

ЭМЕРДЖЕНТНЫЕ СПОСОБНОСТИ В СИСТЕМАХ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Е. А. Рожкова

*студент экономического факультета, Белорусский государственный университет,
г. Минск, Беларусь, e-mail: elen.lenss@yandex.by*

Научный руководитель: Б. Н. Паньшин

*доктор технических наук, профессор, экономический факультет, Белорусский
государственный университет, г. Минск, Беларусь, e-mail: panshin@tut.by*

В последние годы искусственный интеллект набирает обороты, а достижения в области машинного обучения, обработки естественного языка и робототехники приводят к созданию все более сложных и автономных систем. Одним из наиболее интересных аспектов искусственного интеллекта является феномен эмерджентности, который способен произвести революцию в проектировании и понимании интеллектуальных систем. В статье рассматривается возникновение эмерджентных свойств, закон масштабирования, приводятся примеры систем искусственного интеллекта, которые обладают эмерджентными способностями.

Ключевые слова: эмерджентность; закон масштабирования; алгоритмы; системы искусственного интеллекта; производительность; системы Google DeepMind.

EMERGENT CAPABILITIES IN ARTIFICIAL INTELLIGENCE SYSTEMS

E. A. Rozhkova

*Student of the Faculty of Economics, Belarusian State University, Minsk, Belarus,
e-mail: elen.lenss@yandex.by*

Supervisor: B. N. Panshin

*Doctor of Technical Sciences, Professor, Faculty of Economics, Belarusian State
University, Minsk, Belarus, e-mail: panshin@tut.by*

Artificial intelligence has been gaining momentum in recent years, with advances in machine learning, natural language processing and robotics leading to increasingly complex and autonomous systems. One of the most interesting aspects of artificial intelligence is the phenomenon of emergentism, which has the potential to revolutionise the design and

understanding of intelligent systems. This article discusses the emergence of emergent properties, the law of scaling, and provides examples of artificial intelligence systems that have emergent capabilities.

Keywords: emergent; scaling law; algorithms; artificial intelligence systems; performance; Google DeepMind systems.

Эмерджентность означает, что простые объекты при правильном взаимодействии могут порождать высокоорганизованные и адаптивные системы. Способность является эмерджентной, если она отсутствует в мелкомасштабных моделях, но присутствует в более крупных [3]. Такие способности проявляются как неожиданные качества, возникающие в результате взаимодействия отдельных элементов в системах без явного программирования.

Формулировка и развитие данной концепции начались в XX веке. Так, в 1972 г. идея эмерджентности была затронута лауреатом нобелевской премии по физике Ф. Андерсоном в его статье «More Is Different» [2]. В данной работе автор утверждает, что чем сложнее система, тем наиболее вероятно появление новых свойств, которые невозможно предсказать на основе детального понимания отдельных её компонентов. Андерсон подчеркивает, что изучение сложных систем требует изменения научного мышления и иного подхода, признающего важность возможности появления эмерджентных свойств.

Возникновение эмерджентных способностей подчиняется закону масштабирования. Как правило системы масштабируются по трем факторам: объем вычислений, количество параметров модели и размер обучающего набора данных. В случае эмерджентных свойств существует определенная закономерность: при увеличении числа параметров производительность случайна, пока на определенном уровне не проявится новое свойство. Внезапно возникшее свойство способно значительно улучшить производительность. Такое качественное изменение также известно как фазовый переход – резкое изменение общего поведения, которое невозможно предвидеть при изучении менее масштабных систем, поэтому его также называют эмерджентным [3].

В последние годы всё активнее начались изучаться возможности использования эмерджентности при разработке систем искусственного интеллекта (ИИ) с целью создания более гибких, адаптивных и устойчивых машин. При разработке сложных алгоритмов и моделей машинного обучения, программируются основные правила и задачи, но конечные результаты могут превзойти изначальные представления разработчиков.

Рассмотрим наиболее распространённый пример эмерджентных способностей систем ИИ. Системы Google DeepMind (дочерняя компа-

ния Google, занимающаяся разработкой технологий ИИ) могут адаптироваться к изменяющимся условиям, справляться с неопределенностью и соответствующим образом корректировать свои стратегии. Они способны учиться на неудачах, изучать новые подходы и постоянно совершенствовать свою работу. Ярким примером является система AlphaFold, использующая методы глубокого обучения для предсказания структуры белков. Данная система привлекла внимание научного сообщества в 2018 г., когда победила в конкурсе, продемонстрировав новую возможность точного предсказания сворачивания белков – давней проблемы биологии. Также стоит упомянуть алгоритм ИИ AlphaGo, разработанный для игры в го. AlphaGo обладает эмерджентными способностями, такими как интуитивное и стратегическое мышление. В ходе матчей с профессиональными го-игроками алгоритм продемонстрировал новые, ранее неизвестные ходы и тактики, которые не были запрограммированы изначально [1].

Принципы эмерджентности в системах ИИ отражаются в возможности создания сложных алгоритмов и моделей, способных адаптироваться к новым ситуациям и находить неожиданные решения. Эмерджентные способности позволяют создавать умные системы, способные самостоятельно обучаться и развиваться. Понимание и применение принципов эмерджентности в системах ИИ является важным шагом в создании более эффективных систем, способных адаптироваться к постоянно меняющимся условиям.

Библиографические ссылки

1. История DeepMind [Электронный ресурс]. URL: <https://trends.rbc.ru/trends/social/64a51b559a79471af14745ca> (дата обращения: 02.10.2023).
2. Anderson P. W. More is different [Electronic resource]. URL: https://cserobotics.engr.tamu.edu/dshell/cs689/papers/anderson72more_is_different.pdf (date of access: 02.10.2023).
3. Wei J. Emergent abilities of large language models [Electronic resource]. URL: <https://browse.arxiv.org/pdf/2206.07682.pdf> (date of access: 02.10.2023).