

К ОПРЕДЕЛЕНИЮ ПОНЯТИЯ «ЭМЕРДЖЕНТНЫЙ ИНТЕЛЛЕКТ»

П.О. Скобелев

Самарский федеральный исследовательский центр РАН
Россия, 443010, Самара, Студенческий переулок, 3А
E-mail: petr.skobelev@gmail.com

Ключевые слова: эмерджентный интеллект, коллективное принятие решений, конфликты, озарения, онтологии, мультиагентные технологии, цифровые двойники

Аннотация: Несмотря на значительные успехи в области «Искусственного интеллекта» (ИИ) на основе нейронных сетей и методов машинного обучения, вопрос о том, по какому пути пойдет создание систем общего (сильного) интеллекта, способных к целенаправленной творческой деятельности, остается открытым. В настоящем докладе делается попытка уточнить смысл и точнее определить понятие «Эмерджентный интеллект (ЭИ)», которое в последнее время все чаще используется в литературе для описания феноменов коллективного интеллекта в самоорганизующихся системах. В настоящей статье понятие ЭИ связывается с новой методологией решения сложных задач путем выявления и разрешения конфликтов между программными агентами сущностей с противоположными интересами, конкурирующими и кооперирующими на виртуальном рынке мультиагентной системы. Под актом разрешения конфликта («озарением») при этом понимается построение цепочки согласованных решений агентов, приводящая к переходу от одного неустойчивого «конкурентного равновесия» (консенсуса) - к другому, разрешающему противоречие интересов участников. Показывается, что модели, методы и алгоритмы разбора конфликтов и согласованного принятия решений могут базироваться на целевых функциях и протоколах переговоров агентов для поиска решений по взаимным уступкам в ходе достижения консенсуса. Даются примеры разбора конфликтов в задачах управления ресурсами, понимания текстов, извлечения знаний из данных и др. Рассматриваются теоретические основы создания ЭИ и перспективы использования ЭИ в качестве основы «сильного ИИ», а также обсуждаются перспективы развития и применения ЭИ в практических приложениях.

1. Введение

Главной новостью в области «Искусственного интеллекта» (ИИ) в начале прошлого года стал чат-бот ChatGPT от компании OpenAI на основе методов машинного обучения. Число пользователей ChatGPT уже в январе 2023 года достигло 100 миллионов, с более чем 25 миллионов пользователей в день и 615 млн. посетителей (по данным Similarweb.com).

Возникает вопрос, является ли подход ChatGPT наиболее прогрессивным способом создания ИИ для будущего и есть ли шанс, что этот подход в скором времени позволит создать системы «сильного (общего) интеллекта», даже без понимания многих, остающихся мало изученными аспектов интеллекта человека и, прежде всего, его творческого начала?

Цель доклада – предложение альтернативного подхода к созданию ИИ, который может быть назван «Эмерджентный интеллект (ЭИ)», трактующий любую целенаправленную деятельность через цепочки согласованных решений по разрешению конфликтов.

2. На пути к Индустрии 5.0 и Обществу 5.0

Безусловный успех чат-бота ChatGPT подтверждают начавшийся переход к новой формации Индустрии 5.0 и Общества 5.0 [1].

И если в рамках **Индустрии 4.0** решались в основном задачи автоматизации физических процессов для заводов и фабрик, то в **Индустрии 5.0** автоматизации будут подвергаться ментальные процессы, что потребует цифровизации знаний и создания автономных ИИ систем (агентов), играющих роль цифровых двойников предприятий и группировок роботов, коллективно принимающими согласованные решения [2].

Примеры таких новых видов сложных объектов управления могут включать высоко технологичные изделия, такие как самолет, представляющий собой сеть из более, чем 100 тысяч связанных агентов умных частей, многоуровневые группировки автономных малых космических аппаратов, дронов и беспилотных грузовиков, интеллектуальные предприятия, состоящие из программных роботов для различных сервисов, умную инфраструктуру с взаимодействием агентов складов, дорог и светофоров для обеспечения безопасности, эффективности и удобства транспортной среды и т.д.

В этой связи резко обострилась дискуссия о том, что можно считать ИИ и каковы особенности построения систем коллективного (распределенного) интеллекта [3, 4].

Ожидается, что внедрение коллективного ИИ в **Обществе 5.0** приведет, прежде всего, к трансформации управления за счет перехода от централизации – к распределенным системам, от иерархий – к сетям, от жестких бизнес процессов и регламентов – к гибкой самоорганизации, от приказов сверху-вниз – к переговорам, от фиксированной оплаты – к оплате по результату и т.д. Этот подход призван вернуть «Человека» в центр мира ИИ систем, для раскрытия талантов, использования творческого потенциала, воли, знаний и умений, энергии которого и будут договариваться цифровые двойники предприятий и группировок (роев) высокотехнологичных изделий в цифровых эко-системах ИИ будущего.

Таким образом, грядущие изменения требуют нового видения систем ИИ, в основе которых должны лежать процессы самоорганизации и коллективного согласованного принятия решений множеством людей и роботов, обладающих собственными интересами, предпочтениями и ограничениями.

3. Выбор принципов построения ЭИ систем будущего

При выборе новых принципов построения будущих систем ЭИ для Индустрии 5.0 и Общества 5.0 требуется обеспечить перспективу в части создания систем общего (сильного) интеллекта Artificial General Intelligence (AGI).

Вместе с тем, существующие ИИ системы на базе нейронных сетей и машинного обучения, достигшие бесспорного успеха в статистической имитации рассуждений, по своей сути остаются обычными программами. Нужны альтернативные подходы, которые позволят моделировать недетерминированные, конфликтные и противоречивые, нелинейные процессы творческой деятельности и позволят системам ИИ будущего обладать искусственным сознанием и встроенной моделью окружающего мира, целенаправленно рассуждать и действовать на основе общих и предметных знаний, причем с возможностью непрерывного расширения своей базы знаний за счет распознавания образов, абстрагирования, внешнего обучения и самообучения, иметь здравый смысл, воображение для моделирования возможных ситуаций и интуицию на основе опыта и т.д.

Гипотеза настоящей статьи состоит в том, что AGI система не может быть один раз и навсегда кем-то изначально построена, но напротив, такая система должна быть распределенной и самоорганизовываться из базовых элементов подобно тому, как строятся сложные системы в окружающем нас мире (атомы-молекулы-вещества-клетки и т.д.), развиваться и эволюционировать по мере приобретения опыта. Самоорганизация и эволюция системы должны проявлять себя в появлении новых классов и экземпляров элементов и отношений между ними, а также новых свойств как у этих отдельных элементов, так и у системы в целом, которые, в свою очередь, могли и перестраивать исходные базовые элементы, формируя индивидуальную траекторию развития системы.

Один из возможных новых подходов к созданию такого рода AGI систем может базироваться на принципе эмерджентности [4, 5].

Эмерджентность или эмергентность (от англ. гл. emergent – «возникающий, неожиданно появляющийся») в теории систем — появление у системы свойств, не присущих её элементам в отдельности. Эмерджентные структуры присущи многим природным (галактики, снежинки и ураганы), биологическим (вирусы и бактерии, клетки, рой пчел и колония муравьев) и социальным явлениям (формирование мнений в социальных сетях, изобретения и открытия, социальные революции и т.д.). В целом, эмерджентность трактуется как динамическое, не предопределенное заранее, внезапное появление глобального порядка, целостности, структуры, объекта, отношения или свойства на метауровне как результата локальных взаимодействий между элементами системы.

Впервые **эмерджентность** стала предметом исследований в теории сложных адаптивных систем [6], разработанной бельгийском ученым, проф. Ильей Пригожиным, Нобелевским лауреатом по химии (1977). Под **сложной адаптивной системой** понимаются открытые самоорганизующиеся системы, элементы которых обладают определенным уровнем автономности и способны взаимодействовать между собой. В таких системах, базирующихся на нелинейной термодинамике, формируются **«устойчивые неравновесия»** («неустойчивые равновесия»), поддерживаемые непрерывным притоком энергии извне и отводом (диссипацией) энергии на сопротивление (трение). При некоторых условиях в таких неравновесных структурах возникают **автокаталитические реакции**, которые и позволяют из существующего «простого» породить новое «сложное» на основе таких нелинейных явлений, как осцилляции, катастрофы, бифуркации, резонансы и т.д.

Действительно, второй закон термодинамики в физике гласит, что физические системы всегда необратимо обращаются в хаос, но при этом совершенно не объясняет, как появляются сложные системы из начального состояния хаоса и как они себя далее поддерживают. И.Пригожин утверждал, что открытые системы, получающие энергию и материю извне, могут проходить периоды нестабильности и самоорганизации, приводящие к формированию новых систем, чьи характеристики не могут быть заранее предсказаны.

В развитие этой теории американским биологом Стюартом Кауфманом была высказана гипотеза, что происхождение биологических организмов обусловлено вовсе не Дарвинскими принципами природного отбора, а **самоорганизацией и неравновесной динамикой** [7]. Под самоорганизацией при этом понимается процесс, формирующий спонтанный порядок из локальных взаимодействий частей изначально неупорядоченной системы. Дальнейшие исследования показали, что именно так формируются полимеры, в том числе, пептиды, обеспечивающие репродукцию сложных молекул и что любая живая клетка и любой более сложный организм – это также устойчивое неравновесие (аттрактор), что делает эту теорию одним из

кандидатов на объяснение истоков происхождения жизни на Земле. Так, еще в экспериментах Миллера-Юри (1953) было показано, как именно сложные аминокислоты и протеины могут формироваться из базовых элементов в условиях первичной Земли, при подогреве первичного «супа» за счет энергии от молний и вулканов.

Оказалось, что такие взаимодействия характеризуют процессы самоорганизации, протекающие в атомных бомбах и лазерах, именно так «накапливаются» и «взрываются» противоречия в социальных революциях, аналогично совершаются процессы озарений, ведущие к изобретениям и открытиям ученых и инженеров и т.п.

Резонно предположить, что если таким путем протекают столь сложные процессы в окружающей живой и неживой природе, и так зарождалась и формировалась сама жизнь на Земле, то и процессы развития естественного интеллекта имеют ту же природу, а, следовательно, эту же природу должны унаследовать и перспективные ИИ системы.

4. Понятие «Эмерджентного интеллекта»

В последнее время теме коллективного интеллекта (Collective Intelligence) придается все большее внимание в научных публикациях, в частности, в работе [8] проведен анализ около 40 тысяч статей, число которых выросло с десятков в 2000-м году – до примерно 4000 статей в 2021 году. Интересно, что среди указанных статей, посвященных, в основном, машинному обучению, самоорганизующихся ИИ систем не представлено вовсе.

В работе [9], в результате анализа 5600 статей, показано, что большее внимание «Эмерджентному интеллекту» стало уделяться в кибер-физических системах, но и здесь этот вид интеллекта, характерный для самоорганизующихся систем, рассматривается лишь как одно из перспективных направлений, находящихся пока в зачаточном состоянии.

Таким образом, как справедливо отмечается в [4], несмотря на успехи мультиагентных технологий, коллективный эмерджентный ИИ до сих пор фактически никак не представлен в существующих ИИ системах.

В развитие выбранной выше методологии будем называть «**Эмерджентным интеллектом (ЭИ)**» свойство (способность) сложной адаптивной системы решать поставленные задачи **путем выявления и разрешения конфликтов** в ходе конкуренции и кооперации программных агентов на виртуальном рынке системы до достижения неулучшаемого состояния «конкурентного равновесия» (консенсуса), когда ни один из агентов более не может улучшить решение.

Проявлением ЭИ будет построение цепочек согласованного принятия решений по изменению состояний агентов, внезапно, спонтанно и недетерминировано возникающих как при реакции системы на непредвиденные события, так и при внутренней активности (проактивности) агентов в ходе выявления и разрешения конфликтов, при котором устранение выявленного противоречия и достижение нового баланса интересов следует трактовать как самозавершенный акт мышления – «искру» или «вспышку» **озарения**, ценность которых для бизнеса может быть измерена (русским синонимом термина «ЭИ» в перспективе мог бы служить термин «Интеллект озарений» или «Творческий интеллект»).

В основе ЭИ-систем лежит новый класс мультиагентных (распределенных) моделей, методов и алгоритмов для коллективного принятия согласованных решений на виртуальном рынке такого рода систем, в котором конкурируют и кооперируют между собой агенты сущностей с противоположными интересами (подобно началам

«инь и ян»), единство и борьба которых и будут определять сложность поведения системы.

Примером может служить модель сети потребностей и возможностей (ПВ-сети) и метод компенсаций для разрешения конфликтов, нашедший ряд различных применений [5]:

- **Управление ресурсами предприятий:** создаются агенты ПВ-сети пяти основных классов: заказов, процессов (задач), продуктов, ресурсов и предприятия (подразделения) в целом. Задача построения расписаний ресурсов решается путем формирования «конкурентного равновесия» указанных агентов, адаптивно перестраиваемого по событиям, поступающим в реальном времени.
- **Понимание текстов:** в этих задачах ПВ-сеть была разработана для решения задач анализа синтаксиса, морфологии, семантики и прагматики. Ключевыми классами агентов являются агенты слов и смыслов. Поиск соответствия между словами и смыслами по мере прихода новых предложений также осуществляется через разбор конфликтов указанных агентов до достижения «конкурентного равновесия», при котором новые слова и смыслы больше не находят связей между собой.
- **Извлечение знаний из данных:** основными агентами ПВ-сети кластеризации являлись агенты кластеров, которые пытаются максимально расшириться путем поглощения записей, и агенты записей, которые стараются войти в лучшие для них кластеры. Процесс вычислений завершался построением «конкурентного равновесия», в котором ни один кластер или запись не могут более улучшить свое состояние.
- **Цифровые двойники растений:** в этой задаче ПВ-сеть включает агентов стадий роста и развития растений и агентов ресурсов, которые адаптивно пересчитывают по цепочке свои потребности и возможности и формируют «конкурентное равновесие» на основе данных от метеостанций и результатов осмотров посевов растений агрономами.

Модели виртуального рынка в указанных задачах имеют различную сложность, но все они наследуют принципы неравновесной термодинамики, т.к. каждый новый заказ приносит в систему новые виртуальные деньги, являющиеся аналогом энергии в физических системах, которые распределяются между агентами ПВ-сети и частично уходят в налоги, затраты на коммуникацию агентов и т.д. Применяемые методы (протоколы) выявления и разрешения конфликтов строятся на основе функций удовлетворенности и бонусов-штрафов классов агентов, отражающих борьбу и единство противоположностей, и многоитерационных аукцино-подобных переговоров между этими агентами с взаимными уступками и компенсациями для достижения баланса интересов (консенсуса).

При этом ЭИ подход использует онтологии для создания онтологических моделей объекта управления, играющих роль порождающей системы для создания агентов, что позволяет гибко настраивать мультиагентную систему на специфику решаемых задач.

Таким образом, разработанный ЭИ подход показал возможность решения на единой методической основе совершенно различных по своему характеру сложных задач через единство и борьбу агентов противоположностей, достигающих согласия.

В развитие ИО-подхода предполагается проведение следующих разработок:

- Разработка моделей агента целостности, делающего инвестиции в «узкие места» системы для направляемой самоорганизации с целью поддержания общего гомеостаза что позволит сочетать достоинства централизованного и распределенного подходов;

- Развитие онтологий для отделения предметных знаний от исходного кода системы, чтобы встраивать в систему новые возможности для ее самообучения из опыта, без перепрограммирования системы, в том числе, с применением технологий ChatGPT;
- Создание цифровых сетевых платформ для формирования многоуровневых ИО систем, устроенных по принципу «Роя роев».

Разрабатываемые модели и методы призваны обеспечить масштабирование ЭИ систем для создания цифровых двойников крупных предприятий и целых отраслей.

5. Заключение

Предлагаемое новое направление разработки ЭИ систем открывает возможность создания природо-подобного коллективного ИИ, базирующегося на фундаментальных принципах самоорганизации и эволюции, присущих живой природе.

Целью создания ЭИ-систем является большая интеллектуальность в решении поставленных задач и унификация подхода к созданию различного рода ИИ систем, а также большая открытость, оперативность, гибкость и адаптивность, эффективность, производительность, масштабируемость, надежность и живучесть ИИ систем.

Перспективы ЭИ систем связываются с созданием «сильного ИИ» здравого смысла, который может быть интерпретирован как баланс многих интересов.

Работа выполнена при поддержке Министерства науки и высшего образования по соглашению 075-03-2021-651/4.

Список литературы

1. Skobelev P.O., Borovik S.Y. On the way from Industry 4.0 to Industry 5.0: from digital manufacturing to digital society // *Industry 4.0*. 2017. Vol. 2, No. 6. P. 307-311.
2. Galuzin V., Galitskaya A., Grachev S., et al. Autonomous Digital Twin of Enterprise: Method and Toolset for Knowledge-Based Multi-Agent Adaptive Management of Tasks and Resources in Real Time // *Mathematics*. 2022. Vol. 10. P. 1662-1689.
3. Каляев И.А. Как измерить искусственный интеллект? // *Искусственный интеллект и принятие решений*. 2023. Т. 1. С. 95-103.
4. Словохотов Ю.Л., Новиков Д.А. Распределенный интеллект мультиагентных систем. Ч. 1. Основные характеристики и простейшие формы // *Проблемы управления*. 2023, № 5. С. 3-22.
5. Rzevski G., Skobelev P., Zhilyaev A. Emergent Intelligence in Smart Ecosystems: Conflicts Resolution by Reaching Consensus in Resource Management // *Mathematics*. 2022. Vol. 10. P. 1923-1946.
6. Prigogine I. *The End of Certainty: Time, Chaos and the new Laws of Nature*. Free Press. 1997.
7. Kaufman S. *At Home in The Universe: The search for the laws of self organization and complexity*. Oxford University Press, 1995.
8. Berditchevskaia A., Maliaraki E., Stathoulopoulos K. A descriptive analysis of collective intelligence publications since 2000, and the emerging influence of artificial intelligence // *Collective Intelligence*. August-September 2022. P. 1-12.
9. Leitão P., Queiroz J., Sakurada L. Collective Intelligence in Self-Organized Industrial Cyber-Physical Systems // *Electronics*. 2022. Vol. 11. P. 3213.