизнь в близких людях, «инобытие в других»), принимаю идею «метемпсихоза» (переселение душ, «переодушевление») или, наконец, «особое мнение»)

Основная часть 1 – констатирующий этап эксперимента. Испытуемому предлагается последовательно ответить на следующие ответы.

Первый вопрос: как бы участники разговора отнеслись к возможности обрести личное бессмертие в виртуальном пространстве искусственного интеллекта? (здесь в разговоре с испытуемыми еще не различаются ИИ-1 и ИИ-2)

Второе вопрос касается различий между двумя моделями искусственного интеллекта – ИИ-1 и ИИ-2 (оба могущественны, являя собой «сверхразум», но ИИ-1 управляем извне, содействуя при этом достижению важных для вас целей, а ИИ-2 способен ставить и преследовать свои собственные цели).

Заметим, что ИИ-1 существует уже сегодня и, например, обыгрывает гроссмейстеров в шахматы, а ИИ-2 – это фантазия (или утопия), хотя некоторые ученые считают, что она состоится в будущем.

«Отвечая на вопросы, которые будут поставлены, Вы можете исходить из идеи сходства и различий между ИИ-1 и ИИ-2».

Третий вопрос. Здесь требуется проявить фантазию: «Представьте, что Вы – герой научно-фантастического рассказа или фильма, в котором Вам предлагается обрести вечную жизнь, которую гарантировано обеспечит Вам могущественный искусственный интеллект – сверхразум. Вы – согласитесь? (на каких условиях?)»

Основная часть 2:

Независимые переменные (А). «Естественные» различия групп. Сравнение решений о личном бессмертии в ИИ у представителей разных профессиональных и учебных групп – с учетом возраста, пола, здоровья, образования. Среди участников – психологи, философы, математики, социологи, писатели-фантасты, разработчики компьютерных программ, атеисты, верующие (представители разных конфессий), представители разных этнических групп.

Независимые переменные (Б). «Наведенные» различия, направленные на усиление идентификации испытуемого с его виртуальным подобием:

а) *до демоверсии:* испытуемому предлагаются вопросы о его готовности быть продолженным в виртуальном пространстве ИИ, взаимодействующим с реальным окружением, – и это при том, что виртуально-реальное субъекта не имеет визуальной представленности ни на экране, ни в наушниках, ни в ситуации телесной стимуляции (обсуждается лишь абстрактная возможность бессмертия, предоставляемого ИИ);

б) «*живая* демоверсия»: испытуемый видит «себя» на экране, слышит «свой голос», наблюдает «свое поведение» со стороны, прослеживает решение этических дилемм, предпринимаемые виртуальным двойником в «реальном мире», наблюдает последствия «своих действий», испытывает знакомые ему кинестетические ощущения (современные ИИ-1 или в ближайшее время позволят добиться подобных эффектов).

Зависимые переменные: вероятность решения «да» или «нет» (и промежуточных вариантов) при решении вопроса о личной готовности обрести бессмертие в искусственном интеллект ИИ-1 и ИИ-2, а также выдвигаемые испытуемыми условия принятия положительных решений и феноменология переживаний, вместе с внутренними диалогами по этому поводу.

Авторы предстоящего исследования уже сейчас располагают *общей* и некоторыми *эмпирическими гипотезами*, в которых соотносятся указанные переменные. Но эти гипотезы пока мы не формулируются на Новостной ленте идей Журнала (мы проведем со временем дополнительное исследование на эту тему среди читателей, готовых предугадать будущие результаты планируемых экспериментов; впрочем, вы могли бы заглянуть в Приложение, и прислать нам свои ответы в журнал)

Приложение,
необязательное к прочтению
и, тем более, не являющееся руководством к действию

ЧТО ПОЛУЧИТСЯ?

Если читателей заинтересовал предложенный здесь сюжет, то мы попросили коллег ответить на несколько вопросов, вдогон всему сказанному. на тему «что получится?»

Итак: каков процент участников эксперимента, которые подтвердят готовность принять решение о «бессмертии» в искусственном интеллекте (желательно при различении ИИ-1 и ИИ-2). Будем Вам признательны, если Вы ответите хотя бы на один из этих вопросов.

ДО ДЕМОВЕРСИИ:
ИИ-1: _____ %
ИИ-2: _____ %

А как бы ответили ВЫ? Были бы сами готовы принять решение «Да»?

ДО ДЕМОВЕРСИИ
(подчеркните свое решение)
ИИ-1: принимаю, сомневаюсь, допускаю, воздерживаюсь
ИИ-1: принимаю, сомневаюсь, допускаю, воздерживаюсь

А после того, как вы увидели и «ощутили» себя «живущим» в ИИ-1 и ИИ-2?
ИИ-1: принимаю, сомневаюсь, допускаю, воздерживаюсь
ИИ-1: принимаю, сомневаюсь, допускаю, воздерживаюсь

Спасибо, что ответили на поставленные вопросы (если у Вас нашлось пару минут)!