

УДК 004.89:130.3

## ФЕНОМЕНОЛОГИЯ ИНТЕРСУБЪЕКТИВНОСТИ И КОГНИТИВНАЯ АРХИТЕКТУРА ОБЩЕГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА \*

*В.Ю. Дунаев, В.Д. Курганская*

Институт философии, политологии и религиоведения Комитета науки  
Министерства науки и высшего образования Республики Казахстан,  
г. Алматы, Республика Казахстан

Поступила в редакцию: 07.09.25

В окончательном варианте: 05.10.25

**Аннотация.** Одним из необходимых теоретико-методологических ресурсов для построения систем искусственного интеллекта являются феноменологические исследования интенциональных структур субъективной реальности. В статье даётся краткое описание принципов классической феноменологии интерсубъективности и проводится анализ их применения в одном из проектов построения когнитивной архитектуры общего искусственного интеллекта. В феноменологии Э. Гуссерля интенциональные 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 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.

\* Статья подготовлена в рамках финансирования КН МНВО РК (ИРН BR21882302 «Казахстанский социум в условиях цифровой трансформации: перспективы и риски»).

## PHENOMENOLOGY OF INTERSUBJECTIVITY AND COGNITIVE ARCHITECTURE OF ARTIFICIAL INTELLIGENCE

*V.Yu. Dunaev, V.D. Kurganskaya*

Institute for Philosophy, Political Science and Religion Studies of the Science  
Committee of the Ministry of Science and Higher Education of the Republic  
of Kazakhstan,  
Almaty, Republic of Kazakhstan

Original article submitted: 07.09.25

Revision submitted: 05.10.25

---

**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 transcendently 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-onself 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.

---

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 философии в моделировании интеллектуальных функций является трансформация методологии феноменологических анализов трансцендентальных структур интересубъективности в технологию конструирования когнитивной архитектуры общего или сильного искусственного интеллекта.

### **Феноменологическая экспликация понятия интересубъективности**

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» [2, 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 интересубъективности 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 интересубъективность, вселенная монад, объединяющихся в различные сообщества» [3, 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 интересубъективности в модусе данности трансцендентального Его самому себе: интересубъективная феноменология фундирована в солипсистски ограниченной эгологии. Поскольку «другой» в феноменологическом смысле есть «*интенциональная модификация моего объективированного Я, моего первопорядкового мира*» [3, с. 224], то, соответственно, бытие других для меня *создаётся* особой операцией сознания – аппрезентацией или аналогизирующей (уподобляющей) апперцепцией. Проще

говоря, другой есть аналог я, или alter ego. Alter ego 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, отказывавшимся называть другого alter ego, поясняет: «Другой является абсолютно другим, лишь будучи эго, то есть в некотором смысле будучи тем же, что я» [4, 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» [5, с. 259]. 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» [6, 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, принятом феноменологией 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» [7, с. 314].

Интерсубъективность существует не внутри трансцендентального эго, но как фактичность жизненного мира, «как фундаментальная онтологическая категория человеческого существования» [8, 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 [9].

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). 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. 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.

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» [9].

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» [9].

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 участвовать в обсуждении инженерных проблем интеграции этих модулей, налаживания интерфейсов между ними и валидации 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 других структур такого рода интегрированной интерсубъективности. Повторим ещё раз, что неумеренные антропоморфизмы, используемые авторами в изложении идеи и принципов построения когнитивной архитектуры ИИ, не должны служить основанием для её априорной дискредитации. Отнюдь не обязательно разделять с авторами убеждённость в том, что материал, представленный в их статье, «даёт представление о том, как создать самоосознающую (self-aware) систему» [9]. Вместе с тем, на наш взгляд, следует признать перспективность позиции 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» [10, 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. – ISBN 978-5-89353-665-2.
2. Гуссерль, Э. Идеи к чистой феноменологии и феноменологической философии. Книга первая / Э. Гуссерль; пер. с нем. А.В. Михайлова. – Москва: Академический проект, 2009. – 489 с. – ISBN 9768-5-8291-1042-0.
3. Гуссерль, Э. Картезианские размышления / Э. Гуссерль. – Санкт-Петербург: Ювента, 1998. – 315 с. – ISBN 5-87399-059-X.
4. Деррида, Ж. Насилие и метафизика. Очерк мысли Эммануэля Левинаса / Ж. Деррида // Письмо и различие. – Москва: Академический Проект, 2000. – С. 124–248. – ISBN 5-8291-0088-6.
5. Сартр, Ж.П. Бытие и ничто: Опыт феноменологической онтологии / Ж.П. Сартр; пер. с фр. В.И. Колядко. – Москва: Республика, 2000. – 639 с. – ISBN 5-250-02726-6.
6. Гегель, Г.В.Ф. Феноменология духа / Г.В.Ф. Гегель // Сочинения. Том IV. – Москва: Издательство социально-экономической литературы, 1959. – 440 с.
7. Рено, А. Эра индивида. К истории субъективности / А. Рено; пер. с фр. С.Б. Рындина; под ред. Е.А. Самарской. – Санкт-Петербург: Владимир Даль, 2002. – 473 с. – ISBN 5-93615-019-4.
8. Шютц, А. Смысловая структура повседневного мира: очерки по феноменологической социологии / Сост. А.Я. Алхасов; пер. с англ. А.Я. Алхасова, Н.Я. Мазлумяновой. – Москва: Институт Фонда «Общественное мнение», 2003. – 336 с. – ISBN 5-93947-012-2.
9. Toward Self-Aware Robots / R. Chatila, E. Renaudo, M. Andries [et al.] // *Frontiers in Robotics and AI*. – 2018. – Vol. 5. – Article 88. – Pp. 1–20. – DOI: 10.3389/frobt.2018.00088.
10. Ильенков, Э.В. Диалектическая логика: Очерки истории и теории / Э.В. Ильенков. – 2-е изд., доп. – Москва: Политиздат, 1984. – 320 с.

## References

1. Socio-humanitarian aspects of digital transformations and artificial intelligence. Ed. by V.E. Lepskiy, A.N. Raikov. Moscow: Kogito-Center, 2022. 308 p. ISBN 978-5-89353-665-2 (In Russ.).
2. Husserl E. Ideas for pure phenomenology and phenomenological Philosophy. The first book. Transl. from German by A.V. Mikhailov. Moscow: Academic Project, 2009. 489 p. ISBN 9768-5-8291-1042-0 (In Russ.).
3. Husserl E. Cartesian reflections. St. Petersburg: Juventus, 1998. 315 p. ISBN 5-87399-059-X (In Russ.).
4. Derrida J. Violence and metaphysics. An essay on the thought of Emmanuel Levinas. *Writing and difference*. Moscow: Akademicheskii Proekt, 2000. Pp. 124–248. ISBN 5-8291-0088-6 (In Russ.).
5. Sartre J.P. Being and nothing: The Experience of phenomenological ontology / Transl. from French, preface, note by V.I. Kolyadko. Moscow: Respublika, 2000. 639 p. ISBN 5-250-02726-6. (In Russ.).
6. Hegel G.V.F. Phenomenology of the Spirit. Essays. Vol. IV. Moscow: Publishing House of Socio-economic Literature, 1959. 440 p. (In Russ.).
7. Renault A. The era of the individual. Towards the History of Subjectivity. Transl. from French by S.B. Ryndin, ed. by E.A. Samarskaya. St. Petersburg: Vladimir Dahl, 2002. 473 p. ISBN 5-93615-019-4 (In Russ.).

8. Schutz A. The semantic structure of the everyday world: essays on phenomenological sociology. Transl. from English by A.Ya. Alkhasova, N.Ya. Mazlumyanova. Moscow: Institute of the Public Opinion Foundation, 2003. 336 p. ISBN 5-93947-012-2 (In Russ.).
9. Chatila R., Renaudo E., Andries M. [et al.]. Towards Self-Aware Robots. *Frontiers in Robotics and AI*. 2018;5:1-20. DOI: 10.3389/frobt.2018.00088.
10. Ilyenkov E.V. Dialectical logic: Essays on history and theories. 2nd ed. Moscow: Politizdat, 1984. 320 p. (In Russ.).

---

*Информация об авторах*

---

**ДУНАЕВ Владимир Юрьевич** – доктор философских наук, профессор, главный научный сотрудник Центра политических исследований Института философии, политологии и религиоведения Комитета науки Министерства науки и высшего образования Республики Казахстан, г. Алматы, Республика Казахстан; eLibrary SPIN: 3081-7406.  
**E-mail:** vlad.dunaev2011@yandex.kz

**КУРГАНСКАЯ Валентина Дмитриевна** – доктор философских наук, профессор, главный научный сотрудник Центра политических исследований Института философии, политологии и религиоведения Комитета науки Министерства науки и высшего образования Республики Казахстан, г. Алматы, Республика Казахстан; eLibrary SPIN: 5551-0409.  
**E-mail:** vkurganskaya@mail.ru

---

*Information about the authors*

---

**DUNAEV Vladimir Yu.** – Doctor of Philosophy, Professor, Chief Researcher at the Center for Political Studies of the Institute for Philosophy, Political Science and Religious Studies of the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Almaty, Republic of Kazakhstan; eLibrary SPIN: 3081-7406. **E-mail:** vlad.dunaev2011@yandex.kz

**KURGANSKAYA Valentina D.** – Doctor of Philosophy, Professor, Chief Researcher at the Center for Political Studies of the Institute for Philosophy, Political Science and Religious Studies of the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Almaty, Republic of Kazakhstan; eLibrary SPIN: 5551-0409. **E-mail:** vkurganskaya@mail.ru