

УДК: 303.2, 316.6

### 1.3. Социальные представления об Искусственном Интеллекте: методологические аспекты (Часть 2)

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

*Статья является продолжением Части 1, акцентируя проблему методологии исследования, как и центральности элементов структуры социальных представлений об ИИ. С этих позиций представлены различные методы исследования социальных представлений, с изложением полученных с их применением неоднозначных результатов. Показано, что выделенные с помощью метода 10%-ой согласованности ответов ядровые компоненты структуры представляют собой относительно согласованную и стабильную основу представлений респондентов-россиян, репрезентируют фундаментальные свойства объекта представления, а элементы периферии являются обобщением этих свойств и обозначением возможных стратегий поведения. Центральное ядро поощряет консенсус, психологическую общность группы, однако зависит от используемого метода диагностики.*

#### **Введение**

Предлагаемая вниманию читателя публикация продолжает описание результатов эмпирического исследования социальных представлений жителей России об искусственном интеллекте. В отличие от предыдущей статьи, [Ноак, 2021] авторы поднимают здесь вопросы методологического плана, поскольку они значительно влияют на характер получаемых результатов.

Исследования методологического характера сопровождают концепцию социальных представлений с начала выхода книги её автора [Moscovici, 1961]. Методы и инструменты, применяемые последователями Московичи для сбора и анализа эмпирического материала, многочисленны и разнообразны [Бовина, 2022; Markova, 1996; Moliner, 2002].

Используемый нами метод простых свободных ассоциаций с процедурной точки зрения состоит в том, что респондентов просят написать на слово-стимул (исследуемый объект репрезентации) слова, термины, выражения, которые спонтанно приходят им в голову. Грамматическая форма этих слов и выражений в нашем случае не имела значения.

Этот метод позволяет нам получить доступ к скрытым явлениям (измерениям), которые организуют семантическую «вселенную» изучаемого явления и/объекта. Метод простых свободных ассоциаций открывает путь к базовым единицам для описания социальных представлений. [Lahlou, 2011] Одно из преимуществ этого метода заключается также в том, что он позволяет быстро обрабатывать данные, получаемые в результате свободного самовыражения людей, эти данные легко понимать, использовать и анализировать [Abric, 2003].

Структурный подход, используемый нами, постулирует, что получаемые социальные представления представляют собой организованную, иерархическую систему, состоящую из двух интерактивных подсистем. Центральная система (Ядро) связана с коллективной памятью и нормами социальной общности, согласованна, относительно стабильна и не очень чувствительна к непосредственному контексту. Функция её – порождать значение представления и определять его организацию. Периферийная система состоит из элементов, конкретизирующих содержание Ядра, иллюстрирующих их. Ниже представлены результаты проведённого нами исследования с использованием структурного метода.

#### **Основная часть**

Всего в исследовании приняли участие 100 чел., средний возраст – 38 л. (70 % женщин, средний возраст – 37,2 года; 30% мужчин, средний возраст – 39,9 года). Преимущественно россияне (94%, 94 человека из 100), 47% респондентов имеют высшее образование, 15 % – незаконченное высшее, 23% и 12% - среднее специальное и среднее, соответственно). Социальный статус распределён следующим образом: самый большой процент выборки (43%) – это наёмные работники в организации, следом за этим идут работники бюджетной сферы и безработные – 13% и 12% соответственно; школьники и студенты – 6% и 5%, соответственно; предприниматели и пенсионеры – 6% и 5%, соответственно. На «Другое» приходится оставшиеся 10%.

#### **Прототипический анализ с использованием Категорий**

В Части 1 данной статьи был осуществлен анализ содержания социальных представлений (далее – СП) об искусственном интеллекте (далее – ИИ) с частичным применением метода контент-анализа и выделением обобщённых Категорий. Из всего полученного массива ассоциаций (354 ассоциации) путем объединения слов-синонимов и синонимических выражений, а также разных словоформ одного и того же слова были составлены словари понятий, которые затем были обработаны методом прототипического анализа по П. Вержесу.

Под единицей анализа (одной ассоциацией) в данном случае понималось и одно слово (например, польза), и в ряде случаев словосочетание и/или несколько словосочетаний: *повышенная волокита; тупая бюрократия, без возможности понять или разобратся* – то есть некоторая структурно-смысловая единица. В ходе контент-анализа и структурного анализа был выделен ряд тем и Категорий: Эти Катего-

рии близки к понятию «темата», введённому автором концепции [Емельянова, 2018]. Темата – это базовые, родовые понятия, которые коренятся в культуре, поддерживаются коллективной памятью и языком. К ним относятся, например: свой/чужой, естественное–природное/неестественное–искусственное, счастье–радость/несчастье–беда [Емельянова, 2018]. Несмотря на то, что понятие темат было введено в начале 1990-х годов, существует относительно мало исследований, анализирующих роль темат в порождении СП. Мы взяли на себя смелость проделать соответствующую работу, хотя она была сопряжена с определёнными трудностями методического характера. Во-первых, сама процедура выделения темат (мы назвали их Категориями, не будучи уверенными в правильности интерпретации нами термина Московичи и во избежание терминологической путаницы) и разнесения всего корпуса ассоциаций по ним предполагает в ряде случаев необходимость значительной интерпретации исследователем полученных данных, что может привести к содержательным ошибкам. Во-вторых, чем меньше количество выделяемых Категорий планируется исследователем, тем, соответственно, большее обобщение он вынужден осуществлять, жертвуя тем самым индивидуальными нюансами значений. С другой стороны, необходимость последующего структурного анализа всех полученных Категорий по методике Вержеса с выделением Ядра и Периферии, соответственно, требует достаточно небольшого числа элементов. В результате проведённого анализа были выделены следующие Категории.

Таблица 1. Категории, выделенные в ходе прототипического анализа

Название категории	Частота	Среднее значение
Антропоморфизм	68	2,7
Архетипы, тематы	23	3,3
Временные характеристики	17	3
Технологические характеристики	146	2,5
Страх, недоверие	23	2,6
Вера в хорошее	49	3,1
Применение	19	3,7
Сопротивление	8	3,25
<i>медиана</i>	23	3,1

Слева – названия категорий, справа – количество ассоциаций, которые, по мнению авторов, подпадают сюда по значению, а также обозначение среднего ранга всех относящихся к Категории ассоциаций.

К Категории Антропоморфизм были отнесены, среди прочих, следующие слова/словосочетания/выражения: *безмозгий, без чувств, всезнающий, глупость, дружелюбный, женщина, мозг, книгу читает, полное отсутствие человечности, ненормальный...* Во всех ассоциациях этой Категории присутствует наделение ИИ некоторыми человеческими характеристиками – душой, чувствами, качествами личности.

К Категории Временные характеристики были отнесены, среди прочих, следующие слова/словосочетания/выражения: *будущее, современность, современный мир*. Название Категории говорит само за себя, причем ассоциаций ИИ с будущим гораздо больше, чем с настоящим.

К Категории Архетипы были отнесены, среди прочих, следующие слова/словосочетания/выражения: *война, новое опасное увлечение человечества, опасность/опасный, искусственность, смерть, неестественно, ненастоящий предмет, неживой, самодельный предмет, сделанный предмет*. Обратим внимание, что, наряду с характерными для архетипических мотивов тем опасности, смерти, войны, значительное место в этой Категории занимает тематика естественности, живого/искусственности, неживого, ставящая глобальные вопросы бытия, в частности, созданного природой или человеком.

К Категории Технологические характеристики были отнесены, среди прочих, слова/словосочетания/выражения, с одной стороны, указывающие на технические характеристики (*автомат, андроид, квантовый компьютер, провода*), с другой – на технологические (*алгоритм, нейросети, программа, обучение, программирование*)

К Категории Страх, недоверие были отнесены, среди прочих, следующие слова/словосочетания/выражения: *безработица, заговор, захват, обман, плагиат, новое изобретение для замены работы мозга человека, репрессии, тревога, угроза, слежка...* Здесь присутствуют ассоциации, отражающие чувства и эмоции по поводу ИИ, преимущественно негативные, как, например, страх потерять работу или быть обманутым; кроме того, ассоциации этой Категории так или иначе связаны с конспирологическими теориями, разделяемыми некоторыми респондентами.

К Категории Вера в хорошее были отнесены, среди прочих, следующие слова/словосочетания/выражения: *возможности, инновация, лёгкость, научный прорыв, оптимизация, польза, помощник, прогресс, развитие, разносторонний, разнообразно...* Критерием отнесения ассоциаций к этой Категории явилось выражение доверия, позитивного отношения респондентов к феномену ИИ.

К Категории Применение были отнесены, среди прочих, следующие слова/словосочетания/выражения: *алиса, медицина, наука, искусство, красивый дом, творчество, учёба, сервис, сон...* Как мы

видим, сюда входят ассоциации, указывающие на примеры практического использования респондентами ИИ в повседневной жизни.

К Категории Соппротивление были отнесены, среди прочих, следующие слова/словосочетания/выражения: *нет, против, колбаса...* Критерием послужило выражение в них, скорее всего, нежелания респондента идти на контакт в экспериментальной ситуации.

Как мы видим по приведённым примерам, ширина и характер разброса ассоциаций по разным Категориям различна, значительно варьирует и по количеству (от 8 в категории Соппротивление до 146 в Категории Технологические характеристики), и по содержанию (от достаточно единообразной Категории Временные характеристики, до на первый взгляд «сборной солянки» в Категории Соппротивление).

Каждому слову/высказыванию, отнесённому в ту или иную Категорию, соответствует определённый ранг, то есть число порядка в ряду ассоциаций. В процедуре обработки данных подсчитывается общее количество ассоциаций во всех Категориях и вычисляется среднее значение рангов. На следующем шаге обработки определяется медиана всех частот и медиана всех рангов. Наконец, в соответствии с определёнными требованиями, прописанными в исследовательской литературе, Категории разносятся по четырём квадратам (см. рис. 1), образуя структуру СП, состоящую из Ядра, ближней Периферии и дальней Периферии. Ниже графически представлена структура социальных представлений респондентов выборки об ИИ в соответствии с прототипическим анализом П. Вержеса.

<p><b>Квадрат 1 (частота <math>\geq 23</math>; ранг <math>&lt; 3,1</math>)</b> <b>ЯДРО</b></p> <p>технологические характеристики антропоморфизм вера в хорошее страх, недоверие</p>	<p><b>Квадрат 3 (частота <math>\geq 23</math>; ранг <math>\geq 3,1</math>)</b> <b>ПЕРИФЕРИЯ 1 Внешнее влияние</b></p> <p>архетипы, <u>тематы</u></p>
<p><b>Квадрат 2 (частота <math>&lt; 23</math>; ранг <math>\leq 3,1</math>)</b> <b>ПЕРИФЕРИЯ 1 Зона потенциальных изменений</b></p> <p>временные характеристики</p>	<p><b>Квадрат 4 (частота <math>\leq 23</math>; ; ранг <math>\geq 3,1</math>)</b> <b>ПЕРИФЕРИЯ 2</b></p> <p>применение сопротивление</p>

**Рис. 1. Структура СП с использованием Категорий анализа**

Как мы видим, центральная зона представлена элементами как технического/технологического плана, так и характеристиками, присущими только человеку. Уже здесь закладывается та амбивалентность, которая будет присуща так или иначе всей структуре корпуса ассоциаций. Так, в зону ближней Периферии структуры социальных представлений, наряду с категориями, связанными с выражением общего недоверия к ИИ, страхами и сомнениями, вошла тематика вера в возможности ИИ и прогресс, с ним связанный. Дальняя Периферия также была представлена категориями с амбивалентным значением. В результате нами была высказана гипотеза, что амбивалентность социальных представлений об ИИ может быть знаком сосуществования в российской общественности полярных социальных групп в отношении к ИИ.

Неоднозначность и противоречивость полученных данных мотивировала нас на поиски методов, позволяющих ответить на вопросы, касающиеся определяющих факторов такой неоднородности. Мы обратились к комплексу так называемых лексико-метрических методов. Их возможности позволили бы нам семантически подтвердить/опровергнуть выдвинутые гипотезы. Кроме того, эта группа методов, по мнению исследователей [Бовина, 2022; Гуриева, 2021; Стернин, 2017] имеет прямое отношение к выявлению так называемой когнитивной устойчивости представления, что, в свою очередь, дало бы нам знак о правомерности (адекватности) проделанной процедуры построения структуры и выделения Ядра СП.

**К лексико-метрическим методам**, используемым в анализе социальных представлений, относятся методы определения количественных свойств полей представления: 1) ширины поля представления, 2) показатель разнообразия ответов; 3) показатель редкости ответов; 4) коэффициент когнитивной устойчивости представления.

Ширина поля – это количество ассоциаций, употребленных хотя бы один раз: в нашем случае это число равно 183 ассоциациям. Показатель говорит, фактически, о разнообразии представлений. В свою очередь, показатель разнообразия ответов варьирует от 0 до 1 и может свидетельствовать как о разно-

образии лексического запаса изучаемой группы, так и о гетерогенности самой группы. Он представляет собой соотношение ширины поля к общему количеству ассоциаций и в нашем случае составляет 0,52. Таким образом, лексический запас изучаемой группы достаточно разнообразен, и вследствие этого показатель гетерогенности (неоднородности) группы также достаточно высок.

Показатель редкости ответов (Показатели Хапакского индекса (Нарахс) указывает на количество слов, которые упоминаются только одним респондентом: в нашей выборке он составляет 111 (31% от всей выборки). Чем больше единичных ответов в анализируемой базе данных, тем выше варьирование между индивидами (Flament, 2003) Как мы видим, и это показатель достаточно высок (после треть от выборки), таким образом, можно говорить о большом варьировании представлений среди респондентов. С помощью данного показателя можно также вычислить и следующий - коэффициент когнитивной устойчивости представления: это соотношение индивидуальных и общих ответов в ответах исследуемых групп: в нашей выборке индивидуальных ответов 111, а 243 ассоциации встречаются 2 и более раз. Соответственно, коэффициент когнитивной устойчивости составляет 0,45. По шкале оценки предполагается, что, если значения лежат в диапазоне до 0,50 – это означает, что поле представления коллективное и устойчивое; после 0,50 – неустойчивое индивидуальное. Наши значения лежат в диапазоне до 0,50, но фактически почти на границе. Это подтверждает наше предположение о наличии неустойчивости в структуре социальных представлений нашей выборки.

#### **Прототипический анализ с использованием метода 10%-ого порога значений**

В ходе анализа возникло намерение проверить полученную нами структуру СП, используя другие методы анализа, с помощью которых можно было бы уменьшить степень когнитивной неустойчивости представлений [Rateau, 2011; Dany, 2015; Бовина, 2022] Мы поставили следующие задачи: проанализировать возможность разных методов в построении структуры социальных представлений россиян об ИИ; проверить гипотезу о центральности выделенных нами элементов ядра. Представленную выше структуру СП можно охарактеризовать как гипотетическую, что вполне согласуется с мнениями последователей концепции [Бовина, 2022; Dany, 2015; Galli, 2021] Эта структура должна получить свою проверку на последующих этапах исследования. Таким образом, одной из наших задач является рассмотрение т.н. проблемы центральности, или проблемы диагностики идентификации ядра, в практическом ключе, применительно к полученным нами результатам. Ниже представлены результаты этого анализа.

<p><b>(14; 2,8)</b>  <b>Квадрат 1 (частота <math>\geq 14</math>; ранг <math>&lt; 2,8</math>)</b>  <b>ЯДРО</b></p> <p>робот  компьютер</p>	<p><b>Квадрат 3 (частота <math>\geq 14</math>; ранг <math>\geq 2,8</math>)</b>  <b>ПЕРИФЕРИЯ 1 Внешнее влияние</b></p> <p>ум/умный  будущее</p>
<p><b>Квадрат 2 (частота <math>\leq 14</math>; ранг <math>&lt; 2,98</math>)</b>  <b>ПЕРИФЕРИЯ 1 Зона потенциальных изменений</b></p> <p>машина</p>	<p><b>Квадрат 4 (частота <math>&lt; 14</math>; ; ранг <math>\geq 2,8</math>)</b>  <b>ПЕРИФЕРИЯ 2</b></p> <p>интернет  технологии</p>

**Рис. 2. Структура СП с использованием метода 10%-ого порога значений.**

Классический метод обработки корпуса ассоциаций и определения структуры социального представления заключается в учёте двух показателей – частоты появления той или иной ассоциации в группе и её ранга (места в ряду). С показателем частоты появляется доступ к коллективному измерению – сколько раз слово употребляется (цитируется) респондентами выборки. Это индикатор количественной центральности [Moline, 1994].

В ранжировании имеет место индивидуальная операция – как предполагают исследователи, респонденты пишут (или произносят) ассоциации в соответствии с их когнитивной доступностью. Предполагается существование связи между частотой ассоциативного ответа и скоростью его появления. Совпадение этих двух критериев (частоты и ранга) может служить индикатором центральности (ядровости) элемента. Соответственно, различные комбинации их могут определить и другие зоны представления, образующие Периферию. Таким образом, в Квадрате 1 находятся элементы, часто цитируемые и упоминаемые в первых рядах, это Ядро социального представления. Напротив, в Квадрате 4 – элементы, цитируемые реже всего и последними. Это так называемая дальняя Периферия. В Квадратах 2 и 3 – высокая частота ассоциаций и низкий ранг, низкая частота и высокий ранг, соответственно. Это – области ближней Периферии, зона нестабильности, источник трансформаций и перемен.

Изложенные выше критерии организации структуры социальных представлений позволяют, по мнению исследователей, идентифицировать центральные и периферийные элементы, но не позволяют сформулировать *диагноз центральности* [Danu, 2015]. Чтобы приблизиться к его формулировке, мы предприняли попытку использовать несколько методов в выявлении структуры СП об ИИ.

Речь идёт, прежде всего, о вопросе процента согласованности ответов, который стоит учитывать при построении структуры. В литературе этот параметр варьируется от 3 до 10% и более (до 25%).

Ассоциации, преодолевшие 10-процентный порог (иначе говоря, которые назвали не менее 10% респондентов выборки) охватывают в нашей выборке 35% от всех названных респондентами ассоциаций, что существенно превышает случайность. Всего в Ядро попало 64 ассоциации из 354, они представлены 2-мя элементами (рис. 2). Элементы Ядра несут основную смысловую нагрузку об объекте представления: ИИ для респондентов – *это робот и компьютер*. Хотя в количественном отношении (представленность по выборке) эти 2 элемента различаются между собой, тем не менее оба они попадают в Ядро, и это не может быть случайностью, вернее, это может быть НЕ-случайностью. Попробуем проанализировать, что может объединять эти 2 элемента. Общее в роботе и компьютере – использование программного обеспечения при выполнении разнообразных задач. Кроме того, и роботы, и компьютеры имеют датчики для реакции на внешнюю и среду и для выполнения задачи. Далее – и роботы, и компьютеры могут быть подключены к сети, использовать интернет для обмена данными и коммуникации с другими устройствами. И, наконец, при этом они остаются машинами, то есть техническими устройствами, зависящими от человека, подчиняющимися его командам. Таким образом, рядоположение, совместность этих 2-х элементов носит, по-видимому, неслучайный характер.

С другой стороны, одно не есть повторение другого – робот – это не компьютер, и наоборот. Роботы и компьютеры оба являются техническими устройствами, но у них есть ряд существенных различий:

Физическая оболочка: роботы обычно имеют физическое тело, которое может перемещаться и взаимодействовать с окружающей средой. Это может быть механическое тело, оборудованное сенсорами, моторами и другими устройствами для выполнения различных задач. Компьютер, с другой стороны, обычно представляет собой электронное устройство без физической оболочки.

Движение и действия: роботы способны выполнять различные физические действия, такие как передвижение, поднятие и перемещение объектов, взаимодействие с окружающей средой. Компьютеры ограничены в основном обработкой информации и вычислениями.

Автономность: некоторые роботы могут быть способными к автономной работе, что означает, что они могут самостоятельно принимать решения и выполнять задачи без постоянного участия человека. Компьютеры обычно работают на основе программ, которые выполняются в соответствии с инструкциями, заданными человеком, и не обладают автономностью.

Цель использования: роботы часто применяются в физических задачах, таких как промышленное производство, медицинская хирургия, исследования в пространстве и многие другие. Компьютеры, в основном, используются для обработки данных, управления информацией, взаимодействия с пользователем и решения различных задач, связанных с информацией и вычислениями.

В целом, роботы и компьютеры могут совмещать свои функции, и существуют также роботы, которые включают в себя компьютеры для управления и обработки информации.

Таким образом, элементы, вошедшие в Ядро социального представления об искусственном интеллекте, демонстрируют общность значений и в то же время существенные семантические различия, особую семантическую связь с объектом представления (ИИ); они также подчёркивают принцип комплементарности значений. [Vergès, 1994].

Проанализируем элементы, вошедшие в ближнюю и дальнюю Периферию

Первая периферическая зона представлена элементами «*машина*», *ум/умный*, *будущее*. Она является своеобразным обобщением для элементов ядра, соединяя оба элемента (*робота и компьютер*) в одно целое, подчеркивая то общее, что видится в них; позволяет соотнести их с другими подобными явлениями (например, машинами вообще, автоматизированными техническими устройствами, выполняющими интеллектуальные задачи), передать суть социального представления в коммуникации, (отношение человека именно к машинам, а не к другому человеку: первое, к примеру, предполагает отсутствие сложных чувств), а также конкретизировать выбор поведенческих стратегий относительно выбранного объекта (имеется в виду действия и/или поступки человека по отношению к машинам – например, опора на их многофункциональность, мощностные, контролируемые их действия со стороны человека). По мнению исследователей, [Wagner, 1993], одно из перспективных для изучения свойств социальных пред-

ставлений как раз и заключается в том, что они дают знание о возможных стратегиях поведения респондентов: суть СП – не в том, как они отражают реальность, а в том, какие поступки они могут вызвать. Это не объяснительные теории в строгом смысле слова. По словам исследователя, (Вагнер) автор концепции Московичи также считал, что СП не опосредуют стимулы, но сами являются стимулами, и, следовательно, независимыми переменными в эмпирических исследованиях.

С другой стороны, это не просто машины, но *умные* машины, то есть наделённые одной из важнейших характеристик человека – умом, интеллектом. Это потребует от человека определённого поведения, детали которого пока не ясны. С этим связано, видимо, появление элемента *будущее* в ближней периферии: то, что неясно в настоящем, будет определено в будущем, для этого появятся необходимые инструменты.

Элементы, включенные во 2-ую периферическую зону, связывают элементы ядра с выбором поведенческой стратегии *в повседневных ситуациях*. Стратегии поведения представлены в элементах 2-ой периферии (*интернет, технологии*), Это, во-первых, указание на то, что ИИ позволяет, как и *интернету*, осуществлять связь всех со всеми, хранить данные, передавать сообщения и многое другое, во-вторых, подчёркивание важности *технологий* как способов контролируемого достижений намеченной цели.

Проведённый выше анализ структуры социальных представлений об ИИ снова обращает нас к проблеме, которую в литературе получила название *проблема центральности* (Бовина, 2022).

#### **ИИ и проблема центральности.**

П. Вержес, предложивший прототипический анализ структуры социальных представлений, использовал, как указывалось выше, количественный (частоту встречаемости элемента) и качественный (ранг появления элемента) показатели. Для другого последователя концепции – Ж.-К. Абрика – идея П. Вержеса некорректна: он считает, что важные, значимые для респондента мысли и суждения появляются в ассоциативном ряду не в начале, а позже, после определенного временного отрезка, после своего рода преодоления защитных механизмов. Выявлять эти элементы он предлагал другими методами. Как утверждает [Бовина, 2022], на настоящий момент развития концепции различить ранг появления ассоциации и её важность не представляется возможным. Но возможно попробовать использовать разные методы для проверки гипотетического Ядра. Полученные 2 элемента Ядра – это по-прежнему лишь гипотеза. Ее необходимо проверить другими исследованиями и другими методами, некоторые из которых вызывают сложности в применении и пока не представляются возможными. Ограничимся анализом с применением доступных нам методов решения этой проблемы.

Обозначим главные свойства ядерных терминов, выделенные в литературе, и проанализируем с этих позиций характеристики тех элементов, которые вошли у нас в зону Ядра.

1. Элементы Ядра должны быть стабильны [Danu, 2015] и менее склонны к изменениям [Бовина, 2022; Moscovici, 1961] Поэтому одним из методов для возможной проверки является анализ динамики результатов, например, сопоставление с результатами исследований прошлых лет. Мы проводили исследование социальных представлений об ИИ в 2020 году [Ноакк, 2021].

<p><b>Квадрат 1 (частота <math>\geq 4</math>; ранг <math>&lt; 2,8</math>)</b> <b>ЯДРО</b></p> <p>будущее робот</p>	<p><b>Квадрат 3 (частота <math>\geq 4</math>; ранг <math>\geq 2,8</math>)</b> <b>ПЕРИФЕРИЯ 1 Внешнее влияние</b></p> <p>компьютер опасность</p>
<p><b>Квадрат 2 (частота <math>&lt; 4</math>; ранг <math>\leq 2,8</math>)</b> <b>ПЕРИФЕРИЯ 1 Зона потенциальных изменений</b></p> <p>машина технологии умный</p>	<p><b>Квадрат 4 (частота <math>&lt; 4</math>; ранг <math>\geq 2,8</math>)</b> <b>ПЕРИФЕРИЯ 2</b></p> <p>развитие прогресс</p>

**Рис. 3. Результаты исследования 2020г.**

Как мы видим по рис. 3, элемент *робот* входит в ядро структуры СП и в 2020 году. По данным исследования, его также выбрали 50% респондентов. Наряду с *роботом*, ядро СП представляет также элемент *будущее*, (со значительным отставанием), который в исследовании 2023 года входит в зону

ближней Периферии. Элементы *машина*, *технологии*, *умный* в исследовании 2020 года входят в ближнюю Периферию, а к 2023 году перемещаются из зоны 2 в зону 3, оставаясь, тем не менее в так называемой ближней Периферии (кроме *технологии*, которые перемещаются в зону 4 – Область стратегий поведения). Интересно, что элемент *опасность* в исследовании 2020 года также находится в ближней Периферии, в квадрате 3, а к 2023 он вообще исчезает из структуры с порогом значений в 10%, оставаясь лишь в индивидуальных представлениях на уровне структуры общей выборки россиян.

Элемент *компьютер*, который в исследовании 2020 входит в ближнюю Периферию и охватывает примерно 16% респондентов, к 2023 году перемещается в Ядро структуры и представлен 22% выборки, что вполне согласуется с описанием основателями концепции процессов трансформации в рамках ближней Периферии.

Резюмируя изложенное, можно заключить: большая часть элементов исследования 2020 года остаётся и в структуре 2023 года, что подтверждает их центральность; в Ядре по-прежнему элемент *робот*, что указывает на его системоорганизующую роль в структуре СП; перемещение элемента *компьютер* из ближней Периферии в Ядро (в 2023 году) может указывать на увеличение его роли в СП. Элемент *технологии* в исследовании 2023 года обнаруживает себя в квадрате 4 - эта Область, как мы указывали выше, отражает стратегии поведения респондентов по отношению к объекту репрезентации. Таким образом, как элементы Ядра (особенно элемент *робот*), так и элементы Периферии обнаруживают относительную стабильность. Произошедшие трансформации – предмет отдельного описания.

3. Центральность элементов ядра можно проверить с использованием лексико-метрических методов, в частности, [Бовина, 2022; Виноградова, 2016] индекса яркости элементов. Он определяется как отношение количества испытуемых, актуализировавших в экспериментах определенное значение объекта, к общему числу испытуемых, принявших участие в эксперименте.

Индексы яркости приводятся в относительных цифрах, в виде десятичной дроби. Чем ближе десятичный индекс яркости к единице, тем ярче соответствующий признак в языковом сознании носителей языка (Виноградова, 2016; Стернин, 2017). В нашем случае индексы яркости элементов Ядра *робот* и *компьютер* составляют 0,49 и 0,22, соответственно; элементов Периферии – 0,1 (*машина*), 0,18 (*ум/умный*), 0,15 (*будущее*), 0,1 (*интернет*), 0,16 (*технологии*).

Таким образом, по показателю индекса яркости самые высокие значения имеют элементы *робот* и *компьютер*, совершенно обоснованно входя в Ядро структуры и подтверждая гипотезу о центральности этих элементов.

Завершая раздел о методах проверки центральности элементов структуры СП, отметим, что используемые нами методы подтверждают гипотезу центральности лишь отчасти, тем не менее продвигая нас в решении задачи.

#### Гендерные особенности структуры СП

Выше мы писали о том, что неоднозначность и противоречивость полученных на первом этапе исследования данных мотивировали нас на поиски методов, позволяющих ответить на вопросы, касающиеся определяющих факторов такой неоднозначности. Как известно, одним из базовых определяющих факторов различий очень часто выступают гендерные особенности. Мы решили проверить влияние половых различий на изменение структуры СП.

<p><b>Квадрат 1 (частота <math>\geq 11</math>; ранг <math>&lt; 2,5</math>)</b> <b>ЯДРО</b></p> <p>робот компьютер</p>	<p><b>Квадрат 3 (частота <math>\geq 11</math>; ранг <math>\geq 2,5</math>)</b> <b>ПЕРИФЕРИЯ 1 Внешнее влияние</b></p> <p>умный будущее</p>
<p><b>Квадрат 2 (частота <math>\leq 11</math>; ранг <math>&lt; 2,5</math>)</b> <b>ПЕРИФЕРИЯ 1 Зона потенциальных изменений</b></p> <p>машина</p>	<p><b>Квадрат 4 (частота <math>&lt; 11</math>; ранг <math>\geq 2,5</math>)</b> <b>ПЕРИФЕРИЯ 2</b></p> <p>интернет нейросеть</p>

Рис. 4. Женщины. Социальные представления

<p><b>Квадрат 1 (частота <math>\geq 4</math>; ранг <math>&lt; 3</math>)</b> <b>ЯДРО</b></p> <p>робот</p>	<p><b>Квадрат 3 (частота <math>\geq 4</math>; ранг <math>\geq 3</math>)</b> <b>ПЕРИФЕРИЯ 1 Внешнее влияние</b></p> <p>компьютер технологии</p>
<p><b>Квадрат 2 (частота <math>\leq 4</math>; ранг <math>&lt; 3</math>)</b> <b>ПЕРИФЕРИЯ 1 Зона потенциальных изменений</b></p> <p>разум</p>	<p><b>Квадрат 4 (частота <math>&lt; 4</math>; ; ранг <math>\geq 3</math>)</b> <b>ПЕРИФЕРИЯ 2</b></p> <p>будущее</p>

**Рис. 5. Мужчины. Социальные представления**

У женщин структура в целом полнее, там больше элементов, чем у мужчин – (7 элементов), причём 35% (91 из 260) ассоциаций входят в структуру ассоциаций. Сама структура более разветвленная, характеризующаяся бОльшей общностью (рис. 4).

У мужчин всего 5 элементов в структуре, что охватывает 28 ассоциаций в структуре - примерно 28% от всех (из 98) ассоциаций, названных респондентами-мужчинами (рис. 5). Мы видим, что у обеих выборок – и женщин, и мужчин - элемент *робот* входит в ядро, причём, со значительным превосходством над остальными элементами: его выбирают 43% (29 чел.) женщин и 50% (14 чел.) мужчин, соответственно). Далее начинаются различия. Элемент *компьютер*, который в общей выборке также входит в Ядро, испытывает значимое влияние женской части выборки: 24% женщин помещают его в Ядро представления. Мужчины относят его в ближнюю Периферию, наряду с *технологиями* – эти элементы набирают по 15% от всей выборки респондентов-мужчин. У женщин в ближней к Ядру периферии стоит *машина*, словно подчёркивая главное отличие ИИ от человека – его искусственное происхождение и зависимость от его создателя. У мужчин на это месте находится *разум*, как бы, напротив, объединяя и человека, и ИИ. Из общих элементов стоит отметить «будущее» - у женщин этот элемент находится в ближней Периферии, у мужчин – в дальней, связанной со стратегиями поведения. Таким образом, респонденты-мужчины более, чем женщины, склонны отодвигать проблемы реального взаимодействия с ИИ и поведения по отношению к нему, как и принятия глобальных решений, связанных с выбором, выносить за рамки настоящего, обнаруживая своеобразную неготовность решать актуальные вопросы рисков и угроз, а, возможно, и преимуществ тоже. В то же время, они больше, чем респонденты-женщины, озабочены вопросами, связанными с технологиями обращения с ИИ.

Таким образом, структура СП в выборках мужчин и женщин несколько отличается, но больших выводов пока делать рано.

#### **Заключение.**

Определяя ИИ как средство небывалой мощности, мы постулируем необходимость адекватного контроля за процессом его рукотворного развития. Одной из контролирующих и корректирующих сил в данной. как и в других сферах, является массовое сознание, включающее в себя в той или иной степени структурированную совокупность представлений о явлениях мира. Мы предположили, что применение для решения данной задачи такого нечасто используемого диагностического инструмента, как методика простых свободных ассоциаций, позволит несколько продвинуться за границы известного и очевидного.

Что здесь может нас интересовать? Вероятно, степень действительной озабоченности соотечественников проблемами развития ИИ в стране и мире, что способно «просигнализировать» через ассоциации не только на сознательном, но и на бессознательном уровне. В некотором смысле это апелляция к родовой, общечеловеческой ответственности на уровне коллективного бессознательного. Это та самая социальная ответственность, имеющая корнем потребность естественного и духовного выживания, которая побуждает людей к стихийному или упорядоченному переустройству условий своего существования.

Применение для анализа этого явления концепции социальных представлений продиктовано средоточием в нём нескольких характеристик, о которых писал автор концепции: ИИ являются «выпуклым» социальным объектом и характеризуется новизной, этической и эмоциональной неоднозначностью, что стимулирует возникновение в обществе обширных дискуссий, обсуждений, а затем и появление социальных представлений в общественном сознании.

В Части 1 данной статьи было выявлено наличие в выборке респондентов противоположных по своей направленности групп россиян с характерными для них СП об искусственном интеллекте, включающими комплекс мнений и суждений об ИИ, различных установок относительно объекта и, соответственно, неоднозначных стратегий поведения. В результате нами была высказана гипотеза, что амбивалентность социальных представлений об ИИ может быть знаком сосуществования в российском обществе полярных социальных групп в отношении к ИИ.

Применение к полученному корпусу ассоциаций лексико-метрических методов наглядно показало наличие достаточно высокого варьирования социальных представлений между респондентами, гетерогенности нашей выборки, что и привело к когнитивной неустойчивости поля представления. Неоднозначность полученных данных послужило основанием для поиска дальнейших возможностей и методов анализа.

Использование метода прототипического анализа по П. Вержесу с учётом порога в 10% выборки позволило выдвинуть гипотезу, что социальные представления респондентов, вошедших в эту группу, в отличие от выборки в целом, представляют собой мнения, суждения, эмоции и практики, отличающиеся единообразием и внутренней непротиворечивостью. Искусственный интеллект респонденты данной выборки представляют в качестве образа робота и/или компьютера, с акцентом на их технических аспектах и больших возможностях, особенно в будущем развитии, подчеркиванием их интеллектуальной составляющей как основополагающей доминанты прогресса.

Проверка полученной гипотетической структуры социальных представлений (гипотезы центрально-сти) была осуществлена с применением ряда методов: метода анализа сравнительной динамики структуры СП; метода вычисления яркости элементов, метода сравнения структур СП по гендерному признаку. Выявлено, что наибольшие шансы на присутствие в Ядре социального представления об искусственном интеллекте имеет элемент *робот*, на втором месте с большой вероятностью – *компьютер*. Остальные элементы, вошедшие в ближнюю к Ядру Периферию, подчёркивают техническую природу объекта представления, связанную также с его многофункциональностью, возможностью решать сложные интеллектуальные задачи. Элементы, связанные с тревогой и опасностью, присутствующие в исследовании 2020 года, как и в 2023 году, но на уровне всей выборки, уходят из структуры с 10%-порогом общности, оставаясь на уровне индивидуальных интерпретаций. Подчеркнута особая роль элемента *будущее* - и у женщин, и у мужчин этот элемент стоит в Периферии, причём, у мужчин – в дальней Периферии, с которой исследователи связывают обозначение стратегий поведения. Таким образом респонденты-мужчины склонны отодвигать проблемы реального взаимодействия с ИИ, выносить за рамки настоящего, обнаруживая своеобразную неготовность решать актуальные вопросы рисков и угроз, а, возможно, и преимуществ тоже.

Резюмируя изложенное выше, нарисует «портрет» социального представления об ИИ, которое продемонстрировала наиболее (10% порог значений) репрезентативная часть нашей выборки.

Искусственный интеллект представляет собой робота и/или компьютер, предназначенные для выполнения самых многообразных задач разных уровней сложности, как физических, так и высокоинтеллектуальных и подчиняющиеся командам со стороны человека или особым написанным инструкциям, заданным человеком. ИИ может иметь физическое тело, а может и быть встроенным в какое-либо устройство. Это умные (разумные) машины, то есть технические устройства, способные работать на уровне человеческого интеллекта, тем не менее оставаясь под его контролем. Будущее развитие этого явления сулит большие возможности пока неясного характера.

Подчеркнем еще раз, что автор концепции видел главную функцию СП в трансформации чего-то нового, неизвестного, тревожного, пугающего, представляющегося опасным и разрушающим – во что-то знакомое, известное, обычное, позволяющее человеку и группе сохранить картину мира непротиворечивой. Образы робота и компьютера, присутствующие в полученной в нашем исследовании структуре социальных представлений об искусственном интеллекте, позволяют сделать ИИ как явление новое и во многом непонятное и непредсказуемое – понятным, знакомым, существующим в повседневных ситуациях жизни и работы, требующим определённых, опять же знакомых, действий и поступков. В этом смысле социальные представления об ИИ в нашей выборке совершенно точно выполняют свою защитную функцию, оберегая психику от угрозы, с которой она может не справиться, ведь и робот (через литературный и кинодискурс), и компьютер уже давно вошли в повседневную жизнь россиян.

#### Литература

1. Бовина И.Б., Дворянчиков Н.В., Мельникова Д.В., Лаврешкин Н.В. К вопросу об исследовании социальных представлений: взгляд со стороны // Социальная психология и общество. 2022. Том 13. № 3. С. 8—25. DOI: <https://doi.org/10.17759/sps.2022130302>
2. Виноградова О. Е., Стернин И.А. Психолингвистические методы в описании семантики слова. Монография. – Воронеж: издательство «Истоки», 2016. – 160 с. – 200 стр.
3. Гуриева С. Д., Я. Е. Виноградова Социальные представления о феномене предательства: анализ структуры и содержания в разных возрастных группах Вестник Вятского государственного университета, 2021, № 1 (139), стр. 97-114. DOI: 10.25730/VSU.7606.21.010
4. Емельянова Т.П., Вопилова И.Е. Социально-психологический механизм ментального освоения новых технологий //из кн Психология человека как субъекта познания, общения и дея-

- тельности/ Отв. Ред. В.В. Знаков, А.Л.Журавлёв. - М.: Изд-во «Институт психологии РАН», 2018. - стр.687-696.
5. Ноак Н.В., Костина Т.А. Социальные представления об ИИ // Цифровая экономика. №2 (14), 2021, стр. 24-30.
  6. Ноак Волкова Костина Социальные представления российского общества об искусственном интеллекте: пилотное исследование (часть 1) // Цифровая экономика, 2023, №4, стр.23-32
  7. Стернин И.Я. Лесикографическое представление психологического значения слова /Вестник РУДН. 2017, т.15, №2. стр. 204-214. DOI 10.22363/2313-2264-2017-15-2-204-214
  8. Abric, J.-C.: L'analyse structurale des représentations sociales. In: Moscovici, S., Buschini, F. (eds.) Les méthodes des sciences humaines, pp. 375–392. Presses Universitaires de France, Paris (2003b)
  9. Dany L., Urdapilleta I., Lo Monaco G. Free associations and social representations: some reflections on rank-frequency and importance-frequency methods // Quality and Quantity. 2015. Vol. 49. P. 489—507.
  10. Flament C., Rouquette M.-L. Anatomie des idées ordinaires. Paris: Armand Colin, 2003. 256 p.
  11. Galli I., Fasanelli R. From Prototypical Stimuli to Iconographic Stimuli: The Power of Images in the Study of Social Representations // RUDN Journal of Psychology and Pedagogics. 2021.Vol. 18. P. 391—401. DOI:10.22363/2313-1683-2021-18-2-391-401
  12. Jodelet D. The Notion of Common and Social Representations // RUDN Journal of Psychology and Pedagogics. 2021. Vol. 18. P. 299—314. DOI:10.22363/2313-1683-2021-18-2-299-314
  13. Lahlou, S., Abric, J.-C.: What are the "elements" of a representation? Pap. Soc. Represent. 20, 20.1–20.10 (2011)
  14. Lahlou S. Social Representations and Individual Representations: What is the Difference? And Why are Individual Representations Similar? // RUDN Journal of Psychology and Pedagogics. 2021. Vol. 18. P. 315—331. DOI:10.22363/2313-1683-2021-18-2-315-331
  15. Markova I. Towards an epistemology of social representations // Journal for the Theory of Social Behavior. 1996. V. 26. № 2. P. 177–196.
  16. Moliner, P.: Les méthodes de repérage et d'identification du noyau des représentations. In: Guimelli, C. (ed.) Structures et transformations des représentations sociales, pp. 199–232. Delachaux et Niestlé, Neuchâtel (1994)
  17. Moliner, P., Rateau, P., Cohen-Scali, V.: Les représentations sociales : pratique des études de terrain. Presses Universitaires de Rennes, Rennes (2002)
  18. 36. Moliner P., Lo Monaco G. Méthodes d'association verbale pour les sciences humaines et sociales.Grenoble: Presses Universitaires de Grenoble, 2017. 190 p.
  19. Moscovici, S.: La psychanalyse, son image, son public. Presses Universitaires de France, Paris (1961)
  20. Rateau, P., et al.: Social representation theory. In: P Van Lange, P., Kruglanski, A., Higgins, T. (eds.) Handbook of Theories of Social Psychology, vol. 2, pp. 478–498. Sage, Thousand Oaks (2011)
  21. Vergès, P.: Approche du noyau central: propriétés quantitatives et structurales. In: Guimelli, C. (ed.) Structures et transformations des représentations sociales, pp. 233–253. Delachaux et Niestlé, Neuchâtel (1994)
  22. Wagner W. Can Representations Explain Social Behaviour? A Discussion of Social Representations as Rational Systems //Papers on Social Representations - Textes sur les Représentations Sociales (1021-5573) Vol. 2 (3), 236-249 (1993).

#### References in Cyrillics

1. Vojskunjij A.E. Psixologiya i internet. – М.: Akropol'. 2010. – 439 s.
2. Vojskunjij A. E. Psixologiya i iskusstvenny`j intellekt: novy`j 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. Obraz budushhego, social`ny`j optimizm i dolgosrochnaya orientaciya rossijan: 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 potrebitel'ej: issledovanie vzaimosvyazi (na primere produktov pitaniya otechestvenny'x proizvoditelej) // Nacional'ny'j psixologicheskij zhurnal. 2021. № 4(44). S. 39–52. doi: 10.11621/npj.2021.0404
9. Noack 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. Noack 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>

*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)*

*Ноак Наталья 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)*

#### **Ключевые слова**

Социальные представления, проблема центральности, методы диагностики, прототипический анализ, лексико-метрические методы.

***Natalia Noack, Anastasia Volkova, Tatiana Kostina, Social perceptions of Russian society about artificial intelligence: methodological aspects***

#### **Keywords**

Social representations, the problem of centrality, diagnostic methods, prototypical analysis, lexico-metric methods.

DOI: 10.34706/DE-2023-05-03

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

#### **Abstract.**

The article is a continuation of Part 1, emphasizing the problem of research methodology, as well as the centrality of the elements of the structure of social ideas about AI. From these positions, various methods of studying social representations are presented, with an exposition of the ambiguous results obtained with their application. It is shown that the core components of the structure, isolated using the method of 10% consistency of responses, represent a relatively consistent and stable basis for the representations of Russian respondents, represent the fundamental properties of the object of representation, and the elements of the periphery are a generalization of these properties and the designation of possible behavioral strategies. The central core encourages consensus, the psychological community of the group, but depends on the diagnostic method used.