

*Д.И. Дубровский*

## **Развитие искусственного интеллекта и глобальный кризис земной цивилизации (к анализу социогуманитарных проблем)\***

*Дубровский Давид Израилевич* – доктор философских наук, главный научный сотрудник. Институт философии РАН. Российская Федерация, 109240, г. Москва, ул. Гончарная, д. 12, стр. 1; e-mail: ddi29@mail.ru

В статье рассматривается качественно новый этап развития искусственного интеллекта (ИИ), связанный с разработкой общего искусственного интеллекта (в международной номенклатуре сокращенно обозначается **AGI – Artificial General Intelligence**). В отличие от традиционного ИИ, **AGI** существенно приближается по своим функциям к естественному интеллекту (ЕИ), сможет обладать способностью самообучаться, решать широкий круг задач в разных средах, т.е. быть интегральным и автономным. Такой уровень «самостоятельности» **AGI** открывает принципиально новые перспективы для развития информационных технологий, но вместе с тем ставит многие острые социогуманитарные проблемы, связанные с рисками и угрозами утраты контроля за развитием ИИ. Для успешной разработки **AGI** необходимы новые теоретико-методологические подходы, опирающиеся на принципы постнеклассической эпистемологии и результаты нейронаучных и феноменологических исследований сознания. Весьма важно рассмотрение этих вопросов под углом крайнего обострения глобального кризиса мировой цивилизации, обусловленного ее потребительской доминантой и усилиями сохранить ее монополярную структуру со стороны США и их западных союзников. В этой связи требуется также более широкий, философско-антропологический подход к пониманию нынешнего состояния нашей цивилизации и возможностей ее преобразования. Он предполагает учет того, что именуют природой человека, как стойкого комплекса его психических и телесных свойств. Они воспроизводились у всех

---

\* Работа выполнена при поддержке РФФ, проект № 21-18-00184 «Социогуманитарные основания критериев оценки инноваций, использующих цифровые технологии и искусственный интеллект».

народов, во все исторические эпохи, при всех социальных устройствах, что указывает на их биологическую обусловленность. Среди них, наряду с альтруистическими свойствами, можно выделить ряд негативных свойств (такие как неограниченное потребление, агрессивность к себе подобным, чрезмерное эгоистической своеволие). Именно эти свойства, характерные для массового сознания, активно эксплуатировали в своих интересах адепты монополярности. Преодоление принципов и, главное, практики монополярности и тем самым изменение мировой социальной самоорганизации является необходимым условием для подлинно гуманистического этапа антропотехнологической эволюции, способной открыть новые экзистенциальные перспективы преобразования человека и человечества.

**Ключевые слова:** искусственный интеллект (ИИ), естественный интеллект (ЕИ), общий искусственный интеллект (AGI), его интегральность и автономность, социогуманитарные оценки AGI (позитивные и негативные), глобальный кризис земной потребительской цивилизации, природа человека, антропотехнологическая эволюция и ее перспективы

Последние два десятилетия отмечены крупными достижениями в развитии информационных технологий и процессов цифровизации, охвативших практически все основные сферы социальной жизнедеятельности. Это служит предметом широкого обсуждения в философской и научной литературе под углом рассмотрения их социогуманитарной значимости, в том числе анализа явлений негативного характера. Опасения общественности вызывают риски и угрозы, обусловленные исключительно быстрым возрастанием сложности систем искусственного интеллекта (ИИ), трудностями контроля за их функционированием. Это в первую очередь связано с результатами использования машинного обучения, в особенности глубокого обучения нервных сетей, количество параметров которых исчисляется триллионами. Такой масштаб усложнения ведет к тому, что специалисты называют проблемой «Черного ящика», т.е. недостаточной обоснованности, «объяснимости» получаемых на выходе результатов работы с большими данными [Славин, 2021]. А это снижает к ним доверие пользователей: ведь именно эти результаты должны служить в качестве инструментов и рекомендаций для решения всевозможных социальных задач (экономических, образовательных, коммуникативных, медицинских и т.д.). Отмечаются к тому же многие существенные издержки внедрения цифровизации в разных сферах социальной жизни. Как видим, крупные достижения ИИ порождают и столь же крупные проблемы, касающиеся вопросов социальной значимости этих достижений.

Но еще гораздо более серьезные проблемы такого рода возникают сейчас, когда начался качественно новый этап развития ИИ, связанный с разработкой **Общего искусственного интеллекта** (в принятой международной номенклатуре сокращенно – AGI, от Artificial General Intelligence), который по своим функциям приближается к способностям естественного интеллекта (ЕИ). Все существующие, традиционные системы ИИ являются «узкими», т.е. функционируют лишь в ограниченном наборе задач и условий, представляют отдельные виды ИИ, такие как системы компьютерного зрения, обработки естественного языка, машинного обучения, обработки символьной информации и т.п.

Они не способны самообучаться в существенно новых условиях, требуют каждый раз специального программирования. Они не являются автономными, не могут полноценно функционировать без участия человека-оператора.

Что же касается **AGI**, то эта система ИИ по своим функциональным возможностям должна быть не «узкой», а *интегральной* (в смысле *общей*), но в то же время *автономной* (функционирующей независимо от человека-оператора) и *самообучающейся* (способной самостоятельно решать задачи в разных условиях внешней среды). Разработка **AGI** стала мировым трендом, предметом возрастающей конкуренции между крупнейшими корпорациями и государствами-лидерами в области ИИ.

Недавно вышла первая в нашей стране книга, посвященная основательному обзору проблематики **AGI**, анализу состояния ее разработки, основных трудностей и возможных способов их преодоления. В связи с этим важное место в книге занимают вопросы уточнения понятий ИИ и ЕИ, особенно соотношение **AGI** с ЕИ и, что весьма интересно, вопросы о *proto-AGI* и *узком AGI*, которые рассматриваются как стадии продвижения к полноценному **AGI** [Бурцев, Бухвалов, Ведяхин, 2021].

Для успешной разработки **AGI** требуются новые теоретические и методологические подходы, отвечающие принципам постнеклассической эпистемологии [Лекторский, 2001], с учетом опыта философского анализа междисциплинарных проблем и результатов когнитивной науки, включая социогуманитарный контекст [Лекторский, 2006; Лекторский, 2018]. Вместе с этим для моделирования необходимых **AGI** функций требуется исследование и описание специфических когнитивных архитектур естественных процессов мышления. Здесь необходимыми ресурсами для построения моделей **AGI** могут служить в первую очередь положения эпистемологии, касающиеся объяснения и обоснования успешного познавательного процесса, результаты психологических и феноменологических исследований ценностно-смысловых, динамических структур субъективной реальности и результаты новейших нейронаучных исследований сознания. Одна из попыток анализа, обобщения и систематизации этих исследований содержится в моих работах (см.: [Дубровский, 2021; Дубровский, 2022]).

Принципиальное значение для осмысления нового этапа развития ИИ имеют основательные социогуманитарные подходы, на основе которых оцениваются проблемы, результаты и перспективы разработки **AGI**. В этом плане важное значение приобретает предложенная В.Е. Лепским концепция саморазвивающихся рефлексивно-активных сред (кибернетика третьего порядка) [Лепский, 1998; Лепский, 2010] и рассмотрение на ее основе вопросов социогуманитарной эргономики, особенно социогуманитарных критериев оценки новаций цифровой реальности [Лепский, 2018; Лепский, 2021].

Очевидно, что успехи в разработке **AGI** будут иметь стратегическое значение для нашей страны. Это обязывает нас максимально концентрировать научные и организационные усилия, настойчиво добиваться опережения конкурентов. В нынешних условиях предельного обострения противоречий на международной арене это имеет чрезвычайное значение.

Наряду с акцентом на достижении успехов в разработке **AGI**, столь же важно рассматривать те новые широкомасштабные риски и угрозы социогума-

нитарного характера, которое они могут принести человечеству. Ведь система ИИ такого рода должна стать автономной и самообучающейся, а следовательно, способной к «самодетальности»; она сможет легко выходить из-под контроля человека, более того – генерировать собственные цели и даже конкурировать с ним. Подобные сюжеты обсуждаются во многих публикациях. Некоторые авторы (как, например, Ник Бостром) говорят о сингулярном рубеже – когда наделенный интеллектом искусственный агент догонит, а затем и превзойдет человека, причем для этого ему не обязательно приписывать обладание сознанием, качеством субъективной реальности. Теоретически это мыслимо в силу принципа изофункционализма систем. Впрочем, в такой же степени мыслимо, что реально это не удастся осуществить. Тем не менее вполне реальны перспективы развития человека в форме гибридного интеллекта, сочетающего в себе свойства ЕИ с ИИ (уже сейчас у нас такое сочетание проявляется в определенной форме и степени). Можно с уверенностью полагать, что идущая уже несколько десятилетий и набирающая темпы **антропо-технологическая эволюция** способна привести к существенным, возможно, качественным преобразованиям нашего интеллекта, творческих способностей, нашей телесности и в силу этого основного комплекса наших потребностей, а тем самым к трансформации социума. Мы уже были свидетелями того, что буквально за несколько десятилетий в связи с появлением ИИ и развитых на его основе технологий произошел переход от индустриального к информационному обществу, радикально изменилась вся система производства, потребления и коммуникаций (достаточно сравнить 1950-й год с 1990-м, не говоря уже о последнем времени).

Подобная перспектива вполне вероятна в случае создания и развития технологий, связанных с **AGI**, особенно в состоянии резкого обострения глобального кризиса нашей цивилизации. Выход из этого состояния предполагает существенное преобразование мировой социальной самоорганизации, прежде всего конец ее монополярной структуры, породившей фундаментальные противоречия, чреватые третьей мировой войной.

Рассуждая о социогуманитарных аспектах нового этапа развития ИИ, важно расширить рассмотрение этой проблематики в **антропологическом контексте**. Нашу цивилизацию справедливо называют «потребительской», поскольку главный, неуклонный вектор ее развития состоит в том, чтобы еще больше потреблять, чтобы еще больше производить, чтобы еще больше потреблять и т.д. и т.п. Из этого параноидального круга не видно выхода. Главным «двигателем» и охранителем такого «прогресса» выступает пресловутый «золотой миллиард», западные страны во главе с США, которые потребляют неизмеримо больше остальных и навязывают свою систему ценностей и свои интересы всему миру. Но чтобы еще больше производить, нужно производить все новые и новые, более эффективные технологии. А для всего этого нужны новые и новые природные ресурсы, которые ограничены, и чтобы их обрести, необходимо политическое доминирование в мире, нужна власть над теми «нижестоящими» странами, которые ими располагают, и, конечно, нужно, чтобы эти страны тоже потребляли больше продукции, производимой для них западным миром. Везде – круговая зависимость экономических, политических

и иных интересов субъектов разного вида и ранга, начисто лишенных, как правило, подлинного гуманистического экзистенциального смысла. Но это ведет к нарастанию абсурда во всех сферах и на всех уровнях социальной жизнедеятельности, что представляет яркий симптом глобального кризиса нашей цивилизации. Сейчас этот кризис, развивавшийся много лет, приближается к пределу, за которым либо гибель нашей цивилизации, либо переход ее на новый этап развития. В этих условиях понятие социогуманитарной значимости технологий ИИ существенно изменяет свое содержание в плане системы приоритетов, главными из которых становятся жизнь и здоровье людей, выживание человечества. Вместе с этим повышается вероятность того, что развитие «узких» видов ИИ, AGI, разных форм гибридного интеллекта смогут, как уже отмечалось выше, качественно обогатить и ускорить антропотехнологическую эволюцию, представляющую собой форму и средство перехода к новому этапу жизни земной цивилизации.

Но здесь надо еще раз обратить внимание на то принципиальное обстоятельство, что информационные технологии играют в общественном развитии двоякую роль. С одной стороны, они являются средством эффективного решения неоглядного множества насущных практических задач и способны при определенных условиях служить глобальным целям положительного характера, но, с другой стороны, будучи интегральными в том смысле, что являются важным фактором совершенствования и использования большинства современных технологий, они определяют общее состояние и направленность развития производства, потребления, коммуникаций, в конечном итоге основные, в том числе негативные, особенности нынешнего этапа мировой цивилизации. Таким образом, выходит, что они ответственны за неуклонно нарастающий глобальный кризис.

Мы часто слышим, что дело не в самих по себе технологиях, а в том, как их использует человек. Во многих конкретных случаях это справедливо. Но возникает вопрос: почему же абсолютное большинство, многие сотни миллионов людей (независимо от национальной, государственной, религиозной принадлежности, от социального статуса, уровня благосостояния, образования и т.д.) охотно используют их в целях еще большего потребления и осуществления личных или институциональных интересов?

Получается, что ответственность лежит, конечно, на самом человечестве. Это давно и предельно ясно для многих, стало уже неким общим местом, но тем не менее требует осмысления с опорой на более глубокое антропологическое основание, именуемое **«природой человека»**. Это понятие весьма часто используется в философской и научной литературе. Оно обозначает комплекс устойчивых психических и телесных свойств социального индивида, которые обнаруживают инвариантность по отношению ко всем различным эпохам, этносам, социальным устройствам. А это однозначно указывает на их биологическую обусловленность. Взятые по отдельности и в комплексе, эти свойства выражают потребности, способности, склонности, коммуникативные особенности, формы поведения и деятельности человека. За обозримый период истории природа человека не изменилась. Так позволяют думать обширные материалы о людях Древнего Египта и Месопотамии, труды древнегреческих

и древнеримских историков и философов. Читая «Характеры» Теофраста, поражаешься удивительному сходству: прошло почти две с половиной тысячи лет, но перед нами – все те же столь знакомые каждому человеческие типы, формы общения и поведения. Поскольку природа человека биологически обусловлена, она не могла измениться за столь мизерный срок, по сравнению с процессом антропогенеза.

Когда мы размышляем о социогуманитарных проблемах, то часто недооцениваем социальную значимость биологических факторов, то обстоятельство, что наши фундаментальные ценности, такие как жизнь, семья, дети, любовь, здоровье и ряд других, имеют биологическую основу. Природа человека, наряду с альтруистическими свойствами, включает ряд явно негативных свойств, играющих весьма важную роль в социальных процессах. Если говорить кратко, то это – неумеренное потребительство, агрессивность к себе подобным и чрезмерное эгоистическое своеволие. Они в разной степени присущи как отдельным людям, так и коллективным и институциональным субъектам, выражают существенные черты массового сознания, которые определяют цели, формы и результаты деятельности массового человека. Но именно эта деятельность в итоге служит неуклонному нарастанию глобального кризиса земной цивилизации.

Указанные негативные свойства природы человека активно используются сложившейся системой потребительского общества в условиях монополярной структуры мировой цивилизации, которая давно обнаружила свой деструктивный характер. Именно монополярность выступает сейчас главным фактором роста международных конфликтов, мировых экономических дисбалансов и нарушений социальной самоорганизации, т.е. служит главным фактором углубления глобального кризиса нашей цивилизации. В последнее время мы являемся свидетелями если не краха монополярности, то, по крайней мере, резкого падения влияния, управляющей силы тех государств, которые олицетворяли и осуществляли много лет идеи и функции монополярности (имеются в виду, конечно, США и их западные союзники). Крах монополярного типа цивилизации станет началом перестройки мировой системы социальной самоорганизации с новыми возможностями антропотехнологической эволюции, способной открыть новые экзистенциально значимые перспективы преобразования человека и человечества.

### Список литературы

Бурцев, Бухвалов, Ведяхин, 2021 – *Бурцев М.С., Бухвалов О.Л., Ведяхин А.А. и др.* Сильный искусственный интеллект. На подступах к сверхразуму. М.: Интеллектуальная литература, 2021. 232 с.

Дубровский, 2007 – *Дубровский Д.И.* Природа человека и социальные процессы // *Дубровский Д.И.* Сознание, Мозг, Искусственный интеллект. М.: Стратегия-Центр, 2007. С. 249–264.

Дубровский, 2021 – *Дубровский Д.И.* Задача создания Общего искусственного интеллекта и проблема сознания // *Философские науки.* 2021. Т. 64. № 1. С. 13–44.

Дубровский, 2022 – *Дубровский Д.И.* Значение нейронаучных исследований сознания для разработки общего искусственного интеллекта (методологические вопросы) // *Вопросы философии.* 2022. № 2. С. 83–93.

Лекторский, 2001 – Лекторский В.А. Эпистемология классическая и неоклассическая. М.: Эдиториал УРСС, 2001. 256 с.

Лекторский, 2006 – Лекторский В.А. Философия, искусственный интеллект и когнитивная наука // Искусственный интеллект: Междисциплинарный подход / Под ред. Д.И. Дубровского, В.А. Лекторского. М.: ИИнтелЛ, 2006. С. 13–21.

Лекторский, 2018 – Лекторский В.А. Человек и культура. СПб.: СПбГУП, 2018. 640 с.

Лепский, 1998 – Лепский В.Е. Концепция субъектно-ориентированной компьютеризации управленческой деятельности. М.: Ин-т психологии РАН, 1998. 204 с.

Лепский, 2010 – Лепский В.Е. Рефлексивно-активные среды инновационного развития. М.: Когито-Центр, 2010. 280 с.

Лепский, 2021 – Лепский В.Е. Искусственный интеллект в субъектных парадигмах управления // Филос. науки. 2021. Т. 64. № 1. С. 88–101.

Лепский, 2018 – Лепский В.Е. Социогуманитарные критерии оценки новаций цифровой реальности // SocioTime / Социальное время. 2018. № 4 (16). С. 16–26.

Славин, 2021 – Славин Б. Может ли искусственный интеллект быть справедливым // ВИТ. 2021. № 10 (113). С. 32–35.

## **The development of artificial intelligence and the global crisis of earthly civilization (to the analysis of socio-humanitarian problems)\***

*David I. Dubrovsky*

Institute of Philosophy, Russian Academy of Sciences. 12/1 Gonchamnaya Str., Moscow 109240, Russian Federation; e-mail: ddi29@mail.ru

The article considers a qualitatively new stage in the development of artificial intelligence (AI), associated with the development of artificial general intelligence (abbreviated as AGI in the international nomenclature – from Artificial General Intelligence). Unlike traditional AI, AGI is significantly closer in its functions to natural intelligence (EI), it will be able to self-learn, solve a wide range of tasks in different environments, i.e. be integral and autonomous. Such a level of “independence” of AGI opens up fundamentally new prospects for the development of information technologies, but at the same time poses many acute socio-humanitarian problems associated with the risks and threats of losing control over the development of AI. The successful development of AGI requires new theoretical and methodological approaches based on the principles of post-nonclassical epistemology and the results of neuroscientific and phenomenological studies of consciousness. It is very important to consider these issues from the angle of the extreme aggravation of the global crisis of world civilization, due to its consumer dominance and efforts to preserve its monopolar structure from the part of the United States and its Western allies. In this regard, a broader, philosophical-anthropological approach is also required to understand the current state of our civilization and the possibilities for its transformation. It involves taking into account what is called the nature of man, as a stable complex of his mental and bodily properties. They were reproduced among all peoples, in all historical epochs, under all social structures, which indicates their biological conditionality. Among them, along with altruistic properties,

---

\* The work was supported by the Russian Science Foundation, project No. 21-18-00184 “Social and humanitarian foundations for evaluation criteria for innovations using digital technologies and artificial intelligence”.

a number of negative properties can be distinguished (such as unlimited consumerism, aggressiveness towards one's own kind, excessive egoistic self-will). These characteristic properties of mass consciousness were actively exploited adherents of monopolarity in their interests. Overcoming the principles and practices of monopolarity and thereby changing the global social self-organization is a necessary condition for a truly humanistic stage of anthropotechnological evolution, capable of opening up new existential prospects for the transformation of man and mankind.

**Keywords:** artificial intelligence (AI), natural intelligence (EI), artificial general intelligence (AGI), its integrality and autonomy, socio-humanitarian assessments of AGI (positive and negative), global crisis of earthly consumer civilization, human nature, anthropotechnological evolution and its prospects

## References

Burtsev, M.S., Bukhvalov, O.L., Vedyakhin, A.A. et al. *Sil'nyi iskusstvennyi intellekt. Na podstupakh k sverkhrazumu* [Strong artificial intelligence. On the approaches to the supermind]. Moscow: Intellectual Literature Publ., 2021. 232 pp. (In Russian)

Dubrovsky, D.I. "Priroda cheloveka i sotsial'nye protsessy" [Human nature and social processes], in: Dubrovsky D.I., *Soznanie, Mozg, Iskusstvennyi intellekt* [Consciousness, Brain, Artificial Intelligence]. Moscow: Strategiya-Tsentr Publ., 2007, pp. 249–264. (In Russian)

Dubrovsky, D.I. "Zadacha sozdaniya Obshchego iskusstvennogo intellekta i problema soznaniya" [The task of creating a general artificial intelligence and the problem of consciousness], *Filosofskie nauki*, 2021, vol. 64, no. 1, pp. 13–44. (In Russian)

Dubrovsky, D.I. "Znachenie neironauchnykh issledovaniy soznaniya dlya razrabotki obshchego iskusstvennogo intellekta (metodologicheskie voprosy)" [The Importance of Neuroscientific Research of Consciousness for the Development of General Artificial Intelligence (Methodological Issues)], *Voprosy Filosofii*, 2022, no. 2, pp. 83–93. (In Russian)

Lectorsky, V.A. *Ehpistemologiya klassicheskaya i neoklassicheskaya* [Classical and neoclassical epistemology]. Moscow: Editorial URSS Publ., 2001. 256 pp. (In Russian)

Lectorsky, V.A. "Filosofiya, iskusstvennyi intellekt i kognitivnaya nauka" [Philosophy, artificial intelligence and cognitive science], in: *Iskusstvennyi intellekt: Mezhdistsiplinarnyi podkhod* [Artificial intelligence: An interdisciplinary approach], ed. by D.I. Dubrovsky, V.A. Lectorsky. Moscow: IInteLL Publ., 2006, pp. 13–21. (In Russian)

Lectorsky, V.A. *Chelovek i kul'tura* [Man and Culture]. St. Petersburg: SPbUHSS Publ., 2018. 640 pp. (In Russian)

Lepsky, V.E. *Kontseptsiya sub"ektno-orientirovannoi komp'yuterizatsii upravlencheskoi deyatel'nosti* [The concept of subject-oriented computerization of managerial activity]. Moscow: Institute of Psychology of the Russian Academy of Sciences Publ., 1998. 204 pp. (In Russian)

Lepsky, V.E. *Refleksivno-aktivnye sredy innovatsionnogo razvitiya* [Reflexive-active environments of innovative development]. Moscow: Kogito-Center Publ., 2010. 280 pp. (In Russian)

Lepsky, V.E. "Iskusstvennyi intellekt v sub"ektnykh paradigmakh upravleniya" [Artificial intelligence in subjective management paradigms], *Filosofskie nauki*, 2021, vol. 64, no. 1, pp. 88–101. (In Russian)

Lepsky, V.E. "Sotsiogumanitarnye kriterii otsenki novatsii tsifrovoi real'nosti" [Socio-humanitarian criteria for evaluating innovations in digital reality], *SocioTime / Sotsial'noe vremya*, 2018, no. 4 (16), pp. 16–26. (In Russian)

Slavin, B. "Mozhet li iskusstvennyi intellekt byt' spravedlivym" [Can artificial intelligence be fair], *BIT*, 2021, no. 10 (113), pp. 32–35. (In Russian)