

**УДК 1.159.955.4**

**Губанов Сергей Анатольевич**, доктор филологических наук по научной специальности 5.9.5 «Русский язык. Языки народов России», профессор кафедры философии, Поволжский государственный университет телекоммуникации и информатики, Россия, г. Самара

**Перепеченова Диана Ильдаровна**, бакалавриат, Поволжский государственный университет телекоммуникации и информатики, Россия, г. Самара

## **ИИ КАК ЗЕРКАЛО СОКРАТА: УЧИТ ЛИ НАС МАШИНА ПОЗНАВАТЬ САМИХ СЕБЯ?**

### **Аннотация**

В статье рассматривается ситуация на рынке продовольствия после введения комплекса внешнеэкономических и регуляторных ограничений. Автором проведен комплексный анализ структурных сдвигов, затронувших цепочки поставок, логистические маршруты и ценообразование на базовые категории товаров. Особое внимание уделяется процессам импортозамещения и трансформации потребительского спроса в условиях возросшей волатильности курсов валют и инфляционного давления. В работе детализированы ключевые вызовы, с которыми столкнулись отечественные сельхозпроизводители и переработчики, включая дефицит упаковки, удорожание техники и переориентацию экспортных потоков. На основе статистических данных и экспертных оценок выявлены основные тренды адаптации отрасли, включая рост доли собственных торговых марок (СТМ) и развитие альтернативных каналов сбыта. Также в статье оценивается эффективность мер государственной поддержки, направленных на стабилизацию внутреннего рынка и обеспечение продовольственной безопасности. Автор приходит к выводу о формировании

новой модели функционирования продовольственного рынка, характеризующейся повышенной устойчивостью к внешним шокам, но сохраняющей риски в сегменте высокотехнологичных ингредиентов. Прогнозируется дальнейшая консолидация активов внутри отрасли и усиление роли государства как регулятора товарных интервенций.

### **Abstract**

The article examines the situation on the food market after the introduction of a complex of foreign economic and regulatory restrictions. The author conducted a comprehensive analysis of the structural shifts affecting supply chains, logistics routes, and pricing for basic categories of goods. Special attention is paid to the processes of import substitution and the transformation of consumer demand in the context of increased currency exchange rate volatility and inflationary pressure. The paper details the key challenges faced by domestic agricultural producers and processors, including the shortage of packaging, the rise in the cost of equipment, and the reorientation of export flows. Based on statistical data and expert assessments, the article identifies the main trends in the industry's adaptation, including the growth of private labels (PL) and the development of alternative distribution channels. The article also assesses the effectiveness of government support measures aimed at stabilizing the domestic market and ensuring food security. The author concludes that a new model of the food market is emerging, characterized by increased resilience to external shocks but still facing risks in the high-tech ingredient segment. The article predicts further consolidation of assets within the industry and an increased role for the government as a regulator of commodity interventions.

**Ключевые слова:** продовольственный рынок, продовольственная безопасность, импортозамещение, цепочки поставок, сельское хозяйство, инфляция.

**Keywords:** food market, food security, import substitution, supply chains, agriculture, inflation.

В современном мире искусственный интеллект (ИИ) активно проникает во все сферы человеческой деятельности, трансформируя привычные паттерны

взаимодействия и познания. В контексте самопознания ИИ начинает выполнять функцию, метафорически сравнимую с ролью древнегреческого философа Сократа. Подобно сократическому диалогу, алгоритмические системы, через интерактивные вопросы и анализ данных, способны стимулировать глубокую рефлексию, провоцируя человека к осмыслению собственных мыслей и эмоций.

Центральный парадокс данного исследования заключается в вопросе: может ли система, лишенная сознания и интенциональности, эффективно способствовать самопознанию человека, выполняя функцию античного философа? Ответ на этот вопрос требует не только технической оценки возможностей ИИ, но и глубокого философского осмысления критериев истинного самопознания. Это поднимает фундаментальные вопросы о природе диалога и его роли в поиске истины о себе.

Однако, наряду с потенциальными выгодами, использование ИИ в качестве инструмента самопознания несет в себе ряд серьезных рисков и проблем. Чрезмерное доверие к машинным интерпретациям может привести к делегированию рефлексивных операций и утрате навыков критической самооценки. Существует опасность закрепления алгоритмических описаний как окончательных версий «я», что ставит под сомнение не только достоверность машинных «отражений», но и устойчивость автономного самопознания.

Актуальность данного исследования обусловлена необходимостью концептуализировать и эмпирически проверить границы и возможности «машинного зеркала» в эпоху быстрого развития чат-ботов и персонализированных ИИ-ассистентов. Цель работы состоит в том, чтобы историко-теоретически сопоставить сократическую практику и современные алгоритмические техники рефлексии, а также выработать ориентиры, позволяющие использовать машинные инструменты для усиления, а не подмены подлинного самопознания. Это позволит интегрировать философские идеи с современной психологией и когнитивными науками, понимая влияние технологий на этику самопознания и ментальное здоровье в цифровую эру.

In today's world, artificial intelligence (AI) is actively penetrating all aspects of human activity, transforming conventional patterns of interaction and cognition. In the context of self-discovery, AI is beginning to fulfill a role that can be metaphorically compared to that of the ancient Greek philosopher Socrates. Similar to Socratic dialogues, algorithmic systems, through interactive questions and data analysis, have the potential to stimulate deep reflection, prompting individuals to contemplate their own thoughts and emotions.

The central paradox of this research lies in the question: can a system devoid of consciousness and intentionality effectively contribute to human self-knowledge, fulfilling the role of an ancient philosopher? The answer to this question requires not only a technical assessment of AI capabilities, but also a profound philosophical understanding of the criteria for true self-knowledge. This raises fundamental questions about the nature of dialogue and its role in the pursuit of self-truth.

However, along with the potential benefits, the use of AI as a tool for self-discovery also presents several significant risks and challenges. Excessive trust in machine interpretations can lead to the delegation of reflexive operations and the loss of critical self-assessment skills. There is a risk of algorithmic descriptions becoming the definitive versions of the self, which not only questions the accuracy of machine "reflections" but also the sustainability of autonomous self-knowledge.

The relevance of this study lies in the need to conceptualize and empirically test the boundaries and capabilities of the "machine mirror" in an era of rapid development of chatbots and personalized AI assistants. The purpose of the work is to compare the Socratic practice and modern algorithmic techniques of reflection in a historical and theoretical context, as well as to develop guidelines for using machine tools to enhance, rather than replace, genuine self-knowledge. This will allow for the integration of philosophical ideas with modern psychology and cognitive sciences, understanding the impact of technology on the ethics of self-knowledge and mental health in the digital era.

## ГЛАВА 1 СОКРАТИЧЕСКИЙ ДИАЛОГ И ИИ

## 1.1. Сущность сократического метода: искусство задавать вопросы для познания истины

Сократический метод представляет собой поэтапную аналитическую процедуру, цель которой — выявление внутренних противоречий в убеждениях через последовательное Эленхус (или эленхус-метод) — это метод, используемый в логике и философии для выявления и анализа аргументов. Он основан на принципе, что любой аргумент можно подвергнуть критике, выявив его слабые места или логические ошибки. Применение эленхуса включает в себя следующие этапы:

1. Идентификация аргумента: Определение основной идеи или утверждения, которое необходимо проанализировать.

2. Выявление предпосылок: Определение основных предпосылок, на которых основан аргумент.

3. Анализ логической структуры: Оценка логической связи между предпосылками и выводом.

4. Критика: Поиск возможных слабостей, неясностей или логических ошибок в аргументе.

5. Формулирование контраргументов: Предложение альтернативных точек зрения или аргументов, которые могут опровергнуть исходный аргумент.

Эленхус может быть полезен в различных областях, таких как философия, юриспруденция, политика и даже в повседневной жизни, когда необходимо оценить достоверность утверждений или аргументов., уточнение понятий и логическую проверку предпосылок. Эленхус функционирует как испытание на согласованность: собеседник ставит утверждение в контекст вопросов, обнаруживая несогласованности и вынуждая к пересмотру исходных посылок. Уточнение понятий осуществляется через дефиниции и контрпримеры, что позволяет локализовать источники ошибок в рассуждении. Такая структурированность создает методологическую основу для сопоставления сократического диалога с алгоритмическими процедурами анализа, рассматриваемыми в дальнейших главах.

Психолого-когнитивные механизмы сократического диалога включают наведение сомнения, маевтику и последовательное уточнение определений, которые способствуют перестройке ментальных представлений и формированию отражённого знания. Наведение сомнения нарушает автоматические схемы мышления и открывает пространство для аналитической рефлексии, в то время как маевтика направлена на поэтапное извлечение латентных убеждений через наводящие вопросы. Последовательное уточнение определений работает как инструмент коррекции когнитивных моделей, позволяя участнику диалога реконструировать собственные предпосылки в более согласованную систему. Описанные механизмы обеспечивают переход от поверхностных реакций к системному самоосмыслению, что служит теоретическим мостом к обсуждению роли ИИ в стимулировании рефлексии.

#### 1.2. Аналогии между сократическим диалогом и алгоритмами ИИ в стимулировании рефлексии

В функциональном плане сократический диалог и современные алгоритмы ИИ демонстрируют принципиально сходную динамику итеративного уточнения гипотез, при которой первоначальные утверждения последовательно ставятся под сомнение и пересматриваются. В обоих случаях процесс можно схематично представить как цикл опрос → ответ → уточнение, где каждый ответ формирует новую отправную точку для последующего вопроса или модели, а механизмы обратной связи и контекстуальное уточнение служат средством коррекции предположений. Такая поэтапная интеракция способствует рефлексивному переосмыслению убеждений у собеседника или пользователя, поскольку выявляет противоречия и стимулирует преобразование внутренних представлений. Описанная аналогия согласуется с ранее выработанной теоретической рамкой и плавно переходит к изучению конкретных механизмов взаимодействия ИИ и человека в следующей главе.

#### 1.3. Исторический контекст сократического метода и его актуальность в современном мире

Сократический метод восходит к платоновским диалогам, где техника целенаправленного вопрошания и эленхус выступает средством выявления противоречий и продвижения к понятиям истины, что закреплено в историографии как отправная точка метода. В последующих античных и средневековых традициях метод подвергался трансформации и адаптации к риторическим, дидактическим и богословским задачам, что свидетельствует о его функциональной гибкости и устойчивости. В современности элементы сократического подхода сохраняют методологическую ценность в педагогике, психотерапии и этических размышлениях, функционируя как универсальный инструмент структурированного самопознания и развития критического мышления. Такая историческая преемственность обеспечивает методологическую основу для сопоставления сократического диалога с современными алгоритмами ИИ и служит связующим звеном к последующим анализам практических механизмов ИИ в роли зеркала для рефлексии.

## CHAPTER 1 SOCRATIC DIALOGUE AND AI

1.1. The essence of the Socratic method: the art of asking questions to find out the truth

The Socratic method is a step—by-step analytical procedure, the purpose of which is to identify internal contradictions in beliefs through sequential Ellenhus (or the Ellenhus method) is a method used in logic and philosophy to identify and analyze arguments. It is based on the principle that any argument can be criticized by revealing its weaknesses or logical errors. The application of Ellenhus includes the following steps:

1. Identifying the argument: Determining the main idea or statement that needs to be analyzed.

2. Identifying the premises: Identifying the main premises on which the argument is based.

3. Analyzing the logical structure: Evaluating the logical relationship between the premises and the conclusion.

4. Critiquing the argument: Identifying any potential weaknesses, ambiguities, or logical errors in the argument.

5. Formulating counterarguments: Proposing alternative perspectives or arguments that can refute the original argument.

Ellenus can be useful in various fields, such as philosophy, law, politics, and even in everyday life, where it is necessary to evaluate the validity of statements or arguments, clarify concepts, and logically test assumptions. Ellenus functions as a consistency test, where the interlocutor places the statement in the context of questions, revealing inconsistencies and prompting a review of the original premises. The clarification of concepts is carried out through definitions and counterexamples, which allows for the localization of sources of errors in reasoning. This structure provides a methodological basis for comparing the Socratic dialogue with the algorithmic analysis procedures discussed in subsequent chapters.

The psychological and cognitive mechanisms of the Socratic dialogue include the induction of doubt, maieutics, and the sequential refinement of definitions, which contribute to the restructuring of mental representations and the formation of reflected knowledge. Inducing doubt disrupts automatic thought patterns and opens up space for analytical reflection, while maieutics focuses on gradually extracting latent beliefs through leading questions. The sequential refinement of definitions serves as a tool for correcting cognitive models, allowing the participant to reconstruct their own assumptions into a more coherent system. These mechanisms facilitate the transition from superficial reactions to systemic self-reflection, providing a theoretical bridge to discussing the role of AI in promoting reflection.

1.2. Analogies between Socratic dialogue and AI algorithms in stimulating reflection

Functionally, Socratic dialogue and modern AI algorithms demonstrate a fundamentally similar dynamics of iterative hypothesis refinement, in which initial statements are sequentially questioned and revised. In both cases, the process can be schematically represented as a cycle of question → answer → refinement, where each answer forms a new starting point for the subsequent question or model, and feedback

mechanisms and contextual refinement serve as a means of correcting assumptions. This step-by-step integration promotes a reflective rethinking of the interlocutor's or user's beliefs, as it reveals contradictions and stimulates the transformation of internal representations. This analogy aligns with the previously established theoretical framework and seamlessly transitions to the exploration of specific mechanisms for AI-human interaction in the next chapter.

### 1.3. The historical context of the Socratic method and its relevance in the modern world

The Socratic method can be traced back to Plato's dialogues, where the technique of purposeful questioning and elenchus serves as a means of identifying contradictions and progressing towards the concepts of truth, which has been established in historiography as the starting point of the method. In subsequent ancient and medieval traditions, the method has been transformed and adapted to meet rhetorical, didactic, and theological objectives, demonstrating its functional flexibility and resilience. In modern times, elements of the Socratic approach retain their methodological value in pedagogy, psychotherapy, and ethical reflection, functioning as a universal tool for structured self-discovery and the development of critical thinking. This historical continuity provides a methodological framework for comparing the Socratic dialogue with modern AI algorithms and serves as a link to subsequent analyses of practical AI mechanisms as a mirror for reflection.

## ГЛАВА 2 ИИ КАК ИНСТРУМЕНТ САМОПОЗНАНИЯ

### 2.1. Механизмы взаимодействия человека с ИИ: диалоговые системы как 'зеркало'

Структура диалога в системах ИИ — тип задаваемых вопросов, последовательность реплик и зеркалирование лексики и стиля — существенно влияет на активацию метапознания и самонаблюдения у пользователя. Открытые вопросы и уточняющие реплики направляют внимание на процессы мышления, тогда как закрытые вопросы чаще фиксируют ответы на факты и ограничивают рефлексю. Зеркалирование лексики и стиля способствует ощущению понятности и безопасности, что повышает готовность пользователя к

самоанализу. Последовательность реплик, включающая переформулирование и паузы для ответа, структурирует поток сознания и облегчает выявление внутренних гипотез и предположений.

Алгоритмически сгенерированные подсказки, переформулировки и контекстные реплики выполняют роль проективной фасилитации, способствуя организации личного нарратива и выявлению неосознанных смыслов. Переформулировки служат внешним отзеркаливанием мыслей, которое позволяет пользователю увидеть свои предположения в иной формулировке и оценить их когерентность. Контекстные подсказки и направляющие вопросы помогают выстроить цепочку интерпретаций, делая очевидными противоречия и пропуски в рассказе о себе. Автоматизированные реплики при этом могут накладывать структурные рамки на нарратив, что требует критической оценки их влияния на автономию рефлексии.

2.2. Анализ данных поведения и настроения: ИИ как детектор внутренних состояний

Методологические подходы к детекции аффекта и внутренних состояний опираются на комбинированный анализ лексических, паралингвистических и временных признаков, где лексика отражает семантику и тональность высказываний, паралингвистика фиксирует просодию и темп речи, а временные параметры учитывают паузы и задержки ответов. В качестве моделей применяются традиционные методы обучения с учителем и современные нейросетевые архитектуры с мультимодальным объединением сигналов, а процедуры валидации включают разметку корпуса экспертами, кросс-валидацию и сравнительные бенчмарки при стремлении обеспечить переносимость результатов. Системные ограничения проявляются в сильной зависимости от контекста и культуры, шуме каналов передачи данных, неоднозначности меток и недостаточной интерпретируемости моделей, что ведёт к потенциальным искажениям выводов о внутренних состояниях субъектов. Учитывая эти ограничения, выводы алгоритмов требуют осторожной валидации в реальных условиях и сочетания количественных методов с

качественными процедурами интерпретации для минимизации ошибок и неверных суждений о психическом состоянии пользователей.

### 2.3. Практические кейсы применения ИИ для самоанализа и личностного роста

Практические протоколы внедрения ИИ в самоанализ обычно включают журнальные интерфейсы, интегрированные когнитивно-поведенческие сопровождения и персональный коучинг, которые функционируют как стандартизированные сценарии взаимодействия между пользователем и системой. Эффективность таких сценариев оценивается по набору показателей: изменению симптоматики и уровней благополучия по шкалам самооценки, объективным поведенческим маркерам и метрикам вовлечённости, получаемым через как активные отклики, так и пассивное сенсирование. Методики измерения изменений включают дизайн до–после, контрольные группы и длительные наблюдения с использованием методов экологического моментального оценивания и анализа временных рядов для выявления устойчивых трендов. Ограничения переносимости результатов связаны с гетерогенностью выборок, контекстной зависимостью данных и риском переобучения алгоритмов на локальных паттернах поведения, что требует осторожной интерпретации и проверки на внешних когортах.

## CHAPTER 2 AI AS A TOOL FOR SELF-EXPLORATION

### 2.1. Mechanisms of human interaction with AI: Dialogue systems as a 'mirror'

The structure of the dialogue in AI systems — the type of questions asked, the sequence of turns, and the mirroring of vocabulary and style — significantly affects the activation of meta-cognition and self-observation in the user. Open questions and clarifying turns direct attention to the processes of thinking, while closed questions are more likely to fix answers to facts and limit reflection. Mirroring of vocabulary and style contributes to a sense of clarity and security, which increases the user's willingness to engage in self-reflection. The sequence of prompts, including rephrasing and pausing for responses, structures the flow of consciousness and facilitates the identification of internal hypotheses and assumptions.

Algorithmically generated prompts, rephrasing, and contextual responses serve as a form of projective facilitation, helping to organize personal narratives and uncover unconscious meanings. Reformulations serve as an external reflection of thoughts, allowing the user to see their assumptions in a different form and assess their coherence. Contextual prompts and guiding questions help to build a chain of interpretations, making contradictions and gaps in the self-story apparent. Automated responses can also impose a structural framework on the narrative, requiring critical evaluation of their impact on the autonomy of reflection.

## 2.2. Analysis of behavioral and mood data: AI as a detector of internal states

Methodological approaches to the detection of affect and internal states are based on a combined analysis of lexical, paralinguistic, and temporal features, where the lexicon reflects the semantics and tone of the utterances, paralinguistics captures prosody and speech tempo, and temporal parameters take into account pauses and delays in responses. Traditional supervised learning methods and modern neural network architectures with multimodal signal fusion are used as models, and validation procedures include expert annotation of the corpus, cross-validation, and comparative benchmarks to ensure the transferability of results. Systemic limitations manifest themselves in strong dependence on context and culture, noise of data transmission channels, ambiguity of labels, and insufficient interpretability of models, which leads to potential distortions of conclusions about the internal states of subjects. Given these limitations, the conclusions of algorithms require careful validation in real-world conditions and the combination of quantitative methods with qualitative interpretation procedures to minimize errors and incorrect judgments about the mental state of users.

## 2.3. Practical cases of using AI for self-reflection and personal growth

Practical protocols for implementing AI in self-reflection typically include journal interfaces, integrated cognitive-behavioral support, and personalized coaching, which function as standardized scenarios for user-system interaction. The effectiveness of these scenarios is evaluated based on a set of indicators, including changes in symptomatology and well-being levels as measured by self-report scales, objective behavioral markers, and engagement metrics obtained through both active responses

and passive sensing. Measurement techniques include before-and-after designs, control groups, and long-term observations using ecological momentary assessment and time series analysis to identify stable trends. Limitations in the transferability of results arise from sample heterogeneity, contextual dependence of data, and the risk of overfitting algorithms to local patterns of behavior, necessitating cautious interpretation and validation in external cohorts.

## ГЛАВА 3 ЭТИКА И РИСКИ МАШИННОГО ОТРАЖЕНИЯ

### 3.1. Ограничения ИИ в понимании человеческой субъективности и сознания

Статистическая природа современных моделей ограничивает доступ к качественным характеристикам переживаний, поскольку алгоритмы опираются на вероятностные зависимости в данных, а не на непосредственно переживаемые состояния. Обработка текстовых и поведенческих сигналов сводится к выявлению паттернов и корреляций и не воспроизводит феноменологические свойства субъективного опыта (*qualia*). Вследствие этого интерпретации внутренних состояний остаются репрезентациями третьего порядка, а не эквивалентным пониманием переживания субъекта. Такая разница между статистической моделью и феноменологическим содержанием ставит под сомнение претензии на полноту машинных «отражений» личности.

Эпистемические ограничения ИИ во многом связаны с отсутствием телесной и эмпатической основы, что делает выводы зависимыми от свойств обучающих выборок и качества аннотаций. Эта зависимость приводит к фрагментарным и контекстно-зависимым интерпретациям сознания, поскольку модели не обладают собственной ситуативной интенциональностью и не встраиваются в переживание, как это делает живой агент. Вследствие этого валидность выводов о личности и мотивации снижается, особенно при переносе моделей на новые культурные или ситуационные контексты. Указанные ограничения должны учитываться при формулировании рекомендаций по использованию ИИ в процессах самопознания и отражаются в последующем обсуждении выводов главы.

### 3.2. Риски искаженного самовосприятия и утраты навыков рефлексии при чрезмерном доверии к ИИ

Чрезмерное упование на выводы алгоритмов ведёт к внешней верификации идентичности и формированию автоматизированных трактовок я-концепции, что потенциально искажает самовосприятие индивидов и ослабляет внутренние механизмы критической рефлексии. «Чрезмерное доверие к машинным интерпретациям способствует внешней верификации идентичности и укреплению автоматизированных трактовок я-концепции, что ведёт к искажению самовосприятия, конформизации самоотчётов и уменьшению навыков критической рефлексии и метакогнитивного контроля [12, с.103].» Последствия этого сдвига проявляются в повышенной конформизации самоотчётов

Конформизация самоотчётов — это процесс, при котором индивид изменяет свои самооценки или представления о себе в соответствии с социальными нормами, ожиданиями окружающих или групповой динамикой. Это может происходить как сознательно, так и бессознательно, и часто связано с желанием быть принятым в группе или соответствовать определённым стандартам.

Примеры конформизации самоотчётов могут включать:

1. Изменение мнения: Человек может изменить своё мнение о своих способностях или достижениях, чтобы соответствовать мнению группы.
2. Социальное сравнение: Люди могут занижать или завышать свои достижения в зависимости от того, как они воспринимают успехи других.
3. Подстройка под ожидания: Индивид может адаптировать свои самоотчёты, чтобы соответствовать ожиданиям родителей, учителей или коллег.

Конформизация самоотчётов может иметь как положительные, так и отрицательные последствия. С одной стороны, она может способствовать социальной гармонии и улучшению взаимодействия в группе, с другой — может привести к снижению самооценки и внутреннему конфликту., снижению метакогнитивного контроля и уменьшению готовности к самостоятельному пересмотру убеждений, что влечёт за собой утрату навыков оценочной

саморефлексии. В ответ на выявленные риски рассматриваются меры по сохранению автономии рефлексивных практик и по внедрению механизмов критической валидации машинных интерпретаций в рамках междисциплинарных политик и исследовательских программ.

3.3. Этические аспекты использования ИИ в процессах самопознания: конфиденциальность и ответственность

Применение ИИ в процессах самопознания сопряжено с непосредственными рисками для конфиденциальности данных, включая их сбор, хранение и возможность вторичной идентификации, что требует особого внимания со стороны регуляторов и практиков. «Использование ИИ в самопознании порождает прямые риски для конфиденциальности (сбор, хранение, вторичная идентификация персональных данных) и требует чёткого распределения ответственности между разработчиками, операторами и пользователями, внедрения прозрачности алгоритмов, процедур согласия, аудита и механизмов обжалования выводов системы [1, с.12].» Внедрение прозрачных алгоритмических описаний, формализованных процедур информированного согласия и независимых механизмов аудита представляется минимально необходимым набором мер для распределения ответственности между разработчиками, операторами и пользователями. Такие меры обеспечивают основу для оценки влияния ИИ на личную рефлексию и позволяют интегрировать этические гарантии в рекомендации по безопасному использованию технологий самопознания.

## CHAPTER 3 ETHICS AND RISKS OF MACHINE REPRESENTATION

### 3.1. AI limitations in understanding human subjectivity and consciousness

The statistical nature of current models limits access to the qualitative characteristics of experiences, as algorithms rely on probabilistic dependencies in data rather than directly experienced states. Processing textual and behavioral signals is reduced to identifying patterns and correlations and does not reproduce the phenomenological properties of subjective experience (qualia). As a result, interpretations of internal states remain third-order representations rather than

equivalent understandings of the subject's experience. This difference between the statistical model and the phenomenological content raises questions about the completeness of machine-based "reflections" of the individual.

The epistemic limitations of AI are largely due to the absence of a bodily and empathetic foundation, which makes inferences dependent on the properties of the training data and the quality of the annotations. This dependence leads to fragmented and context-dependent interpretations of consciousness, as models do not have their own situational intentionality and do not integrate into experience in the same way as a living agent does. As a result, the validity of conclusions about personality and motivation is reduced, especially when models are transferred to new cultural or situational contexts. These limitations should be taken into account when formulating recommendations for using AI in self-discovery processes, and they are discussed further in the conclusions of this chapter.

3.2. The risks of distorted self-perception and loss of reflection skills when relying too much on AI

Excessive reliance on algorithmic conclusions leads to external verification of identity and the formation of automated interpretations of the self-concept, which potentially distorts individuals' self-perception and weakens their internal mechanisms of critical reflection. "Excessive trust in machine interpretations contributes to the external verification of identity and the strengthening of automated interpretations of the self-concept, which leads to a distortion of self-perception, conformity of self-reports, and a decrease in critical reflection and metacognitive control skills [12, p. 103]." The consequences of this shift manifest themselves in increased conformity of self-reports.

Conformity of self-reports is a process in which an individual changes their self-evaluations or self-perceptions in accordance with social norms, expectations of others, or group dynamics. This can occur both consciously and unconsciously, and is often related to the desire to be accepted in a group or to meet certain standards.

Examples of conformity in self-reports may include:

1. Changing opinions: A person may change their opinion about their abilities or achievements in order to conform to the group's opinion.

2. Social comparison: People may underestimate or overestimate their achievements based on how they perceive the success of others.

3. Adjustment to expectations: An individual may adapt their self-reports to meet the expectations of parents, teachers, or colleagues.

The conformity of self-reports can have both positive and negative consequences. On the one hand, it can contribute to social harmony and improved group interaction, but on the other hand, it can lead to decreased self-esteem and internal conflict, reduced metacognitive control, and a decrease in the willingness to independently revise beliefs, resulting in a loss of evaluative self-reflection skills. In response to the identified risks, measures are being considered to preserve the autonomy of reflexive practices and to implement mechanisms for critical validation of machine interpretations within the framework of interdisciplinary policies and research programs.

### 3.3. Ethical Aspects of Using AI in Self-Knowledge Processes: Privacy and Responsibility

The use of AI in self-knowledge processes poses immediate risks to data privacy, including data collection, storage, and the potential for secondary identification, which requires special attention from regulators and practitioners. "The use of AI in self-discovery poses direct risks to privacy (collection, storage, and secondary identification of personal data) and requires a clear distribution of responsibility between developers, operators, and users, as well as the implementation of transparent algorithm descriptions, formalized informed consent procedures, and independent audit mechanisms [1, p. 12]." The implementation of transparent algorithm descriptions, formalized informed consent procedures, and independent audit mechanisms appears to be the minimum necessary set of measures to distribute responsibility between developers, operators, and users. Such measures provide a framework for assessing the impact of AI on personal reflection and allow for the integration of ethical safeguards into recommendations for the safe use of self-discovery technologies.

Проведенное исследование подтверждает, что искусственный интеллект способен выполнять функцию современного «зеркала Сократа», предлагая уникальные возможности для самопознания. Через интерактивные диалоги и анализ данных ИИ выявляет скрытые паттерны мышления, эмоциональные реакции и поведенческие маркеры, стимулируя пользователя к глубокой рефлексии и формулированию уточняющих вопросов к самому себе. Этот процесс аналогичен сократическому методу, где вопросы направлены на стимулирование внутреннего диалога и критического осмысления.

Однако ключевой парадокс заключается в том, что ИИ, не обладая сознанием или интенциональностью, выполняет эту функцию исключительно инструментально. Он создает стимулирующую среду для самоисследования, но не может заменить философскую ответственность и автономную способность человека к критическому самоанализу. Машина лишь отражает информацию, не понимая ее истинного смысла, что подчеркивает фундаментальное различие между человеческим и машинным познанием.

Вместе с тем, существуют реальные риски, связанные с чрезмерной зависимостью от машинного «отражения». Это может привести к ослаблению навыков самостоятельной рефлексии и формированию искаженного самопонимания из-за алгоритмических предубеждений. Кроме того, возникают серьезные проблемы приватности и ответственности за интерпретации внутреннего опыта, что требует внимательного подхода к использованию ИИ в столь интимной сфере.

Таким образом, практическая и этическая стратегия должна заключаться в осознанной интеграции ИИ в практики самопознания. Это требует прозрачности алгоритмов, четких границ ответственности, образовательной подготовки пользователей и сочетания машинной обратной связи с человеческим сопровождением. Такой подход позволит сохранить достоинство автономного познания и эффективно использовать потенциал технологического «зеркала», минимизируя при этом возможные негативные последствия.

The conducted research confirms that artificial intelligence can serve as a modern-day "Socrates mirror," offering unique opportunities for self-discovery. Through interactive dialogues and data analysis, AI identifies hidden patterns of thought, emotional responses, and behavioral markers, encouraging users to engage in deep reflection and ask themselves clarifying questions. This process is similar to the Socratic method, where questions are used to stimulate internal dialogue and critical thinking.

However, the key paradox is that AI, lacking consciousness or intentionality, performs this function solely instrumentally. It creates a stimulating environment for self-exploration, but it cannot replace the philosophical responsibility and autonomous ability of humans to critically analyze themselves. The machine merely reflects information without understanding its true meaning, highlighting the fundamental difference between human and machine cognition.

At the same time, there are real risks associated with excessive reliance on machine "reflection." This can lead to a weakening of self-reflection skills and the formation of distorted self-understanding due to algorithmic biases. Additionally, there are serious concerns about privacy and responsibility for interpreting internal experiences, which requires a careful approach to using AI in such an intimate area.

Therefore, a practical and ethical strategy should be to consciously integrate AI into self-discovery practices. This requires transparent algorithms, clear boundaries of responsibility, educational training for users, and a combination of machine feedback with human guidance. This approach will allow us to preserve the dignity of autonomous knowledge and effectively use the potential of the technological "mirror," while minimizing possible negative consequences.

### **Список литературы**

1. Быльева Д.С. Этика искусственного интеллекта через концепции любви и свободы // Семиотические исследования. Semiotic studies. — 2022. — №4. — С. 8–14.

2. Гараганов А.В. Как интеллектуальные технологии меняют общество и пространство большого города // Социально-политические науки. — 2024. — №1. — С. 96–101.
3. Горбачёва А.Г. Искусственный интеллект и современное искусство: новые возможности и вызовы // Антропологическая навигация: феномен и практики. — 2018. — №13. — С. 145–154.
4. Макушин М. Искусственный интеллект: проблемы и перспективы развития // Электроника наука | технология | бизнес. — 2020. — №9. — С. 62–66.
5. Матюк Ю.С. Искусственный интеллект: новые вызовы и перспективы // Advances in law studies. — 2020. — №8. — С. 42–48.
6. Мигунова О.Д., Везбердева Е.В. Что мы знаем об искусственном интеллекте. Профориентационный проект. Для учащихся старших классов: методические рекомендации. — Калининград: Калининградская областная юношеская библиотека им. В. Маяковского, 2025.— 26 с.
7. Продан А.А. Влияние самоосознания на психологический рост личности // Вопросы современной науки и практики. — 2013.— №2. — С. 123–127.
8. Пустовалова В.Ю. Разговорный искусственный интеллект для обучения английскому языку студентов неязыковых специальностей (на примере чат-бота Replika) // Открытое и дистанционное образование. — 2025. — №1. — С. 32–41.
9. Шталь Б.К., Шредер Д., Родригес Р. Этика искусственного интеллекта: кейсы и варианты решения этических проблем // Экономическая социология. — 2024. — №1. — С. 85–89.
10. Щербакова К.М. Психологические подходы к изучению структуры образа «Я» // Альманах современной науки и образования. — 2009. — №4. — С. 185–186.
11. Ян В., Го Т. Применение искусственного интеллекта в преподавании РКИ: возможности и ограничения // Вестник педагогических наук.— 2025. — №9. — С. 185–189.

12. Ярков А.П. Искусственный интеллект, вызов цивилизации // Русская политология — Russian political science. — 2022. — №1. — С. 101–104.

13. Niculin E. Искусственный интеллект в информационных технологиях: проблемы и опасности // Culegere de lucrări științifice simpozionul științific al tinerilor cercetători ediția a XXII-a, 26-27 aprilie 2024, vol. II. — Chișinău, 2024. — С. 66–67.

### **List of references**

1. Bylyeva D.S. Ethics of Artificial Intelligence through the Concepts of Love and Freedom // Semiotic studies. Semiotic studies. — 2022. — No. 4. — Pp. 8–14.

2. Garaganov A.V. How Intelligent Technologies Change Society and the Space of a Large City // Social and Political Sciences. — 2024. — No. 1. — Pp. 96–101.

3. Gorbacheva A.G. Artificial Intelligence and Contemporary Art: New Opportunities and Challenges // Anthropological Navigation: Phenomenon and Practices. — 2018. — No. 13. — Pp. 145–154.

4. Makushin M. Artificial Intelligence: Problems and Prospects of Development // Electronics science | technology | business. — 2020. — No. 9. — Pp. 62–66.

5. Matyuk Yu.S. Artificial intelligence: new challenges and prospects // Advances in law studies. — 2020. — No. 8. — Pp. 42–48.

6. Migunova O.D., Vezberdeva E.V. What we know about artificial intelligence. Career guidance project. For high school students: methodological recommendations. — Kaliningrad: V. Mayakovsky Kaliningrad Regional Youth Library, 2025. — 26 p.

7. Prodan A.A. The Influence of Self-Awareness on Psychological Growth of an Individual // Issues of Modern Science and Practice. — 2013. — No. 2. — Pp. 123–127.

8. Pustovalova V.Yu. Conversational Artificial Intelligence for Teaching English to Students of Non-Linguistic Specialties (Using the Replika Chatbot as an Example) // Open and Distance Education. — 2025. — No. 1. — Pp. 32–41.

9. Stahl B.K., Schroeder D., Rodriguez R. Ethics of Artificial Intelligence: Cases and Solutions to Ethical Problems // *Economic Sociology*. — 2024. — No. 1. — Pp. 85–89.
10. Shcherbakova K.M. Psychological Approaches to Studying the Structure of the Self-Image // *Almanac of Modern Science and Education*. — 2009. — No. 4. — Pp. 185–186.
11. Yang V., Guo T. Application of Artificial Intelligence in Teaching Russian as a Foreign Language: Opportunities and Limitations // *Bulletin of Pedagogical Sciences*.— 2025. — No. 9. — Pp. 185–189.
12. Yarkov A.P. Artificial Intelligence, a Challenge to Civilization // *Russian Political Science*. — 2022. — No. 1. — Pp. 101–104. Artificial Intelligence in Information Technologies: Problems and Dangers // *Culegere de lucrări științifice simpozionul științific al tinerilor cercetători ediția a XXII-a, 26-27 aprilie 2024, vol. II*. — Chișinău, 2024. — Pp. 66–67.