МРНТИ 02.11.21 https://doi.org/10.48010/2025.3/1999-5911.06

ФЕНОМЕНОЛОГИЯ ИНТЕРСУБЪЕКТИВНОСТИ И КОГНИТИВНАЯ АРХИТЕКТУРА ОБЩЕГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА*

[1]Дунаев В.Ю., [2]Курганская В.Д.,[3]Сагикызы А.*
[1,2,3]Институт философии, политологии и религиоведения КН МНВО РК
*(*Алматы, Казахстан)
[1]vlad.dunaev2011@yandex.kz, [2]vkurganskaya@mail.ru,[3]ayazhan@list.ru

[1]Dunaev V., [2]KurganskayaV., [3]Sagikyzy A.
[1,2,3]Institute for Philosophy, Political Science and Religious Studies
of the CS MSHE RK (Almaty, Kazakhstan)
[1]vlad.dunaev2011@yandex.kz, [2]vkurganskaya@mail.ru,[3]ayazhan@list.ru

Аннотация. Одним из необходимых теоретико-методологических ресурсов для построения систем искусственного интеллекта являются феноменологические исследования интенциональных структур субъективной реальности. В статье даётся краткое изложение принципов классической феноменологии интерсубъективности и проводится анализ их применения в одном из проектов построения когнитивной архитектуры общего искусственного интеллекта. В феноменологии Э. Гуссерляинтенциональные структуры трансцендентальной интерсубъективности наделяются статусом онтологической и смысловой первоосновы реальности. Однако в концепции Э. Гуссерля эти структуры интерпретируются как имманентные взаимосвязи поля трансцендентально чистого сознания. Для Гегеля структуру формообразований сознания образует развитие совокупной духовной культуры человечества. Первичным, фундаментальным фактом опыта самосознания является не cogito, но опыт соотношения бытия-для-себя и бытия-для-иного. В самом своём бытии я зависим от другого, моё сознание в самом его средоточии должно быть опосредствовано другим сознанием.

Подход Гегеля к феноменологии интерсубъективности и её роли в формировании сознания можно рассматривать как философское обоснование концепции построения когнитивной архитектуры интерсубъективности в интернациональном проекте создания системы искусственного интеллекта RoboErgoSum. В этом проекте объективные качества предметов выявляются через возможные действия с этими предметами. Для успешных совместных действий агенты взаимодействия должны уметь координировать свои намерения, планы, цели и действия. В этом же процессе возникает восприятие себя как отличного от окружающей среды – что является основой человеческого самосознания и его роботизированной модели. Разработчики проекта настаивают на том, что в их проекте роботы приобретают сознание и самосознание, формируют

* *Данное исследование финансировалось Комитетом науки Министерства науки и высшего образования Республики Казахстан (ИРН BR21882302 «Казахстанский социум в условиях цифровой трансформации: перспективы и риски»)*

* *Автор-корреспондент* - Сагикызы А ayazhan@list.ru

способности к построению внутренних моделей мира, к рассуждению и размышлению, самоконтролю, целеполаганию и мотивации, достигают понимания происходящего и т.д. В статье делается вывод, что такого рода антропоморфизмы допустимы лишь в качестве метафорического языка дискурса. На деле речь должна идти о моделировании высших психических функций человека (перцептивных, познавательных, аффективных, прогностических) в когнитивной архитектуре искусственного интеллекта. Решающая роль в формировании структур такого рода интегрированной интерсубъективности отводится процессу обучения искусственного интеллекта на основе координации действий робота с действиями человека.

Ключевые слова: интерсубъективность, феноменология, искусственный интеллект, когнитивная архитектура, сознание, робот.

Введение

В коллективной монографии известных российских исследователей, посвящённой социогуманитарным аспектам проблемы искусственного интеллекта, отмечается: «Для моделирования необходимых ОИИ функций требуется исследование и описание специфических когнитивных архитектур естественных процессов мышления. Здесь необходимыми ресурсами для построения моделей ОИИ могут служить результаты психологических и феноменологических исследований ценностно-смысловых, динамических структур субъективной реальности» [1, с. 70-71]. Одним из наиболее перспективных направлений актуализации ресурсов классической и постклассической философии в моделировании интеллектуальных функций является трансформация методологии феноменологических анализов трансцендентальных структур интерсубъективности в технологию конструирования когнитивной архитектуры общего или сильного искусственного интеллекта.

Методы исследования

В проведённом авторами настоящей статьи исследовании применяемых специалистами парадигм, теорий, моделей и алгоритмов конструирования когнитивных структур искусственного интеллекта использованы методы логического и семантического анализа; диалектико-логические принципы объективности, системности, конкретности; методы концептуального моделирования; методология системного, структурно-функционального и факторного анализа; методология интердисциплинарного подхода; методы и концептуальные подходы когнитологии.

Понятие когнитивной архитектуры и основные подходы к её построению

«Когнитивная архитектура – это гипотеза о фиксированных структурах, которые обеспечивают работу разума в естественных или искусственных системах, и о том, как они взаимодействуют – в сочетании со знаниями и навыками, воплощёнными в архитектуре, – для обеспечения разумного поведения в различных сложных средах» [2]. Под когнитивной архитектурой

ISSN 1999-5911 (Print)
ISSN 2960-2173 (Online)

искусственного интеллекта (ИИ) понимаются базовая инфраструктура и общие принципы построения систем обработки данных, позволяющие моделировать ментальные процессы (мышление, восприятие, воображение, эмоции, аффекты и т.д.). К основным строительным блокам всех типов и моделей когнитивной архитектуры ИИ относятся:

– кратковременная и долгосрочная память;

– элементы этих видов памяти и их организация в ментальные структуры;

– процессы и механизмы приобретения, представления, организации, использования и изменения этих элементов и структур [3, p. 141].

Базы знаний, убеждений, ценностей, целей, правил и т.д. могут изменяться и по-разному использоваться при сохранении когнитивной архитектуры. П. Лэнгли, Дж. Лэрд и С. Роджерс проводят прямую аналогию когнитивной архитектуры с архитектурой здания, состоящего из постоянных частей, таких как его фундамент, крыша и помещения, а не с мебелью и техникой, которые можно перемещать или заменять.

Исследования когнитивных архитектур направлены на решение задач моделирования ИИ на системном уровне, а не на уровне решения специализированных задач. Первоначально когнитивные архитектуры были ориентированы на обработку информации посредством моделирования таких функций, как рассуждение, планирование, изучение и решение проблем. В последние годы функционал ряда когнитивных архитектур был расширен и распространён на моделирование восприятия, действий, аффективных состояний (мотивацию, эмоции, отношения), взаимодействия с другими интеллектуальными агентами и окружающей средой. Для поддержания когнитивных архитектур строго необходимы только несколько функций – таких как распознавание и принятие решений. Однако, требуется полный набор функциональных возможностей, чтобы КА могла охватить полный спектр интеллектуальной деятельности человеческих уровня [3, p. 145].

К настоящему времени специалистами разработан и реализован ряд когнитивных архитектур: ACT-R, AIS,ATLANTIS, APEX, Soar, PRODIGY, Emile,EPIC, CIRCA,LIDA, ICARUS, PreAct, FORR, CLARION и т.д. К основным из них относятся ACT-R и SOAR.

Когнитивная архитектура ACT-R (Adaptive Control of Thought - Rational), разработанная в Университете Карнеги-Меллона Джоном Р. Андерсоном и Кристианом Лебьером, связана с моделированием человеческого мышления и поведения. ACT-R состоит из набора модулей, обрабатывающих различные типы информации. К ним относятся модули долгосрочной и кратковременной памяти, сенсорные модули для обработки визуальных данных, двигательные модули для действий, модуль намерений для целей и декларативный модуль для долгосрочных декларативных знаний. При выполнении задач модули восприятия, внимания, памяти и действий синхронизируются и функционируют как единое целое, в идеале порождая и поддерживая синергетические эффекты (хотя «заявления о синергии в когнитивных системах трудно проверить эмпирически» [3, p. 153]). В основе работы ACT-R лежат правила принятия решений и выполнения действий в виде условных инструкций: если выполнено определённое условие, то должно выполняться и соответствующее действие.

SOAR (State, Operator, AndResult – текущее состояние, оператор и результат) – одна из наиболее продуктивных когнитивных архитектур, имитирующих основные аспекты человеческого интеллекта.В основе SOAR лежит реализация единой теории познания Аллена Ньюэлла [4]. По мысли разработчиков SOAR Дж. Лэрда, А. Ньюэлла и П. Розенблума когнитивная архитектура интегрированного многозадачного интеллекта должна быть способна выполнять полный комплекс когнитивных функций, включая семантическую память, воображение, эмоции.

Ключевыми функционалами когнитивной архитектуры SOAR являются:

– хранящийся в долговременной памяти и активируемый в зависимости от ситуации набор правил представления информации, указывающих условия, при которых следует предпринимать те или иные действия (правила вида «если – то»);

– динамическая память, сохраняющая цели, текущее состояние, промежуточные результаты рассуждений, и постоянно обновляющаяся по мере поступления и обработки новой информации;

– механизм обучения с подкреплением, на основе которого КА создаёт новые алгоритмы и правила;

– методы теоретических рассуждений о принятии решений в динамических средах, в условиях неопределённости и неполноты данных.

Всё многообразие когнитивных архитектур сильного или общего искусственного интеллекта (AGI) группируется вокруг двух парадигм его построения: восходящей и нисходящей. Восходящая решает задачу моделирования базовых элементов и функций биологического субстрата человеческого мышления – мозга и центральной нервной системы человека, посредством чего достигается цель создания интеллекта как эмерджентного свойства этого субстрата: «Представляется логичным при создании будущего сильного ИИ “учиться у мозга”, поскольку никакой другой системы, обладающей сильным интеллектом, мы не знаем» [5]. Сторонники такого рода заключений не принимают во внимание того принципиально значимого факта, что мыслит не мозг, а человек при помощи мозга. Нисходящая парадигма стремится к моделированию высших когнитивных функций, ставя себе целью создание интеллекта perse [6, с. 25]. При этом построение такого рода когнитивных архитектур, как правило, ведётся на основе алгоритмов работы с символами, которые масштабируются «до уровня автономного поведения в реальном мире» [7]. В данной парадигме совершается инверсия причинно-следственной зависимости: упускается из виду тот факт, что формирование символов и их смыслов происходит в процессах интерсубъективных взаимодействий в реальном, «жизненном мире» (Lebenswelt). Между тем как «Сфера чистого интеллигибельного разума с его априоризмом и аподиктичностью есть, говоря языком науки, сфера потенциалистской логики, аксиоматики и дедукции, в конечном счёте, беспредметного, математического, количественно-топологического моделирования мира» [8, с. 204].

Таким образом, острие проблемы построения когнитивной архитектуры сильного ИИ заключается в разработке феноменологии присущих «естественному» интеллекту когнитивных функций, формирующихся в пространстве интерсубъективности.

ISSN 1999-5911 (Print)
ISSN 2960-2173 (Online)

Феноменологическая экспликация понятия интерсубъективности

Согласно Э. Гуссерлю, феноменологическая редукция («феноменологическое эпохе») заключается в систематическом исключении («заключении в скобки») всякой «объективирующей» позиции. Поле феноменологии как формы трансцендентальной философии конституируется тем обстоятельством, что в любом акте психического переживания человеку дан не мир «в себе и для себя», но мир, поскольку он возникает в сознании в качестве мира «субъекта». В трансцендентальной установке природа, общество и вообще вся вселенная («весь мир вещей, живых существ, людей, включая и нас самих» [8, с. 154] заключаются в скобки. В осадок выпадает чистое сознание, его имманентные взаимосвязи, интенциональные акты (ноэзис) и их корреляты (ноэмы), не затрагиваемые феноменологическим выключением.

Выявление на основе феноменологических процедур интенциональной структуры трансцендентальной интерсубъективности как «конкретной первоосновы» смысла любой реальности, трансцендируемой сознанием, и определение специфики её феноменологического поля становится одной из основных задач трансцендентально-эйдетической феноменологии как, по определению Э. Гуссерля, намеченной ещё Лейбницем универсальной онтологии, априорной науки об универсуме всевозможных форм существования. «Само по себе первое бытие, предшествующее всякой объективности мира и несущее её на себе, есть трансцендентальная интерсубъективность, вселенная монад, объединяющихся в различные сообщества» [10, с. 291].

Интерсубъективность как открытое сообщество монад конституирована во мне, в интенциональных структурах моего Эго, и вместе с тем, как такое сообщество, которое, будучи конституировано и в каждой другой монаде, в её субъективном модусе несёт в себе тот же самый объективный мир.

Согласно Э. Гуссерлю, феноменологический анализ рассматривает априорные сущностные формы и инвариантные структуры интерсубъективности в модусе данности трансцендентального Ego самому себе: интерсубъективная феноменология фундирована в солипсистски ограниченной эгологии. Поскольку «другой» в феноменологическом смысле есть «интенциональная модификация моего объективированного Я, моего первопорядкового мира» [10, с. 224], то, соответственно, бытие других для меня создаётся особой операцией сознания – аппрезентацией или аналогизирующей (уподобляющей) апперцепцией. Проще говоря, другой есть аналог я, или alterego. Alterego в эйдетико-трансцендентальном синтаксисе феноменологии Э. Гуссерля – это не «второе я» в смысле дублирования моей эмпирической идентичности, но вторая эго-структура, почему он и является другим, а не вещью. Ж. Деррида в полемике с Э. Левинасом, отказывавшемся называть другого alterego, поясняет: «Другой является абсолютно другим лишь будучи эго, то есть, в некотором смысле, будучи тем же, что я» [11, с. 193]. Поэтому конститутивные анализы другого как аналога или модификации моего первопорядкового Я не вносят каких-либо принципиальных конститутивных определений в само это Я, в трансцендентальное Эго.

Решение проблемы интерсубъективности, данное Гегелем в «Феноменологии духа», считает Ж.П. Сартр, представляет «значительный прогресс по сравнению

с Гуссерлем» [12, с. 259]. Э. Гуссерль принимает интенциональные структуры как априорно предпосланные эйдетически-феноменологическому анализу имманентные взаимосвязи поля трансцендентально чистого сознания. Для Гегеля структуру формообразований сознания (систему духовных феноменов) образует развитие совокупной духовной культуры человечества, в процессе которого возникают и изменяются как сознание, так и его предметы. При редукции сознания к трансцендентальному. Эго возможность такого подхода утрачивается.

Для Гегеля первичным, фундаментальным фактом опыта самосознания является не cogito, но опыт соотношения бытия-для-себя и бытия-для-иного. В самом своём бытии я зависим от другого, моё сознание в самом его средоточии должно быть опосредствовано другим сознанием, а бытие-для-другого становится необходимым условием для-себя-бытия. «Самосознание есть в себе и для себя потому и благодаря тому, что оно есть в себе и для себя для некоторого другого [самосознания], т.е. оно есть только как нечто признанное» [13, с. 99]. Возникающее в самосознании понятие Духа есть единство свободных и для себя сущих самосознаний: «я», которое есть «мы», и «мы», которое есть «я».

Ряд направлений современной философии принимает сторону Гегеля в вопросе о феноменологической экспликации понятия интерсубъективности. Согласно А. Рено, классическая концепция субъективности достигает высшей точки в монадологическом определении Я, принятым феноменологий Э. Гуссерля. Новый тип осмысления заключается в инверсии этого убеждения, т.е. в положении о том, что конституирование субъекта обусловлено интерсубъективностью: «Постмонадологическое восстановление субъекта происходит через новое открытие интерсубъективности как первичного условия субъективности» [14, с. 314].

Интерсубъективность существует не внутри трансцендентального эго, но как фактичность жизненного мира, «как фундаментальная онтологическая категория человеческого существования» [15, с. 82]. Задачей феноменологического анализа является не конструирование онтологии в терминах интенциональных актов трансцендентального эго, но раскрытие смысловой структуры обыденной интерсубъективности как данности жизненного мира в естественной установке.

Когнитивная архитектура интерсубъективности в проекте RoboErgoSum

«Современные когнитивные архитектуры и традиционные подходы к ИИ практически игнорируют решение субъектных проблем ОИИ, или оставляют их функционально нераскрытыми» [1, с. 190]. Однако из этого общего правила есть исключения. Одним из них является интернациональный проект создания системы искусственного интеллекта RoboErgoSum [16].

В традиционном для робототехники подходе восприятие рассматривается как изолированный процесс наблюдения. Авторы проекта RoboErgoSum объединяют сенсомоторную репрезентацию объекта и возможные действия агента с ним. Взаимодействуя с окружающей средой, робот определяет зависимости между объектами, потенциальными действиями и эффектами (изменениями окружающей среды и самого агента, индуцированные его действиями).

ISSN 1999-5911 (Print)
ISSN 2960-2173 (Online)

Объективные качества предметов выявляются через возможные действия с этими предметами. В этом же процессе возникает восприятие себя как отличного от окружающей среды – что является основой человеческого самосознания и его роботизированной модели. Авторы настаивают на том, что в их проекте роботы приобретают сознание и самосознание, формируют способности к построению внутренних моделей мира, к рассуждению и размышлению, самоконтролю, целеполаганию и мотивации, достигают понимания происходящего и т.д. Со своей стороны, мы полагаем, что такого рода антропоморфизмы допустимы лишь в качестве метафорического языка дискурса. На деле речь должна идти именно о моделировании высших психических функций человека в когнитивной архитектуре искусственного интеллекта. Решающее значение в этом моделировании отводится процессу обучения искусственного интеллекта на основе координации действий робота (и принятия им решений) с действиями человека. В данном обстоятельстве заключается принципиальное отличие проекта RoboErgoSum от концепций создания искусственных нейросетей генеративного ИИ на основе фреймовой семантики Марвина Минского и методов обучения с подкреплением так называемых «Больших языковых моделей» (LLM).

Для успешных совместных действий агенты взаимодействия должны уметь координировать свои намерения, планы, цели и действия: «Способность координировать различные стратегии принятия решений и обучения с подкреплением (рассматриваемая как основной адаптационный процесс принятия решений) может стать первым шагом к (i) большей автономии и адаптации робота, а также к (ii) способности робота анализировать эффективность своих процессов принятия решений и использовать этот анализ для изменения не только своего поведения, но и способа, которым он формирует своё поведение» [16].

У людей существует множество способов налаживания межличностной координации действий, от автоматических и непреднамеренных до чрезвычайно сложных, рефлексивно опосредованных форм коммуникативных практик. Аналогичные проблемы координации решаются и в процессах совместных действий человека и робота.Робот должен обладать способностью к созданию представлений о собственных мотивах и действиях, а также о намерениях и ментальных состояниях человека, с которым он взаимодействует. При этом он должен уметь делать выводы о том, как каждое из этих представлений развивается в ходе развёртывания совместного действия. «Робот также должен понимать и учитывать влияние своих собственных действий на психическое состояние своих партнёров» [16].

В целом когнитивная архитектура ИИ в проекте RoboErgoSum выстраивается как система взаимодействия комплекса модулей:

Модуль сенсорного восприятия, содержащий изначально встроенный набор перцептивных способностей для восприятия окружающей среды (визуальное восприятие и проприоцепция).

Двигательный модуль, содержащий набор доступных роботу действий, которые позволяют ему взаимодействовать с окружающей средой.

Модуль сенсомоторного обучения обрабатывает входные данные (обнаруженные объекты, выполненные действия, измеренные эффекты) для определения того, какие действия доступны роботу в данной ситуации.

Модуль пространственной ориентации генерирует и хранит символические данные о воспринимаемой среде. Затем эти данные используются на этапе планирования действий соответствующими модулями: модулем планирования задач с учётом потребностей человека и модулем планирования движений и манипуляций с учётом потребностей человека.

Система контроля взаимодействует с перечисленными модулями, чтобы решить, какую систему планирования действий использовать, как выполнять корректировку плана и отслеживать активность людей, с которыми взаимодействует робот.

Модуль мотивации управляет набором целей, которые должны быть достигнуты роботом.

Разумеется, здесь мы не можем входить в обсуждение инженерных проблем интеграции этих модулей, налаживания интерфейсов между ними и валидизации всей когнитивной архитектуры проекта RoboErgoSum. Для нас важен вывод разработчиков проекта о том, что эта архитектура выстраивается на основе координации действий человека и робота и реализуется как моделирование перцептивных, познавательных, аффективных, прогностических и т.д. структур такого рода интегрированной интерсубъективности.

Заключение

В заключение нашего обзора повторим ещё раз, что неумеренные антропоморфизмы, используемые разработчиками проекта RoboErgoSum в изложении идеи и принципов построения когнитивной архитектуры ИИ, не должны служить основанием для её априорной дискредитации. Отнюдь не обязательно разделять с исполнителями этого проекта убеждённость в том, что материал, представленный в их статье, «даёт представление о том, как создать самоосознающую (self-aware) систему» [16]. Вместе с тем, на наш взгляд, следует признать перспективность принципиальной позиции коллектива исполнителей проекта RoboErgoSum, согласно которой решающую роль в создании систем общего искусственного интеллекта должны сыграть разработки когнитивной архитектуры интерсубъективности как первоосновы формирования фундаментальной особенности человеческой жизнедеятельности: «Способности человека (как существа мыслящего) смотреть на самого себя как бы со “стороны”, как на нечто “другое”, как на особый предмет (объект), или, иными словами, превращать схемы своей собственной деятельности в объект её же самой» [17, с. 152].

Список литературы

1 Социогуманитарные аспекты цифровых трансформаций и искусственного интеллекта / Под ред. В.Е. Лепского, А.Н. Райкова. – Москва: Когито-Центр, 2022. – 308 с.

2 Cognitive Architecture // Institute for Creative Technologies. 2024. [Электронный ресурс]. – URL: https://cogarch.ict.usc.edu [дата доступа: 25.07.25].

3 Pat Langley, John E. Laird, Seth Rogers. Cognitive Architectures: Research Issues and Challenges // Cognitive Systems Research. 2009. No 10. P. 141–160.

4 Newell, Allen. Unified Theories of Cognition. – Harvard University Press, Cambridge, Massachusetts, 1990. 549 p.

ISSN 1999-5911 (Print)
ISSN 2960-2173 (Online)

5 Шумский С.А. Новые архитектуры сильного ИИ, основанные на принципах работы мозга [Электронный ресурс]. – URL: https://clck.ru/3NcMMt [дата доступа: 23.07.25].

6 Душкин Р.В. На пути к сильному искусственному интеллекту: когнитивная архитектура, основанная на психофизиологическом фундаменте и гибридных принципах // Программные системы и вычислительные методы. – 2021. № 1. С. 22-32. DOI: 10.7256/2454-0714.2021.1.34243

7 Краткий анализ существующих подходов к сильному ИИ. Часть 2. Когнитивные архитектуры [Электронный ресурс]. – URL: https://habr.com/ru/articles/145467/ [дата доступа: 02.08.25].

8 Кутырев В.А. Философия Чистого Разума Канта как спекулятивная предпосылка космизма, виртуальности и Искусственного Интеллекта (трансцендентальный дигитализм contra теллурический реализм) // Вопросы философии. 2023. № 2, С. 201–209. DOI: https://doi.org/10.21146/0042-8744-2023-2-201-209.

9 Гуссерль Э. Идеи к чистой феноменологии и феноменологической философии. Книга первая / Пер. с нем. А.В. Михайлова; Вступ. ст. В.А. Куренного. – Москва: Академический проект, 2009. – 489 с.

10 Гуссерль Э. Картезианские размышления. – Санкт-Петербург: Ювента, 1998. – 315 с.

11 Деррида Ж. Насилие и метафизика. Очерк мысли Эммануэля Левинаса // Деррида Ж. Письмо и различие. – Москва: Академический Проект, 2000. – С. 124–248.

12 Сартр Ж.П. Бытие и ничто: Опыт феноменологической онтологии / Пер. с фр., предисл., примеч. В.И. Колядко. – Москва: Республика, 2000. – 639 с.

13 Гегель Г.В.Ф. Феноменология духа // Гегель Г.В.Ф. Сочинения. Том IV. – Москва: Издательство социально-экономической литературы, 1959. – 440 с.

14 Рено А. Эра индивида. К истории субъективности / Перевод с французского С.Б. Рындина под редакцией Е.А. Самарской. – Санкт-Петербург: Владимир Даль, 2002. – 473 с.

15 Шютц А. Смысловая структура повседневного мира: очерки по феноменологической социологии / Сост. А.Я. Алхасов; Пер. с англ. А.Я. Алхасова, Н.Я. Мазлумяновой; Научи, ред. перевода Г.С. Батыгин. – Москва: Институт Фонда «Общественное мнение», 2003. – 336 с.

16 Toward Self-Aware Robots / Chatila R., Renaudo E., Andries M. [at al.]. // Frontiers in Robotics and AI. 2018. Vol. 5. Article 88. P. 1-20. doi: 10.3389/frobt.2018.00088.

17 Ильенков Э.В. Диалектическая логика: Очерки истории и теории. – 2-е изд., доп. – Москва: Политиздат, 1984. – 320 с.

Transliteration

1 Sotsiogumanitarnye aspekty tsifrovykh transformatsii i iskusstvennogo intellekta [Socio-humanitarian aspects of digital transformations and artificial intelligence] / Pod red. V.E. Lepskogo, A.N. Raikova. – Moskva: Kogito-Tsentr, 2022. – 308 s.

2 Cognitive Architecture. Institute for Creative Technologies. 2024. [Electronic resource]. – URL: https://cogarch.ict.usc.edu [date of access: 07/25/25].

3 Pat Langley, John E. Laird, Seth Rogers. Cognitive Architectures: Research Issues and Challenges. Cognitive Systems Research. 2009. No 10. P. 141–160.

4 Newell, Allen. Unified Theories of Cognition. Harvard University Press, Cambridge, Massachusetts, 1990. 549 p.

5 Husserl E. Idei k chistoi fenomenologii i fenomenologicheskoi filosofii. Kniga pervaya [Ideas for pure phenomenology and phenomenological Philosophy. The first book] / Per. s nem. A.V. Mikhailova; Vstup. st. V.A. Kurennogo. – Moskva: Akademicheskii proekt, 2009. – 489 s.

6 Husserl E. Kartezianskie razmyshleniya [Cartesian meditations]. – Sankt-Peterburg: Yuventa, 1998. 315 s.

7 Derrida J. Nasilie i metafizika. Ocherk mysli Emmanuelya Levinasa [Violence and metaphysics. An essay on the thought of Emmanuel Levinas] // Derrida Zh. Pis'mo i razlichie. – Moskva: Akademicheskii Proekt, 2000. – S. 124–248.

8 Shumskiy S.A. Novye arkhitektury sil'nogo II, osnovannye na printsipakh raboty mozga [New architectures of strong AI based on the principles of brain work] [Elektronnyi resurs]. – URL: https://clck.ru/3NcMMt [date of access: 07/23/25].

9 Dushkin R.V. Na puti k sil'nomu iskusstvennomu intellektu: kognitivnaya arkhitektura, osnovannaya na psikhofiziologicheskom fundamente i gibridnykh printsipakh [Towards strong artificial intelligence: a cognitive architecture based on a psychophysiological foundation and hybrid principles] // Programmnye sistemy i vychislitel'nye metody. – 2021. № 1. S. 22–32. DOI: 10.7256/2454-0714.2021.1.34243

10 Kratkii analiz sushchestvuyushchikh podkhodov k sil'nomu II. Chast' 2. Kognitivnye arkhitektury [A brief analysis of existing approaches to strong AI. Part 2. Cognitive architectures] [Electronic resource]. – URL: https://habr.com/ru/articles/145467/ [date of access: 08/02/25].

11 Kutyrev V.A. Filosofiya Chistogo Razuma Kanta kak spekulyativnaya predposylka kosmizma, virtual'nosti i Iskusstvennogo Intellekta (transtsendental'nyi digitalizm contra telluricheskii realizm) [Kant's Philosophy of Pure Reason as a speculative premise of cosmism, virtuality and Artificial Intelligence (transcendental digitalism contra telluric realism)] // Voprosy filosofii. 2023. № 2, S. 201–209. DOI: https://doi.org/10.21146/0042-8744-2023-2-201-209.

12 Sartre J.P. Bytie i nichto: Opyt fenomenologicheskoi ontologii [Being and nothing: The Experience of phenomenological ontology] / Per. s fr., predisl., primech. V.I. Kolyadko. – Moskva: Respublika, 2000. – 639 s.

13 Hegel G.V.F. Fenomenologiya dukha [Phenomenology of the Spirit] // Hegel G.V.F. Sochineniya. Tom IV. – Moskva: Izdatel'stvo sotsial'no-ekonomicheskoi literatury, 1959. – 440 s.

14 Renault A. Era individa. K istorii sub"ektivnosti [The era of the individual. Towards the History of Subjectivity] / Perevod s frantsuzskogo S.B. Ryndina pod redaktsiei E.A. Samarskoi. – Sankt-Peterburg: Vladimir Dal', 2002. – 473 s.

15 Schutz A. Smyslovaya struktura povsednevnogo mira: ocherki po fenomenologicheskoi sotsiologii [The semantic structure of the everyday world: essays on phenomenological sociology] / Sost. A.Ya. Alkhasov; Per. s angl. A.Ya. Alkhasova, N.Ya. Mazlumyanovoi; Nauchi, red. perevoda G.S. Batygin. – Moskva: Institut Fonda «Obshchestvennoe mnenie», 2003. – 336 s.

16 Towards Self-Aware Robots / Chatila R., Renaudo E., Andries M. [at al.]. Frontiers in Robotics and AI. 2018. Vol. 5. Article 88. P. 1-20. doi: 10.3389/frobt.2018.00088.

17 Ilyenkov E.V. Dialekticheskaya logika: Ocherki istorii i teorii. [Dialectical logic: Essays on history and theories]. – 2-e izd., dop. – Moskva: Politizdat, 1984. – 320 s.

Дунаев В.Ю., Курганская В.Д., Сағиқызы А.

Субъективтіліктің феноменологиясы және жалпы жасанды интеллекттің когнитивті архитектурасы

Аңдатпа. Жасанды интеллект жүйелерін құру үшін қажетті теориялық және әдіснамалық ресурстардың бірі субъективті шындықтың қасақана құрылымдарын феноменологиялық зерттеу болып табылады. Мақалада субъективтіліктің классикалық феноменологиясының принциптерінің қысқаша мазмұны келтірілген және оларды жалпы жасанды интеллекттің когнитивті архитектурасын құру жобаларының бірінде қолдануға талдау жасалады. Э. Гуссерльдің феноменологиясында трансцендентальды субъективтіліктің қасақана құрылымдары шындықтың онтологиялық және семантикалық

ISSN 1999-5911 (Print)
ISSN 2960-2173 (Online)

негізі мәртебесіне ие. Алайда, Э. Гуссерль тұжырымдамасында бұл құрылымдар трансцендентальды таза сана өрісінің имманентті байланыстары ретінде түсіндіріледі. Гегель үшін сана формаларының құрылымы адамзаттың жиынтық рухани мәдениетінің дамуын құрайды. Өзін-өзі тану тәжірибесінің негізгі, негізгі фактісі-бұл cogito емес, бірақ болмыстың-өзі үшін және болмыстың-басқа үшін-қатынасы. Менің болмысымда мен екіншісіне тәуелдімін, оның шоғырлануындағы менің санам басқа санамен делдал болуы керек.

Гегельдің субъективтілік феноменологиясына көзқарасы және оның сананы қалыптастырудағы рөлі roboergosum жасанды интеллект жүйесін құрудың халықаралық жобасында субъективтіліктің когнитивті архитектурасын құру тұжырымдамасының философиялық негіздемесі ретінде қарастырылуы мүмкін. Бұл жобада объектілердің объективті қасиеттері осы объектілермен мүмкін болатын әрекеттер арқылы анықталады. Табысты бірлескен іс-қимыл үшін өзара әрекеттесу агенттері өздерінің ниеттерін, жоспарларын, мақсаттары мен әрекеттерін үйлестіре білуі керек. Дәл осы процесте өзін қоршаған ортадан өзгеше деп қабылдау пайда болады-бұл адамның өзін-өзі тануының және оның роботтық моделінің негізі. Жобаны әзірлеушілер өздерінің жобаларында Роботтар сана мен өзін-өзі тануға ие болады, әлемнің ішкі модельдерін құру, ойлау және ойлау, өзін-өзі бақылау, мақсат қою және мотивация қабілеттерін қалыптастырады, не болып жатқанын түсінуге қол жеткізеді және т.б. мақалада антропоморфизмнің бұл түрі тек дискурстың метафоралық тілі ретінде рұқсат етіледі деген қорытынды жасалады. Шындығында, біз жасанды интеллекттің когнитивті архитектурасында адамның жоғары психикалық функцияларын (перцептивті, танымдық, аффективті, болжамдық) модельдеу туралы айтуымыз керек. Интеграцияланған субъективтіліктің осы түрінің құрылымдарын қалыптастыруда шешуші рөл роботтың іс-әрекетін адамның іс-әрекетімен үйлестіру негізінде жасанды интеллектті оқыту процесіне беріледі.

Түйін сөздер: субъективтілік, феноменология, жасанды интеллект, когнитивті сәулет, сана, робот

Dunaev V.Yu., Kurganskaya V.D., Sagikyzy A.
Phenomenology of intersubjectivity and cognitive architecture of artificial intelligence

Abstract. Phenomenological studies of intentional structures of subjective reality are one of the necessary theoretical and methodological resources for building artificial intelligence systems. The article provides a summary of the principles of the classical phenomenology of intersubjectivity and analyzes their application in one of the projects for building a cognitive architecture of general artificial intelligence. In E. Husserl's phenomenology, the intentional structures of transcendental intersubjectivity are given the status of the ontological and semantic primary basis of reality. However, in E. Husserl's concept, these structures are interpreted as immanent interrelations of the field of transcendentally pure consciousness. For Hegel, the structure of the formations of consciousness is formed by the development of the total spiritual culture of mankind. The primary, fundamental fact of the experience of self-awareness is not the cogito, but the experience of being-for-oneself and being-for-another. In my very being, I am dependent on the other, my consciousness at its very center must be mediated by another consciousness.

Hegel's approach to the phenomenology of intersubjectivity and its role in the formation of consciousness can be considered as a philosophical justification for the concept of building a cognitive architecture of intersubjectivity in the international project of creating an artificial intelligence system RoboErgoSum. In this project, the objective qualities of objects are revealed through possible actions with these objects. For successful joint actions, interaction agents must be able to coordinate their intentions, plans, goals, and actions. In the same process,

there is a perception of oneself as different from the environment, which is the basis of human self–awareness and its robotic model. The project developers insist that in their project robots acquire consciousness and self-awareness, develop the ability to build internal models of the world, to reason and reflect, self-control, goal setting and motivation, achieve an understanding of what is happening, etc. The article concludes that such anthropomorphisms are acceptable only as a metaphorical language of discourse. In fact, we should be talking about modeling higher human mental functions (perceptual, cognitive, affective, predictive) in the cognitive architecture of artificial intelligence. A crucial role in the formation of structures of this kind of integrated intersubjectivity is assigned to the process of artificial intelligence training based on the coordination of robot actions with human actions.

Keywords: intersubjectivity, phenomenology, artificial intelligence, cognitive architecture, consciousness, robot

Поступила: 15.07.2025
Принята:01.09.2025