



ТЕХНОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ПСИХОЛОГИИ

УШАКОВ Д.В.

*Институт психологии Российской академии наук (ФГБУН ИП РАН), г. Москва, Российская Федерация
ORCID: <https://orcid.org/0000-0001-9716-1545>, e-mail: dv.ushakov@gmail.com*

В статье рассматриваются бурно растущие технологии искусственного интеллекта, направленные на познание и поддержку человека. Излагаются основные достижения, возникающие проблемы и перспективные пути развития. Аргументируется, что формирующиеся технологии искусственного интеллекта, направленные на человека, хотя и обещают большое увеличение психологического комфорта, но и угрожают серьезными рисками. Предлагается ориентироваться на концепцию «цифровых ангелов», под которыми понимаются технологии искусственного интеллекта, созданные для защиты интересов своего владельца. Цифровые ангелы должны совмещать три роли — цифровых ассистентов, позволяющих организовывать человеку его жизнь; цифровых коммуникаторов, обеспечивающих связь с другими людьми и со средой; и цифровых конфидентов, которым человек может безусловно доверять.

Ключевые слова: искусственный интеллект, большие языковые модели, автоматическое распознавание свойств и состояний человека, цифровые двойники, цифровые ангелы.

Финансирование. Статья подготовлена при поддержке РФФ, грант № 22-18-00704.

Для цитаты: Ушаков Д.В. Технологии искусственного интеллекта в психологии // Экспериментальная психология. 2024. Том 17. № 4. С. 182—189. DOI: <https://doi.org/10.17759/exppsy.2024170412>

ARTIFICIAL INTELLIGENCE TECHNOLOGIES IN PSYCHOLOGY

DMITRY V. USHAKOV

*Institute of Psychology of the Russian Academy of Sciences, Moscow, Russia
ORCID: <https://orcid.org/0000-0001-9716-1545>, e-mail: dv.ushakov@gmail.com*

The article examines the rapidly growing artificial intelligence technologies aimed at understanding and supporting humans. It outlines the main achievements, emerging problems, and promising development paths. It is argued that while these emerging human-oriented artificial intelligence technologies promise a significant increase in psychological comfort, they also pose serious risks. The concept of “digital angels” is proposed—artificial intelligence technologies created to protect the interests of their owner. Digital angels should combine three roles: digital assistants that help individuals organize their lives; digital communicators that ensure connection with other people and the environment; and digital confidants that a person can unconditionally trust.

Keywords: artificial intelligence, large language models, automatic recognition of human properties and states, digital twins, digital angels.

CC BY-NC



Funding. The article was prepared with the support of the Russian Science Foundation, project number № 22-18-00704.

For citation: Ushakov D.V. Artificial Intelligence Technologies in Psychology. *Экспериментальная психология = Experimental Psychology (Russia)*, 2024. Vol. 17, no. 4, pp. 182–189. DOI: <https://doi.org/10.17759/exppsy.2024170412> (In Russ.).

Революция в искусственном интеллекте в связи с развитием нейросетей — важнейшее технологическое событие последнего десятилетия, которое обещает привести к коренным изменениям в гуманитарных науках. Причина этого двояка. Во-первых, на сегодняшний день технологии искусственного интеллекта предоставляют мощные инструменты выявления закономерностей и моделирования человека и феноменов человеческой культуры, созданных точными науками. Возникает новый уровень возможностей оценки состояния людей, групп, культуры, а также предсказания в их отношении. Во-вторых, технологии искусственного интеллекта позволяют создавать гибкие инструменты взаимодействия с человеком, которые приближаются по тонкости и многогранности к человеческому общению.

Все это в совокупности позволяет рассматривать искусственный интеллект как перспективное средство не только для прикладных, но и для фундаментальных научных целей в гуманитарных науках [4].

Достижения искусственного интеллекта в психологии

Применение машинного обучения в психологии началось с задач распознавания свойств и состояний человека. При этом в первую очередь работы коснулись наиболее исследованных психологических таксономий, таких как Большая пятерка личностных свойств. Машинное обучение было применено и к распознаванию свойств когнитивной системы, важнейшим из которых является интеллект. Кроме индивидуальных свойств, активно исследуются психологические состояния. Эта тема имеет большое практическое значение в разных областях — для напряженных профессий, в школе, для персонала корпораций, для оценки психологического состояния больших социальных групп, регионов и т. д.

Распознавать индивидуальные особенности и состояния можно на основании разных типов информации. Один из них — видеоинтервью в стандартизированных условиях. В этом случае могут анализироваться данные, относящиеся к содержанию порождаемой испытуемым речи, ее акустике, а также к мимике, позе и жестике [2]. Другой распространенный тип данных — активность в социальных сетях [1]. Здесь используются текст, порожденный испытуемым, его перепосты, лайки и другая подобная информация. Для оценки состояний могут использоваться психофизиологические данные, такие как КГР, ЭЭГ, оценки сердечного ритма. Критерием при обучении служат результаты психологических тестов, а также экспертные оценки.

Фундаментальное значение может иметь применение методов объяснимого искусственного интеллекта, которые выявляют неизвестные ранее психологам особенности поведения людей с различными чертами [7; 8].

Автоматический анализ коммуникации — не просто более сложная задача по сравнению с автоматической диагностикой свойств и состояний людей. Это еще и подход, позво-



ляющий вывести последнюю на более высокий уровень и избавиться от отмеченных выше проблем и недостатков.

Примером анализа коммуникации является интен-анализ, разработанный в Институте психологии РАН [5]. Интен-анализ выделяет коммуникативные акты в соответствии с их интенциями, или намерениями говорящих. Интенция может быть направлена на себя (например самопрезентация), на собеседника (похвала, высмеивание и т.д.), на третьих лиц. Интенции говорящих определенным образом оцениваются их коммуникативными партнерами, хотя понимание со стороны далеко не всегда соответствует реальным интенциям. Соответственно реакция коммуникативных партнеров строится на основе оценки ситуации и интенций говорящих. Коммуникация, таким образом, закономерно формируется из переплетения различно направленных интенциональных актов коммуникантов. Эмоциональные реакции включены в коммуникацию и логически связаны, как с интенциями говорящих, так и с воспринимаемыми ими интенциями их коммуникативных партнеров.

Исходно интен-анализ разработан в связи с процедурами экспертной оценки, однако развитие методов машинного обучения и в особенности технологий обработки естественных языков (Natural Language Processing — NLP) открыло возможности его автоматизации. Автоматическое приписывание интенций высказываниям коммуникантов представляет интерес не только само по себе, но и позволяет развить методы автоматического анализа и предсказания хода человеческого взаимодействия, в том числе включенных в него состояний и свойств людей. Дело в том, что анализ коммуникации позволяет применить методы обучения без учителя, связанные с обучением предсказанию коммуникативных ходов взаимодействующих людей. Применяемый метод аналогичен тому, что используется в области обработки естественного языка, где параллелью к тексту является последовательность коммуникативных действий общающихся людей. Элементы (слова или коммуникативные акты) по очереди маскируются, а система обучается предсказанию зашумленных элементов.

Наиболее существенным результатом этого подхода становятся эмбединги, которые фактически отражают пространства состояний и свойств людей. Этим снимаются противоречия и недостатки предшествующих подходов к оценке свойств и состояний людей. Оценка состояния проводится не только на основании их внешних проявлений, но и с учетом коммуникативной ситуации, в которой человек находится. Это делает оценку существенно более точной. Коммуникативным взаимодействием ситуация для человека, безусловно, не ограничивается. Поэтому следующим шагом должен стать более широкий учет происходящего с человеком, способный интегрировать всю доступную информацию о человеке.

Цифровые двойники

Цифровые двойники фактически означают моделирование определенных сторон человеческой деятельности методами информатики. С помощью моделирования осуществимо предсказание. Сегодня просматриваются возможности создания цифровых двойников, как когнитивной системы, так и поведения в различных социальных ситуациях.

Одна из сфер, где в большой степени востребовано моделирование и прогнозирование развития когнитивной системы, — это образование, а именно проблема создания индивидуальных образовательных траекторий. Доказательные исследования образования показывают, что наиболее важным способом ускорить и улучшить обучение является занятие с индивидуальным репетитором, превосходящее все другие разрабатываемые учеными и практиками педагогические методы. Это и не удивительно, если учесть, что в классе из 20 учеников, от-



ражающем генеральную выборку, т. е. без специального подбора, оптимальная скорость движения для первого и последнего учеников различается в среднем в 8 раз.

Подходы к созданию систем индивидуализации образовательных траекторий можно разделить на три уровня. Первый уровень — это создание экспертной системы, отражающей представления опытных и успешных педагогов. Например, педагоги разделяют задания по определенному учебному курсу на группы, и система предъявляет ученику задания следующей группы, после того как он достиг заданного уровня успешности в решении заданий предыдущей группы. Понятно, что такой подход не позволяет достичь чего-либо большего по сравнению с преподаванием педагогом, но может разгрузить последнего, что, впрочем, уже неплохо.

Второй уровень предполагает машинное обучение на массивах данных по ученикам, осваивающим тот или иной учебный курс. При этом подходе нейросеть учится предсказывать, способен ли ученик при определенном профиле предшествующих достижений освоить тот или иной учебный фрагмент или решить ту или иную задачу. Такого рода подход реализуется, например, на базе концептуальных схем, предложенных Фальманом в Канаде.

Наконец, на третьем, наиболее высоком и пока практически неосвоенном, уровне возникает работа с цифровыми двойниками. Ученику в этом случае предъявляется задание, специально сконструированное для него на основе его когнитивного развития в соответствующей области. Пример реализации такого подхода можно увидеть в области тестов интеллекта, генерируемых компьютером. Там задание, предъявляемое испытуемому, основано на понимании факторов сложности его решения в соотношении с оценкой предыдущих действий испытуемого. Достаточно прост перенос подхода на область шахмат, где существует большой набор партий, сыгранных шахматистами разного уровня при разных контролях времени, а также возможность их компьютерного анализа с быстрым и точным выявлением ошибок. Это позволяет с учетом методов объяснимого искусственного интеллекта создавать цифровых двойников шахматистов заданного уровня квалификации и подбирать тестовые и обучающие задания для каждого уровня, действующие с максимальной эффективностью. Распространение этих подходов на реальные учебные дисциплины неизбежно, но требует большой работы.

Для моделирования и прогнозирования социального поведения человека просматривается другой путь, а именно использование больших языковых моделей, которые фактически содержат в себе культурный опыт человечества, зафиксированный в текстах. Разворачивание этого культурного опыта позволяет извлекать информацию о способах поведения человека в различных ситуациях.

Общее моделирование человеческих аттитюдов и поведения позволяет предсказать, исходя из специфических сведений о человеке, его поведение в широком круге ситуаций — от приобретения товаров и политической позиции до семейной жизни. Недавний пример — работа, в которой были смоделированы данные по более чем 1000 респондентов [9]. В ней респонденты проходили двухчасовое структурированное интервью, которое, кстати, проводилось ботом на основе большой языковой модели. Далее формировались цифровые двойники респондентов. Это достигалось путем загрузки данных интервью в большую языковую модель в качестве запроса. Адекватность полученных цифровых двойников оценивалась путем сравнения ответов респондента и его цифрового двойника по целому ряду тестов: Большой пятерке, игр из сферы бихевиоральной экономики (дилемма узника и т. д.) и нескольких социально-экономических экспериментов. Точность предсказания поведения человека его цифровым двойником составляла от 0,6 до 0,8.



Хотя люди создали большие языковые модели, способы хранения в них информации нам не вполне понятны. Эти модели порой «галлюцинируют» или выдают заведомо ложную информацию. Поэтому применять создаваемые с их помощью двойники нужно с использованием методов доверенного искусственного интеллекта, четко очерчивая как способы получения информации, так и область достоверного применения.

На моделирование социального поведения направлен и описанный выше анализ коммуникации, причем потенциально он может дать больше, чем применение больших языковых моделей. Дело в том, что большие языковые модели могут использовать, пусть даже комбинируя, только то знание, которое содержится в человеческих текстах. Но люди далеко не все знают о самих себе, и исследование их взаимодействия методами машинного обучения может внести много нового в это знание.

Перспективы

Проведенный обзор показывает, что искусственный интеллект серьезно занялся изучением людей и постепенно приобретает способность быстро и точно оценивать их свойства и состояния, предсказывать поведение и организовывать воздействие. С одной стороны, все это обещает человечеству невиданный доселе психологический комфорт, а с другой — открывает пугающие перспективы влияния и подчинения. Как человечество может справиться с этими вызовами? Этот вопрос следует рассмотреть в более широком контексте.

Где-то со второй половины XX в. человечество вступило в период, когда стали просматриваться перспективы усиления интеллектуальных возможностей за счет технологий. Обсуждаются две области таких технологий. Во-первых, это повышение биологического разума, т. е. создание технологическим путем человека сверхразумного. Во-вторых, конструирование искусственного сверхразума.

Какое-то время назад первый путь казался более реалистичным. Сегодня расшифрован геном человека и уже работают методы редактирования генома, такие как CRISPR. Хотя мы склонны считать человеческий мозг в высшей степени совершенным органом, возможно все же, что он не вершина совершенства. Во всяком случае вороны или попугаи с мозгом весом 10–20 грамм сопоставимы по уровню интеллекта с родственными нам шимпанзе, мозг которых в 20 или 40 раз тяжелее, и при этом обладают, по-видимому, более совершенным метаболизмом нейронов, чем млекопитающие [6]. Дальше мы, безусловно, будем больше узнавать о роли генов, о том, как они связаны с нашим интеллектом, а методы редактирования генома будут совершенствоваться. Вроде бы пути к технологическому совершенствованию интеллекта человека в принципе просматриваются. Тем не менее сегодня внесение изменений в геном человека находится под строгим этическим запретом. Китайский ученый, совершивший вмешательство в геном человеческого эмбриона, получил тюремный срок, хотя и действовал вроде бы в целях сохранения здоровья. Поэтому сегодня перспективы биологического изменения человека и увеличения его интеллекта весьма туманны, хотя и полностью отрицать такую возможность нельзя.

А вот в другом направлении — в сфере искусственного интеллекта — в последнее десятилетие произошел колоссальный рост, что заставило всерьез задуматься о перспективах появления так называемого сильного искусственного интеллекта, т. е. такого, который во многих или даже в большинстве проявлений сильнее человеческого разума. В каких-то областях искусственный интеллект уже превзошел человеческий, например в играх в шахма-



ты, в го и т. д. Сегодня он близко подошел к человеку и в некоторых видах работы с текстами. Причем очевидно, что это только начало.

Появление на горизонте туманных контуров сильного искусственного интеллекта обеспокоило и вызвало к жизни прогнозы и призывы со стороны таких влиятельных людей, как Б. Гейтс, С. Хокинг, И. Маск, лидеры крупнейших ИТ корпораций, которые предложили ряд подходов к тому, как человечеству выстраивать отношения с нарождающимися интеллектуальными технологиями.

Один из подходов, который, однако, вряд ли наберет много сторонников, состоит в том, что человечество должно сойти с исторической сцены, создав более мощный, чем оно само, искусственный интеллект и уступив ему свое место. Как выражался Мишель Фуко, человек исчезнет, как след ноги на песке после прилива. Сюда примыкает и сценарий «восстания машин». Гейтс и Хокинг сравнили перспективы человечества, сосуществующего с сильным искусственным интеллектом, с судьбой муравейника, который затапливают при строительстве плотины не из-за желания зла, а просто в связи с необходимостью решения крупных задач.

Однако восстание машин, к счастью, выглядит сомнительным в эволюционной перспективе. Для восстания машинам требуется не только обладать выдающимися когнитивными возможностями, но и мотивационной системой, завязанной в том числе на гомеостатические потребности тела, а также выходом на материальные и идеальные инструменты человеческой деятельности. Воспроизводство всех этих элементов означает фактически сотворение искусственной жизни, причем в ее разумной форме. В принципе, наверно, такое возможно, однако непонятно, зачем человечеству может потребоваться создать искусственную жизнь и уступить ей место, в то время как альтернативой является совершенствование собственной жизни с помощью внедряемых в нее технологий.

Противоположная идея заключается в том, чтобы заморозить работы в области сильного искусственного интеллекта с целью не допустить вреда для человечества. Заморозка может быть временной, но возможен и полный запрет на развитие определенных направлений технологии. В то же время остановка технологического движения вряд ли возможна, если объективно анализировать современную эволюционную ситуацию. Силы, приводящие в движение эволюцию, а именно расширенное самовоспроизведение структур, в полной мере сохраняются, а технологии искусственного интеллекта выступают мощным средством повышения эффективности решения задач. Поскольку человечество многосубъектно, проблематичность подхода заключается в готовности всех субъектов принять этот запрет.

Из сказанного возникает видение наиболее желательной перспективы — это сосуществование людей, придающих смыслы происходящему, с сильным искусственным интеллектом, превосходящим их в вычислительных возможностях. Такое видение перспектив заставляет задаться вопросом, как человеку оказаться субъектом, т. е. активным по отношению к этому искусственному интеллекту? Как избежать использования новых мощных технологий для манипуляций и закабаления одних людей другими?

Цифровой ангел

Из сказанного следует, что независимость личности в эпоху сильного искусственного интеллекта возможна только в том случае, если у каждого человека будет работающий исключительно на него искусственный интеллект для обеспечения его интересов и свободы. Такой искусственный интеллект мы предложили назвать цифровым ангелом [3]. Человек



в мире сильного искусственного интеллекта должен быть подобен флагманскому кораблю, окруженному своей эскадрой, которая позволит взаимодействовать с чудным новым миром с его колоссальными вычислительными возможностями.

Цифровой ангел должен обладать тремя ипостасями. Во-первых, он должен быть электронным ассистентом, т. е. обеспечивать новый уровень комфорта и организации нашей жизни. Уже сегодня мы наблюдаем, как постепенно складывается рынок технологий, которые позволяют нам следить за здоровьем и психологическим состоянием, организовывать быт, составлять планы путешествий, помогают в решении задач образования, профориентации и многое другое.

Другая важнейшая ипостась цифрового ангела — его работа в качестве электронного координатора, или медиатора. Причем речь идет не только о взаимодействии в непосредственном круге общения. Мы знаем, что технологическое развитие привело к таким феноменам, как дипфейки, и к различным новым видам мошенничества. Цифровые ангелы в своей ипостаси электронных координаторов должны также защищать нас от недостоверной информации, анализировать в наших интересах направляемую нам рекламу и т. д.

Наконец, цифровой ангел теряет свой смысл без третьей ипостаси — электронного конфиденанта, которому опекаемый им человек может безоговорочно доверять. Для обеспечения эффективной работы в качестве ассистента и координатора электронные системы должны обладать максимально возможной информацией, и только их безусловная надежность в предотвращении распространения информации и направленность на защиту интересов конкретного человека могут служить базой для исполнения ими своих функций.

Представляется, что в рамках такой перспективы происходящее сегодня развитие технологий искусственного интеллекта психологического содержания приобретает оптимистичный гуманистический смысл.

Литература

1. *Валуева Е.А., Григорьев А.А., Лаптева Е.М., Панфилова А.С., Ушаков Д.В.* Когнитивная сложность и коммуникативный контекст: отражение интеллекта пользователей в текстах социальных сетей // Психологический журнал. 2023. Том 44. № 1. С. 70–80. DOI:10.31857/S020595920024352-3
2. *Рафикова А.С., Валуева Е.А., Панфилова А.С.* Голос и психологические свойства человека: обзор современных исследований // Психология. Журнал Высшей школы экономики. 2022. Том 19. № 1. С. 195–215. DOI:10.17323/1813-8918-2022-1-195-215
3. *Степанов С.Ю., Оржековский П.А., Ушаков Д.В.* Оценка ученика: на пути к цифровому образованию. Концептуально-математическая модель // Народное образование. 2019. № 1. С. 130–139.
4. *Ушаков Д.В.* Искусственный интеллект как инструмент психологического исследования // Сибирский психологический журнал. 2023. № 90. С. 188–200. DOI:10.17223/17267080/90/11
5. *Ушакова Т.Н., Павлова Н.Д., Латынов В.В., Цеццов В.А., Алексеев К.И.* Слово в действии. Интенанализ политических дискуссий. М., 2000.
6. *Eugen K. von, Endepols H., Drzezga A., Neumaier B., Gunturkun O., Backes H., Strochens F.* Avian neurons consume three times less glucose than mammalian neurons // Current Biology. 2022. Vol. 32. P. 4306–4313.
7. *Panfilova A.S., Turdakov D.Y.* Applying explainable artificial intelligence methods to models for diagnosing personal traits and cognitive abilities by social network data // Sci. Rep. 2024. Vol. 14. Article 5369. DOI:10.1038/s41598-024-56080-8
8. *Panfilova A.S., Valueva E.A., Ilyin I.Y.* The application of explainable artificial intelligence methods to models for automatic creativity assessment // Front. Artif. Intell. 2024. Vol. 7. Article 1310518. DOI:10.3389/frai.2024.1310518
9. *Park J.S., Zou C.Q., Shaw A., Hill B.M., Cai C., Morris M.R., Willer R., Liang P., Bernstein M.S.* Generative agent simulation of 1 000 people. 2024. DOI:10.48550/arXiv.2411.10109



References

1. Valueva E.A., Grigor'ev A.A., Lapteva E.M., Panfilova A.S., Ushakov D.V. Kognitivnaya slozhnost' i kommunikativnyj kontekst: otrazhenie intellekta pol'zovatelej v tekstah social'nyh setej. *Psihologicheskij zhurnal*, 2023. Vol. 44, no. 1, pp. 70–80. DOI:10.31857/S020595920024352-3
2. Rafikova A.S., Valueva E.A., Panfilova A.S. Golos i psihologicheskie svoystva cheloveka: obzor sovremennyh issledovaniy. *Psihologiya. ZHurnal Vysshej shkoly ekonomiki*, 2022. Vol. 19, no. 1, pp. 195–215. DOI:10.17323/1813-8918-2022-1-195-215
3. Stepanov S.Yu., Orzhekovskij P.A., Ushakov D.V. Ocenka uchenika: na puti k cifrovomu obrazovaniyu. Konceptual'no-matematicheskaya model'. *Narodnoe obrazovanie*, 2019. No. 1, pp. 130–139.
4. Ushakov D.V. Iskusstvennyj intellekt kak instrument psihologicheskogo issledovaniya. *Sibirskij psihologicheskij zhurnal*, 2023. No. 90, pp. 188–200. DOI:10.17223/17267080/90/11
5. Ushakova T.N., Pavlova N.D., Latynov V.V., Sepcov V.A., Alekseev K.I. Slovo v dejstvii. Intent-analiz politicheskikh diskussij. M.: 2000.
6. Eugen K. von, Endepols H., Drzezga A., Neumaier B., Gunturkun O., Backes H., Strockens F. Avian neurons consume three times less glucose than mammalian neurons. *Current Biology*, 2022. Vol. 32, pp. 4306–4313.
7. Panfilova A.S., Turdakov D.Y. Applying explainable artificial intelligence methods to models for diagnosing personal traits and cognitive abilities by social network data. *Sci. Rep.*, 2024. Vol. 14, Article 5369. DOI:10.1038/s41598-024-56080-8
8. Panfilova A.S., Valueva E.A., Ilyin I.Y. The application of explainable artificial intelligence methods to models for automatic creativity assessment. *Front. Artif. Intell.*, 2024. Vol. 7, Article 1310518. DOI:10.3389/frai.2024.1310518
9. Park J.S., Zou C.Q., Shaw A., Hill B.M., Cai C., Morris M.R., Willer R., Liang P., Bernstein M.S. Generative agent simulation of 1 000 people. 2024. DOI:10.48550/arXiv.2411.10109

Информация об авторах

Ушаков Дмитрий Викторович, академик РАН, директор, Институт психологии Российской академии наук (ФГБУН ИП РАН), г. Москва, Российская Федерация, ORCID: <https://orcid.org/0000-0001-9716-1545>, e-mail: dv.usakov@gmail.com

Information about the authors

Dmitry V. Ushakov, Effective Member of the Russian Academy of Sciences, Director, Institute of Psychology of the Russian Academy of Sciences, Moscow, Russia, ORCID: <https://orcid.org/0000-0001-9716-1545>, e-mail: dv.usakov@gmail.com

Получена 28.10.2024

Принята в печать 01.12.2024

Received 28.10.2024

Accepted 01.12.2024