

УДК: 303.2, 316.6

### 1.3. Социальные представления российского общества об искусственном интеллекте: пилотное исследование (часть 1)

Ноакк Н.В., Волкова А.Д., Костина Т.А.  
ЦЭМИ РАН, Москва, России.

*Дано описание пилотного исследования социальных представлений российской выборки об искусственном интеллекте с использованием метода простых свободных ассоциаций и структурного анализа по П. Вержесу. Наличие амбивалентности социальных представлений об исследуемом объекте как на уровне больших групп, так и на уровне отдельной личности может привести к резкой социальной поляризации. Отсутствие концептуализации искусственного интеллекта у большинства респондентов, понимания его реальных угроз, чёткого представления о границах проникновения ИИ в практику повседневного опыта, этических и социальных аспектах его внедрения говорит о том, что эти аспекты в настоящее время упускаются из виду в общественном дискурсе. Негативная эмоциональная оценка последствий внедрения искусственного интеллекта больше свойственна респондентам, отдающим предпочтение ценностям общества. Присутствие архетипических мотивов в ответах выводит проблематику ИИ на уровень глубинных общечеловеческих проблем развития человечества и цивилизации в целом.*

#### Введение

Перефразируя известное высказывание, можно с уверенностью утверждать: по планете бродит призрак искусственного интеллекта. О нем сообщается в новостях, пишутся романы, снимаются кинофильмы; лавинообразно растёт количество конференций, публикаций, видеокурсов и интервью.

Актуальность исследования вызвана как недостаточностью разработки темы в отечественных научных источниках, так и стремительным развитием новых цифровых технологий, в частности, искусственного интеллекта, и в связи с этим – происходящими изменениями и в целом человеческого общества, и отдельного человека, системы его коммуникаций, его качеств и психических функций. Как справедливо указано в [Ясин, 2022] «на настоящий момент разработки в данной области остаются недостаточными, многие аспекты этого взаимодействия не раскрыты с точки зрения психологии», несмотря на то, что в последнее время появляются статьи и книги обзорного характера, преимущественно, зарубежные (см. [Montag, Diefenbach 2018], раскрывающие различные аспекты воздействия цифровых миров на человеческую психику, включая фундаментальные трансформации в саморефлексии и самооценке, а также формирование «я» [Montag 2018; Diefenbach, Christoforakos 2017], влияние на эмоциональные потребности и человеческий мозг, на внимание [Sariyska, Reuter, Lachmann, Montag 2015], на систему коммуникаций, на возникновение стресса, тревожности и депрессивных состояний [Riedl, 2012; Elhai, Dvorak, Levine, Hall, 2017]. Итак, несмотря на лавинообразно увеличивающийся поток научных публикаций по теме искусственного интеллекта (далее - ИИ), как зарубежных, так и особенно в последнее время отечественных [Войскунский, 2010, 2018; Ясин, 2022], а также в значительной степени - работы Т.А. Нестика, А.Л. Журавлева и других авторов (около 22 тысяч публикаций на страницах e-library), вопросы структуры представлений россиян относительно этого явления пока не завоевали должного внимания в научных публикациях. Требуют исследования, на наш взгляд, такие аспекты указанной выше темы, как влияние средств массовой информации на формирование социальных представлений об ИИ, действие культурного (в частности, литературного) дискурса на содержание социальных представлений, связь последних с архетипическими образами и мотивами. Кроме того, мы не нашли развернутых исследований, построенных на концепции социальных представлений, используемой нами и предлагающей, предположительно, значительные возможности для изучения новых социальных явлений. Исключением является работа [Нестик, 2018], где автор, используя метод ассоциативного эксперимента, предлагал респондентам ответить на вопросы о возможностях и рисках, возникающих в связи с развитием искусственного интеллекта.

Каковы социальные представления россиян об искусственном интеллекте? Связываются ли они многочисленными ассоциациями со значимыми сферами жизни личности и матричным слоем коллективного бессознательного, что, полагаем, доказывает более серьёзное отношение к явлению ИИ? Или находятся на поверхности сознания в состоянии расхожих штампов – вне существенной связи с разнообразным опытом человека (как онтогенетическим, так и филогенетическим)?

Некоторые акценты нашей статьи обусловлены более ранними нашими публикациями, посвящёнными новым цифровым технологиям. [Ноакк, Костина, 2021а, 2021б]. Кроме того, выбор предмета исследования, как и основной концепции и методологии, также продиктован многолетними исследовательскими интересами авторов в области отношения людей к цифровым продуктам.

В описанном ниже пилотном исследовании мы ставим несколько задач – выявить социальные представления (далее – СП) россиян различного возраста, пола, социального и финансового положения об этом новом (уже не совсем новом) явлении, проявлениях искусственного интеллекта в их жизни и деятельности, о последствиях, рисках и преимуществах его использования; проанализировать структуру полученных социальных представлений в соответствии с выбранной методикой, выявить ядровые и периферийные элементы, основные темы и категории; вычленив эмоциональные компоненты в ответах;

соотнести полученные данные с ценностными предпочтениями респондентов. Отдельной задачей стоит выявление архетипических мотивов, тем и образов в ответах респондентов.

Пилотное исследование, как мы предполагаем, поможет определить наиболее перспективные направления будущих (более узких и более глубоких) исследований, а также выдвинуть гипотезы, требующие последующего анализа. Результаты исследования будут представлены в двух статьях (частях). Часть 1 посвящена общим вопросам методологии, а также некоторым ключевым темам и категориям, определившимся в ходе первоначального анализа. В Части 2 будет описан структурный анализ данных с использованием метода П. Вержеса.

### **Описание исследования, методологии и выборки, основных результатов**

#### **Методология.**

В основе нашего исследования лежит концепция С. Московичи, имеющая значительный интерпретационный потенциал, который позволяет использовать ее для решения задач по выявлению и анализу СП пользователей о новых продуктах цифровой реальности. Теория социальных представлений, предложенная С. Московичи, является ключевой социально-психологической концепцией, существующей почти шесть десятилетий. Подход имеет богатый потенциал для анализа трансформаций, происходящих с человеком в цифровом мире. Теория СП позволяет ответить на вопрос, как человек строит объяснение новому явлению и своему поведению в соответствии с ним. [Емельянова, 2016]. Привлекательность этой концепции объясняется еще и тем, что объектом анализа исследователей становятся не только концепции, идеи, овладевающие людьми, но и образы, символы и метафоры. Это позволяет выявить более глубокие слои психического, связанные с взаимодействием сознательной и бессознательной сфер. Одним из несомненных преимуществ концепции является акцент на динамике СП, рассмотрении их как изменяющегося в ходе развития реальности феномена, выделении фаз зарождения, стабилизации, трансформации [Ноакк, Костина, 2021а].

В качестве методов сбора данных был использован опрос (анкета), в который одним из пунктов вошла методика простых свободных ассоциаций. Из методов обработки полученных данных были использованы контент-анализ, статистический анализ, метод структурного анализа по П. Вержесу.

Характер поставленной задачи продиктовал жанр изложения – описание полученных результатов с точки зрения целесообразности углубления будущего поиска. Отдельный параграф посвящен анализу структуры социальных представлений по методике П. Вержеса, но, как оказалось, этот анализ не вмещается в рамки одной статьи.

#### **Описание методики сбора данных**

Исследование проводилось онлайн, с использованием цифровой платформы [anketolog.ru](http://anketolog.ru). Для сбора данных была использована специально разработанная методика.

Назначение методики – выявить социальные представления респондентов об искусственном интеллекте; осуществить контент-анализ полученных данных, выявить основные темы и категории, а также общий эмоциональный фон (позитивные/негативные значения). Соотнести полученные данные с оценкой респондентами значимости определённых ценностей, личных и общественных, а также с рядом социально-психологических характеристик, в частности, возрастными.

Структура. Анкета состояла из 4 структурных частей. Первая часть включала в себя ряд пунктов, касающихся социально-психологических характеристик (возраст, гражданство, пол, образование, род деятельности); вторая была посвящена ассоциативному опросу по социальным представлениям респондентов об искусственном интеллекте, практике его использования в повседневной жизни; в третьей респондентам предлагалось дать ответы на вопрос «Кто я»; в четвертой необходимо было оценить значимость ряда предлагаемых ценностей. [Волкова и др., 2021].

Анкета включала в себя как общую инструкцию, так и инструкции к каждой из структурных частей. Общая инструкция представлена ниже.

*Здравствуйте! Выделите, пожалуйста, 15 минут своего времени на заполнение следующей анкеты. Это научное исследование. Гарантируем Вам полную конфиденциальность предоставленной Вами информации. Анкета полностью анонимна, а результаты исследования будут использоваться исключительно в обобщенном виде после статистической обработки. Важно: отвечайте по возможности быстро, не задумываясь.*

Пример инструкции ко второй части анкеты, касающейся ассоциативного опроса по социальным представлениям об искусственном интеллекте, приведен ниже.

*1 Напишите, пожалуйста, не менее 5 ассоциаций к слову / словосочетанию, «Искусственный интеллект». Грамматическая форма слова-ассоциации может быть любой – существительное, глагол, прилагательное, наречие.*

*2 Напишите, пожалуйста, примеры искусственного интеллекта, которые Вы знаете.*

*3 Выберите, пожалуйста, пример искусственного интеллекта из предыдущего вопроса и укажите, как часто Вы его используете.*

Пример инструкции к четвертой части анкеты представлен ниже.

*Оцените от 1 до 7 значимость для Вас следующих ценностей (1 минимальная, 7 – максимальная значимость). Для каждой ценности необходимо выбрать только одно цифровое значение.*

Профессиональный рост  
Здоровье, эмоциональное состояние, отдых  
Общение с родными, коллегами, друзьями  
Материальное благополучие  
Хорошие бытовые условия  
Смысл жизни, моральные и личностные качества  
Состояние внешней среды  
Ценности общества  
Семья, любовь, близкие люди  
Автономия

#### Полученные результаты. Обсуждение.

Опрос проводился среди так называемых «обычных пользователей», не обладающих специальным знанием в области информационных технологий. Выборка включала 100 человек.

Всего было получено 376 ассоциаций. Такое количество определяется тем, что некоторые респонденты приводили не 5 ассоциаций, как того требовала инструкция, а меньше. Под единицей (ассоциацией) в данном случае понимается и одно слово (например, польза), и в ряде случаев словосочетание и/или несколько словосочетаний: *повышенная волокита; тупая бюрократия, без возможности понять или разобраться* – то есть некоторая структурно-смысловая единица. В ходе контент-анализа и структурного анализа был выделен ряд тем и категорий: *Архетипы, Позитивные ассоциации, Негативные ассоциации, Антропоморфизм, Время, Влияние на человечество* и др. Подробный разбор выделенных тем, как и их критериев будет дан в следующей публикации. Здесь остановимся на одной из наиболее отчётливо представленных – теме (категории) *будущего*.

#### ИИ и категория будущего

Как справедливо указано в [Журавлев, Нестик, 2019], формирование цифровой экономики происходит в ситуации «шока будущего», когда ещё отсутствуют психологические механизмы договорного процесса относительно новых цифровых явлений, которые помогли бы и большим социальным группам, и конкретным личностям выработать правила «присвоения», овладения новыми явлениями. Это было одной из причин нашего внимания к данной категории. Другой причиной было то, что в ходе анализа нами была выделена сравнительно гомогенная группа респондентов, употребивших категорию *будущего* в своих ассоциативных рядах: 15 человек, 11 женщин, 4 мужчин. Возраст: 6 чел. – 17-30 лет; 5 чел. - 31-40 лет; 3 человека - 41-50 лет; 1 чел. – 51-60 лет. Как уже было сказано, в отличие от выборки в целом, эта группа демонстрирует большее единство в используемых ассоциациях. Начать с того, что ассоциации здесь несут в целом или позитивный, или нейтральный характер. Далее, почти у 60% респондентов этой группы употребляется в ассоциативных рядах, наряду с *будущим*, ассоциация *робот*. Встречаются также: *машина, компьютер, алгоритм, сервис* - то есть подчёркиваются технические/технологические («не-человеческие») аспекты ИИ. Можно предположить, что будущее развитие ИИ связывается значительной частью данной выборки с развитием робототехники. Возможно, что это обусловлено представленностью данной тематической связи в художественных произведениях (литературных и кино-), в материалах СМИ. Респондентам, скорее всего, легче представить ИИ как некоторый технический объект, имеющий визуальное сходство с человеком, но тем не менее всего лишь машину, которую и создал человек, и он же может ею управлять и контролировать ее работу. В целом, ассоциации наших респондентов, употребляющих категорию *будущего* в своих ответах, носят достаточно неопределённый характер. Это согласуется с данными ряда исследований, в частности, общероссийских опросов Левада-центра. Согласно последним, будущее представляется россиянам как «смутное», «призрачное», «туманное», «неопределённое». Одна из причин такой неопределённости может состоять в том, что будущее страны мало связано с личным будущим [Нестик, 2021].

Присутствие ассоциации *робот* в социальных представлениях о будущем вполне объяснимо – создатели произведений литературы и искусства (в основном, кино) уже давно разрабатывают эту тему, мягко воздействуя на умы (воображение) читателей и зрителей, готовя их к спокойному принятию, одобрению, освоению этих явлений.

Большая часть российской читательской и зрительской аудитории, безусловно, знакома с рассказами и романами Айзека Азимова, как и с кинофильмами, снятыми по ним. Роботы в его произведениях существуют рядом с людьми, взаимодействуют с ними, выполняют различные, сложные в том числе, функции, в главном и целом помогая человеку в наиболее трудных задачах. При этом, что очень важно, они остаются в полном его подчинении. Даже если в некоторых случаях они начинают делать что-то самостоятельно или, казалось бы, открыто конфликтовать с человеком - как, например, в рассказе «Логика», когда робот наотрез отказывается считать, что он создан людьми, утверждая, что его создал Господь. Заметим, что Азимов доводит ситуацию до предела: робот не только провозглашает своё превосходство над людьми по интеллектуальным и физическим возможностям, но и заключает двух главных героев под домашний арест, при этом беря руководство космической станцией на себя, как и выполнение важной миссии – предотвращение катастрофы на Земле. Но со своей задачей он справляется блестяще: оказывается, что он таким образом следовал Первому закону робототехники, так как выполнил эту задачу точнее, чем люди. А то, что он говорит и делает в отношении двух испытателей – в данном случае не

играет никакой роли. В другом рассказе, когда робот открыто не повинует человека и даже, казалось бы – угрожает его жизни («Как потерялся робот») сами люди изменяют Первый закон так, что робот «может своим бездействием допустить причинение вреда человеку». В результате появляется робот, который может «обидеться» на своего создателя, то есть фактически наделённый человеческими эмоциями. Более того, он даже может попытаться силой остановить человека. Но и в этом случае автор от лица главной героини утверждает – робот не может нанести человеку вред.

Интересно, что, несмотря на то, что речь идёт о литературных произведениях и вымышленном мире – в социальных представлениях этот вымышленный мир может терять свои границы с реальным миром, то есть восприниматься как *объективная реальность* [Войскунский, 2010, 2018]. Причём, речь идёт не о том процессе, который А.Войскунский в своем масштабном труде причислял к феноменам Presence и который часто присутствует при восприятии интересной захватывающей книги.

Как показали исследования, сдвиг к консервативным ценностям и авторитарным установкам снижает готовность к диалогу по поводу совместного будущего.

Коллективный образ будущего — это представления о будущем социальной группы, разделяемые ее членами. [Нестик, 2021 ] Нам показалось чрезвычайно интересным выделение в содержании коллективного образа будущего следующих шести феноменов: *Страхи, Цели и планы, Надежды, Ожидания, Мечты, Идеалы*. Мы решили применить к ассоциативным рядам наших респондентов перечень значений этих феноменов и вот что получили (в некоторых случаях, в силу невозможности однозначного отнесения ассоциативного ряда к определённому феномену, обозначали несколько феноменов для каждого анализируемого ряда; цифры обозначают количество определяемых как относящихся к данному феномену ассоциативных рядов социальных представлений).

*Страхи* (понимается как разделяемые членами группы представления об угрозах для индивидуального и коллективного благополучия или даже для самого существования группы и организации) – 3; *Цели и планы* (понимается как объединяющие и координирующие усилия группы в ходе совместной деятельности) – 7; *Надежды* (понимается как представления членов группы о вероятном положительном для них исходе уже наблюдаемых или ожидаемых событий ближайшего и среднесрочного будущего) – 4; *Ожидания* (понимается как представления о значимых событиях ближайшего будущего, которые члены группы не могут контролировать, но стремятся учесть в своих планах) – 8; *Мечты* (представления о желательных для группы изменениях в обществе, которые не предполагают немедленной и полной реализации в действительности); – 3; *Идеалы* (представления о принципиально недостижимой, но желательной для группы ситуации). – 0. Таким образом, можно отметить как наиболее встречаемые в этой группе феномены коллективного будущего: 1. *Цели и задачи* (46%); и 2. *Ожидания* (54%). Хотелось обратить внимание на первый из указанных феноменов (*Цели и задачи*) – он может говорить о том, что респонденты этой группы ориентируются в своих представлениях на ситуации совместной с ИИ деятельности, в которых может понадобиться взаимодействие, распределение усилий и соответствующих задач между человеком и ИИ. Частота употребления второго из феноменов (*Ожидания*), как видно из его значения, может предполагать *невозможность контроля за деятельностью ИИ со стороны человека*. Здесь мы видим проявление той противоречивости (амбивалентности) социальных представлений об ИИ респондентов нашей выборки, о которой пойдёт речь ниже.

В общем и целом, полученные в ходе изучения описанной группы респондентов результаты дают возможность выдвинуть ряд гипотез, которые потребуют проверки в следующих исследованиях.

- респонденты, употребляющие категорию будущего, склонны видеть в ИИ не самостоятельную сущность со своим интеллектом, а преимущественно машину; употребляемые «человеческие» характеристики (мыслит, мозг, беспристрастность) не меняют в целом указанного значения ассоциативного ряда;
- респонденты этой группы относят развитие и функционирование ИИ в будущее, тем самым как бы откладывая возможные проблемные вопросы на потом;
- те респонденты, которые употребляют категорию будущего при описании ИИ, склонны избегать ответственности за социальные явления, игнорируют необходимость сиюминутного решения злободневных социальных проблем;
- россияне этой группы склонны видеть за ИИ будущее в значении прогрессивного развития технологии;
- респонденты данной группы акцентируют позитивные аспекты в развитии ИИ. Возможная опасность предполагается имплицитно, но высвечивается наряду с очевидной пользой для человека (имеется в виду, например, такой ассоциативный ряд (№18): будущее, польза, опасность, персональные данные, плагиат, где польза – на 2-ом месте, в ядровой компоненте, опасность и другое – в периферии);

#### **ИИ и возраст, эмоциональная оценка (лингвистический анализ)**

Здесь представлены результаты анализа полученных данных по критерию позитивной и негативной оценки ИИ (Таблица 1). Как мы видим, большая часть ассоциаций об ИИ во всех возрастных группах нашей выборки носила нейтральные значения. Наибольшее число ассоциаций позитивного значения обнаружено в возрастной группе 41–50 лет – 28%. Наименьшее – в группе 31–40 лет – 13%. Возрастные группы 17–30 лет и 51–80 лет представлены 22% и 18%, соответственно. Что касается ассоциаций негативного характера, то наибольшее число обнаружено в группе 51–80 лет – 18% всех ассоциаций, наименьшее – в группе 17–30 лет – 3% всех ассоциаций. Второе и третье места по количеству

употреблённых ассоциаций негативного значения разделили группа 41–50 лет и группа 31–40 лет, 16% и 12%, соответственно. Забегая вперед, следует сказать, что при осуществлении структурного анализа полученных ассоциаций по П. Вержесу, выявлено, что негативные ассоциации входят в Ядро структуры социальных представлений. Задача изучения эмоциональных компонентов отношения россиян к искусственному интеллекту требует более развернутых исследований.

**Таблица 1. Эмоциональная оценка и лингвистические характеристики ассоциаций в возрастных группах.**

Группа	Кол-во респондентов	Всего ассоциация	Позитивные ассоциации	Нейтральные ассоциации	Негативные ассоциации	Конкретность	Абстрактность	Соотнесённость с человеком	Грамм. форма - существительное	Грамм. форма - другое
17-30	17	68	22%	75%	3%	1%	99%	9%	65	12
31-40	36	134	13%	75%	12%	4%	96%	10%	134	27
41-50	25	82	28%	56%	16%	1%	99%	5%	66	31
51-80	16	60	18%	63%	18%	8%	92%	5%	64	59

В ходе исследования было осуществлено разделение ассоциаций по таким лингвистическим характеристикам, как конкретность-абстрактность значений, грамматическая форма (существительное/другое), соотнесённость значения с человеком. Сравнение не показало видимых различий в возрастных группах по первой из указанных характеристик.



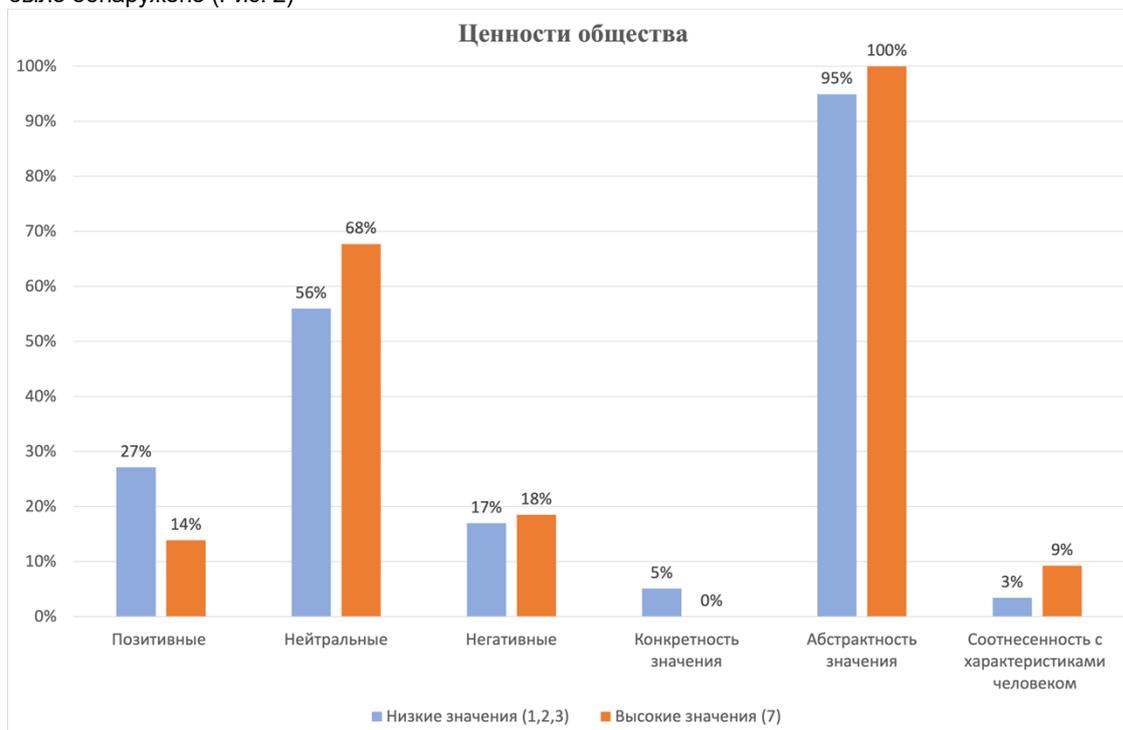
**Рис. 1 Эмоциональная оценка**

Грамматическую форму существительного в ассоциативных рядах употребляют почти вдвое больше респондентов возрастной группы 31-40 лет, по сравнению с другими группами; эти же респонденты, как и участники группы 17-30 лет, вдвое превосходят две другие группы по процентному соотношению ассоциаций, связанных с человеком и его характеристиками, хотя 10% и 9% - это не очень много (по 14 и 7 ассоциаций из 134 и 68 ассоциаций, соответственно). В целом по выборке можно отметить превалирование абстрактных значений употребляемых ассоциаций в ущерб конкретным (Рис. 1).

**ИИ и ценности**

Одной из гипотез нашего пилотного исследования было предположение, что предпочтение респондентами определённых ценностей из предлагаемого нами списка будет связано с эмоциональной окраской их ассоциаций. Мы взяли значения респондентов по ценности «Ценности общества», причём в качестве высоких приняли значение в 7 баллов (предельный балл), в качестве низких – значения 1, 2, 3 балла. Были выделены 2 группы с соответствующими значениями. Обнаружено, что у респондентов, проставивших низкие баллы по ценности «Ценности общества», ассоциаций с позитивными значениями в 2 раза больше, чем соответствующих ассоциаций в группе респондентов, проставивших высокие значения по соответствующей ценности (28% и 14%, соответственно). Иначе говоря, россияне, для которых ценности общественной жизни имеют важное значение, более склонны негативно оценивать последствия ИИ, по сравнению с группой россиян, оценивших данные ценности как менее значимые.

Эти же 2 группы были проанализированы по критериям – наличия ассоциаций с конкретными /абстрактными значениями и соотнесённости с «человеческими характеристиками». Видимых различий не было обнаружено (Рис. 2)



**Рис. 2** Связь значимости ценности «Ценности общества» с позитивными, нейтральными, негативными значениями социальных представлений, абстрактностью-конкретностью, соотнесённостью с характеристиками человека

#### ИИ и архетипические мотивы.

Ряд ассоциаций респондентов нашей выборки об ИИ связан с архетипическими образами естественности/искусственности, зла, опасности, угрозы/добра, пользы, помощи, ума/глупости, человечности, души/ бесчеловечности, бездушия: понятие архетипических образов употребляется нами в юнгианском понимании как неких коллективных паттернов, типов, образцов архаического характера, включающих мифологические мотивы, принадлежащих человечеству в целом и, таким образом, коллективных по своей природе [Юнг, 1994].

Так, почти 20% выборки (21 человек) подчеркивает в своих ответах архетип искусственности (неестественности)- в разном контексте – прямое называние (*неживой, неестественный, искусственный разум*), указание на отсутствие чувств, человечности, отсутствие души (*полное отсутствие человечности, без чувств, бездушный*)); на то, что это создано человеком (*сделанный предмет, синтезированный, искусственное сознание, созданное людьми*). Сюда же можно отнести ассоциации, объединённые нами в категорию реальности/нереальности, вымышленности/подделки: *не настоящий предмет, реальность, вымысел, подделка, не существующий*.

Такие ассоциации, как *война, смерть, опасность, угроза* позволяют предположить, что размышления и рассуждения об ИИ идут на глубинном, общечеловеческом уровне. Здесь задействованы, как утверждает К. Юнг, процессы бессознательного, а значит, коллективное бессознательное начинает констеллироваться, что может привести к большим социальным потрясениям: «Подобные движения очень заразительны – заражение происходит потому, что во время активации коллективного бессознательного человек перестаёт быть самим собой. Он не просто участвует в движении, он и есть само движение» [Юнг, 1994].

Связь социальных представлений об ИИ и архетипических мотивов требует, безусловно, более развернутого анализа, что и будет осуществлено в следующей статье.

#### ИИ и антропоморфизм.

Другая категория, которая выделилась в процессе контент-анализа ассоциаций на понятие ИИ и также связана с архетипическими мотивами, может быть названа антропоморфизмом— уподоблением человеку, наделением ИИ человеческими качествами.

• *мозг, думает, (искусственный) разум, помощник, глупость, умно/-ый, беспристрастность, мыслить, (технический) разум, симпатичный, дружелюбный, всезнающий, ум, безмозглый, решает, добрый, честный, сильный, (робот) умница, мозги (искусственные), умный (компьютер), без чувств.*

Можно сказать, что этими респондентами ИИ воспринимается как некое существо, личность, персоналия, обладающая характером, мышлением и другими человеческими/личностными признаками.

Среди данного ряда ассоциаций в свою очередь можно выделить подгруппу ассоциаций, касающихся умственных характеристик (*мозг, разум, умный, глупость, мыслить*), т.е. описывающих интеллектуальную деятельность именно в человеческом понимании. С другой стороны, само понятие «ИИ» содержит значение интеллектуальной деятельности, поэтому эти ассоциации могут быть вполне закономерны.

Что нам кажется более любопытным результатом – это то, что часть ассоциаций касается не интеллектуальных функций, а качеств характера и прочих «человеческих» характеристик (*помощник, беспристрастность, симпатичный, дружелюбный, добрый, честный, сильный*). Таким образом, комплекс «антропоморфических» ассоциаций к понятию ИИ становится более разносторонним, респонденты данной подгруппы практически наделяют ИИ характером, моральными качествами, социальными ролями.

Продолжая рассуждение, можно предположить, что таким образом у человека может возникать отношение к ИИ не как к программе, инструменту, предмету, чему-то неодушевленному, а как к личности; могут формироваться соответствующие паттерны отношения и поведения по отношению к ИИ – обида, радость, жестокость, уважение и т. п.

#### **ИИ и структурный анализ по П. Вержесу.**

Одним из наиболее развитых подходов в рамках концепции СП является структурный подход. В структуре социальных представлений выделяют три компонента: информация, отношение и поле репрезентации. Структурный анализ в узком смысле этого понятия предполагает выделение Ядра СП и периферийной части. В отличие от результатов исследования, представленного в [Нестик, 2018] мы намеренно не разделяли вопрос об ИИ на «возможности» и «угрозы», чтобы посмотреть, как сами респонденты видят позитивные и негативные стороны исследуемого явления.

Структурный анализ социальных представлений об ИИ россиян, вошедших в выборку, был принят применительно ко всей нашей выборке. Именно с этим, на наш взгляд, связан противоречивый характер полученных по нему результатов. Им будет посвящена отдельная статья (часть 2). Здесь укажем, что и Ядро, и Периферия структуры социальных представлений о слове/словосочетании «Искусственный интеллект» в ответах респондентов несут на себе сильный отпечаток двойственности, амбивалентности. Этой двойственностью буквально пропитаны все части структуры. Это и сильно мешает, и в то же время даёт направление будущего анализу. Так, в Ядро социальных представлений (напомним, что здесь располагаются ассоциации, получившие первые ранги (поставленные респондентами на первые места) и наиболее часто встречающиеся в нашей выборке) наряду с ассоциациями, обозначающими робота и компьютер, как и их элементы/функционал (*программа, алгоритм, прибор, провода, интернет, сеть*), то есть технические/технологические характеристики – входят такие ассоциации, как *мозг, разум*, связанные с чисто человеческими характеристиками. В зону ближней периферии структуры социальных представлений, наряду с категориями, связанными с выражением общего недоверия к ИИ, страхами и сомнениями (*подделка; плагиат; против; полное отсутствие человечности; тревога; репрессии; сбой*) – входит тематика вера в возможности ИИ и прогресс, с ним связанный (*будет развиваться; будущее; научный прогресс; научный прорыв*). Дальняя Периферия также представлена категориями с амбивалентным значением – с одной стороны, здесь присутствует тематика акцентирования искусственности происхождения ИИ, с другой – он (ИИ) наделяется характеристиками, которые больше подходят человеку, другу – *любимый, симпатичный, красивый*. Амбивалентность социальных представлений об ИИ может быть знаком сосуществования в российском обществе полярных социальных групп в отношении к ИИ, что, в свою очередь, может привести к конфликтам и/или социальным противостояниям, а также создавать значительные трудности в ситуациях, требующих взаимодействия. Антропоморфизм в отношении ИИ позволяет выдвинуть гипотезу о размывании границ естественного/искусственного относительно ИИ, что не может не вызывать опасений со стороны психологов.

В целом можно сделать вывод, что структура социальных представлений об ИИ в российском обществе не сложилась, носит противоречивый характер, находится в динамике становления. Выскажем предположение, что утверждение автора концепции С.Московичи о фазах становления структуры социальных представлений (зарождения, стабилизации, трансформации) в случае с ИИ может не работать в традиционном смысле, поскольку изменения происходят постоянно и немислимо стремительно и границы между фазами размываются.

Целесообразными представляется с этой точки зрения проведение структурного анализа не в целом по всей выборке, а в отдельных не очень больших социальных группах. В этом случае, возможно, можно будет более четко определить структурные особенности социальных представлений различных групп, как и взаимодействие этих групп друг с другом. Так, например, было бы интересным сделать структурный анализ социальных представлений об ИИ в группе респондентов, употребивших ассоциацию *будущее* в своих ассоциативных рядах, а также в группе людей, профессионально связанных с тематикой искусственного интеллекта.

#### **Этические/моральные (социальные) аспекты отношения к ИИ.**

В ходе пилотного исследования мы задались вопросом, как наши респонденты представляют себе последствия внедрения ИИ, связанные с этическими/моральными аспектами. Предположительно, ряд респондентов в своих ассоциативных рядах должны были коснуться этих вопросов.

В эту категорию мы отнесли следующие ассоциации: *новое опасное увлечение человечества, опасность, слежка, захват, обман, репрессии, безработица*. Как мы видим, некоторых наших респондентов беспокоят такие аспекты социальной жизни, как возможность потери рабочих мест из-за замены людей искусственным интеллектом; как генерирование ложной информации, в том числе в СМИ; как, видимо, опасность попадания ИИ не в те руки и связанные с этим наблюдения за людьми и/или прослушивание разговоров; как, наконец, возможность несправедливого наказания. В целом ассоциации этой категории составляют очень малый процент от общего количества. Можно утверждать, что в ассоциациях выборки уделено мало внимания вопросам этики, морали и социальных последствий внедрения ИИ – свидетельство того, что эти аспекты в настоящее время упускаются из виду в общественном дискурсе.

### **Заключение**

Полученные ассоциации отражают общественное восприятие и чувства разных социальных групп российского общества по отношению к технологии искусственного интеллекта. Россияне воспринимают ИИ в первую очередь как технологию будущего, которая может принести как прогресс, так и потенциальные опасности. Это отражает общую тревогу и одновременное восхищение перед быстро развивающейся технологией.

Противоречивость свойственна отношению к ИИ в плане оценки некоторых его характеристик (параметров). Ряд респондентов характеризует ИИ как инструмент, описывая параметры его скорости, полезности, удобства. Другие склонны проводить сопоставление ИИ с человеком – акцентируя его «человеческие/нечеловеческие» характеристики, тем самым вводя тематику ИИ в антропоморфный дискурс, отождествляя ИИ с человеческими качествами или способностями. Таким образом, мы наблюдаем поляризованность восприятия ИИ – и как инструмента и/или технологии, и как сущности с собственным сознанием и чувствами.

Несмотря на множество позитивных ассоциаций, таких как *помощник, удобный, прогресс*, существуют и ряды негативных ассоциаций, таких как *опасность, глупость, смерть*. Это также указывает на двойственное восприятие ИИ в обществе: с одной стороны, это полезный инструмент, с другой — потенциальная угроза. Тем не менее количество ассоциаций во втором значении в общем числе невелико, что может говорить об отсутствии у большинства выборки (возможно, и у россиян в целом) понимания реальных угроз. Ассоциации не включают многие технические и специализированные аспекты ИИ, что может свидетельствовать о том, что у большинства россиян нашей выборки отсутствует глубокое понимание этой области. В ассоциациях выборки уделено мало внимания вопросам этики, морали и социальных последствий внедрения ИИ – свидетельство того, что эти аспекты в настоящее время упускаются из виду в общественном дискурсе.

Выводы: социальные представления об ИИ россиян из числа неспециалистов по информационным технологиям, вошедших в выборку, носят противоречивый характер. Предположительно, отношение к искусственному интеллекту находится в фазе становления. Возможно, оно резко поляризовано, разнесено по социальным (общественным) группам, что может привести к конфликтным ситуациям в будущем. Обращает на себя внимание отсутствие детального и глубокого понимания и концептуализации искусственного интеллекта у большинства людей, понимания его реальных угроз, чёткого представления о границах проникновения ИИ в практику повседневного опыта, этических и социальных аспектах его внедрения. Полезность и вред ИИ воспринимается «в целом», что может свидетельствовать, скорее, о воздействии социального стереотипа. Представления о будущем развитии ИИ как о роботизированном также могут отражать влияние литературного и кинематографического дискурса.

Выявлено, что россияне, для которых ценности общественной жизни имеют важное значение, склонны более негативно оценивать последствия ИИ. Связь ряда затронутых в ответах тем с архетипическими мотивами выводит проблематику ИИ на уровень глубинных общечеловеческих проблем развития человечества и цивилизации в целом.

В качестве будущих перспективных направлений изучения искусственного интеллекта укажем следующие: архетипические образы в социальных представлениях; психологические механизмы формирования социальных представлений об искусственном интеллекте, в частности, появления антропоморфических тенденций в его восприятии; роль эмоциональных характеристик в отношении россиян к ИИ; структурный анализ социальных представлений об ИИ в различных (по полу, возрасту, социальному и финансовому положению, профессии) социальных группах.

### **Литература**

1. Войскунский А.Е. Психология и интернет. – М.: Акрополь. 2010. – 439 с.
2. Войскунский А. Е. Психология и искусственный интеллект: новый этап взаимодействия // Психология человека как субъекта познания, общения и деятельности / под ред. В. В. Знакова, А. Л. Журавлева. М.: Издательство Института психологии РАН, 2018. С. 2094–2101.
3. Емельянова Т.П. Социальные представления: История, теория и эмпирические исследования. – М.: Изд-во «Институт психологии РАН», 2016. – 476 с.
4. Журавлев А. Л., Нестик Т. А. Социально-психологические последствия внедрения новых технологий: перспективные направления исследований // Психологический журнал. 2019. № 5 (40). С. 35–47. <https://doi.org/10.31857/S020595920006074-7>

5. Нестик Т.А. Образ будущего, социальный оптимизм и долгосрочная ориентация россиян: социально-психологический анализ // Социодиггер, 2021, октябрь-ноябрь. Том 2. Выпуск 9(14). – Стр. 6-48.
6. Нестик Т.А. Социально-психологические предикторы отношения личности к искусственному интеллекту и роботизации // Социальная и экономическая психология. – 2018. – С. 406–428.
7. Нестик Т.А. Социально-психологические аспекты отношения человека к новым технологиям // Актуальные проблемы психологии труда, инженерной психологии и эргономики. – 2018. – С. 50–73.
8. Волкова А.Д., Иванова Н.Л., Патоша О.И. Социальная идентичность и предпочтения потребителей: исследование взаимосвязи (на примере продуктов питания отечественных производителей) // Национальный психологический журнал. 2021. № 4(44). С. 39–52. doi: 10.11621/npj.2021.0404
9. Ноакк Н.В., Костина Т.А. Эмпирическое исследование социальных представлений о смарт-счетчиках // Цифровая экономика, 1(13), 2021. – стр. 34–40.
10. Ноакк Н.В., Костина Т.А. Социальные представления о ИИ // Цифровая экономика, №2 (14), 2021. – стр. 24–30.
11. Юнг К.-Г. Аналитическая психология / Перевод и редакция В.В. Зеленского. Санкт-Петербург, МЦНК и Т «Кентавр». Институт Личности, ИЧП «Палантир», 1994. – 136 стр.
12. Ясин М. И. Представления молодежи об искусственном интеллекте и отношении к нему // Известия Саратовского университета. Новая серия: Философия. Психология. Педагогика. 2022. Т. 22, вып. 2. С. 197–201. <https://doi.org/10.18500/18197671-2022-22-2-197-201>
13. Diefenbach, S., Christoforakos, L.; Ullrich, D. Digitale Disbalance—Herausforderungen der Smartphone-Ära. *Wirtschaftspsychol. Aktuell* 2017, 3, 36–42.
14. Elhai, J.D.; Dvorak, R.D.; Levine, J.C.; Hall, B.J. Problematic smartphone use: A conceptual overview and systematic review of relations with anxiety and depression psychopathology. *J. Affect. Disord.* 2017, 251–259.
15. Montag, C. *Homo Digitalis*; Springer: Berlin, Germany, 2018; ISBN 978-3658200251
16. Montag C., Diefenbach S. Towards Homo Digitalis: Important Research Issues for Psychology and the Neurosciences at the Dawn of the Internet of Things and the Digital Society // *Sustainability*. 2018. Vol. 10 (2). P. 415–436. <https://doi.org/10.3390/su10020415>
17. Riedl, R. On the biology of technostress: Literature review and research agenda. *ACM SIGMIS Database* 2012, 44, 18–55.
18. Sariyska, R.; Reuter, M.; Lachmann, B.; Montag, C. Attention deficit/hyperactivity disorder is a better predictor for problematic Internet use than depression: Evidence from Germany. *J. Addict. Res. Ther.* 2015, 6, 1–6.
19. Sindermann, C.; Luo, R.; Zhao, Z.; Li, Q.; Li, M.; Kendrick, K.M.; Panksepp, J.; Montag, C. High ANGER and low agreeableness predict vengefulness in German and Chinese participants. *Personal. Individ. Differ.* 2018, 121, 184–192.

#### References in Cyrillics

1. Воjskunjij A.E. Psixologiya i internet. – М.: Akropol'. 2010. – 439 s.
2. Воjskunjij A. E. Psixologiya i iskusstvennyj intellekt: novyj e`tap starogo vzaimodej-stviya // Psixologiya cheloveka kak sub``ekta poznaniya, obshheniya i deyatel`nosti / pod red. V. V. Znakova, A. L. Zhuravleva. М. : Izdatel`stvo Instituta psixologii RAN, 2018. S. 2094–2101.
3. Emel`yanova T.P. Social`ny`e predstavleniya: Istoriya, teoriya i e`mpiricheskie issledova-niya. – М.: Izd-vo «Institut psixologii RAN», 2016. – 476 s.
4. Zhuravlev A. L., Nestik T. A. Social`no-psixologicheskie posledstviya vnedreniya novy`x texnologij: perspektivny`e napravleniya issledovanij // Psixologicheskij zhurnal. 2019. № 5 (40). S. 35–47. <https://doi.org/10.31857/S020595920006074-7>
5. Nestik T.A. Образ budushhego, social`ny`j optimizm i dolgosrochnaya orientaciya rossiyan: social`no-psixologicheskij analiz // Sociodigger, 2021, oktyabr`-noyabr`. Tom 2. Vy`pusk 9(14). – Str. 6-48.
6. Nestik T.A. Social`no-psixologicheskie prediktory` otnosheniya lichnosti k iskusstvennomu intellektu i robotizacii // Social`naya i e`konomicheskaya psixologiya. – 2018. – S. 406-428.
7. Nestik T.A. Social`no-psixologicheskie aspekty` otnosheniya cheloveka k novy`m texnologiyam // Aktual`ny`e problemy` psixologii truda, inzhenernoj psixologii i e`rgonomiki. – 2018. – S. 50-73.
8. Volkova A.D., Ivanova N.L., Patosha O.I. Social`naya identichnost` i predpochteniya potrebitelej: issledovanie vzaimosvyazi (na primere produktov pitaniya otechestvenny`x proiz-voditelej) // Nacional`ny`j psixologicheskij zhurnal. 2021. № 4(44). S. 39–52. doi: 10.11621/npj.2021.0404
9. Noakk N.V., Kostina T.A. E`mpir. issledovanie social`ny`x predstavlenij o smart-schetchikax // Cifrovaya e`konomika, 1(13), 2021. – str. 34-40.
10. Noakk N.V., Kostina T.A. Social`ny`e predstavleniya o II/ Cifrovaya e`konomika, №2 (14), 2021. – str. 24-30.
11. Yung K.-G. Analiticheskaya psixologiya/Perevod i redakciya V.V. Zelenskogo. Sankt-Peterburg, MCZnK i T «Kentavr». Institut Lichnosti, IChP «Palantir», 1994. – 136 str.

12. Yasin M. I. Predstavleniya molodezhi ob iskusstvennom intellekte i otnoshenie k nemu// Izvestiya Saratovskogo universiteta. Novaya seriya. Seriya: Filosofiya. Psixologiya. Pedago-gika. 2022. T. 22, vy`p. 2. S. 197–201. <https://doi.org/10.18500/18197671-2022-22-2-197-201>

*Ноакк Наталия Вадимовна – к.психол.н., ведущий научный сотрудник  
Федерального государственного бюджетного учреждения науки Центральный экономико-  
математический институт Российской академии наук г. Москва, Россия.*

ORCID 0000-0001-8696-5767

[n.noack@mail.ru](mailto:n.noack@mail.ru)

*Волкова Анастасия Дмитриевна – младший научный сотрудник  
Федерального государственного бюджетного учреждения науки  
Центральный экономико-математический институт*

*Российской академии наук г. Москва, Россия*

SPIN РИНЦ: 1470-2650

ORCID: 0000-0002-4216-9328

[volkova.nst@mail.ru](mailto:volkova.nst@mail.ru)

*Костина Татьяна Анатольевна*

*Федеральное государственное бюджетное учреждение науки*

*Центральный экономико-математический институт Российской академии наук г. Москва, Россия*

[kostina1@yandex.ru](mailto:kostina1@yandex.ru)

#### **Ключевые слова**

Социальные представления, простые свободные ассоциации, феномены будущего, архетипические мотивы, робототехника, амбивалентность, позитивное и негативное значение, ценности общества.

***Natalia Noack, Anastasia Volkova, Tatiana Kostina, Social perceptions of Russian society about artificial intelligence: a pilot study***

#### **Keywords**

Social representations, simple free associations, phenomena of the future, archetypal motives, robotics, ambivalence, positive and negative meanings, values of society.

DOI: 10.34706/DE-2023-04-03

JEL classification: C61 – методы оптимизации, модели программирования, динамический анализ

#### **Abstract.**

The description of a pilot study of the social representations of the Russian sample about artificial intelligence using the method of simple free associations and structural analysis according to P. Verges is given. The presence of ambivalence of social ideas about the object under study both at the level of large groups and at the level of an individual can lead to a sharp social polarization. The lack of conceptualization of artificial intelligence in the majority of respondents, understanding of its real threats, a clear idea of the boundaries of AI penetration into the practice of everyday experience, ethical and social aspects of its implementation suggests that these aspects are currently overlooked in public discourse. A negative emotional assessment of the consequences of the introduction of artificial intelligence is more characteristic of respondents who prefer the values of society. The presence of archetypal motives in the answers brings the problems of AI to the level of deep universal problems of human development and civilization as a whole.