СПЕЦИФИКА АББРЕВИАТУР В АНГЛОЯЗЫЧНЫХ НАИМЕНОВАНИЯХ СФЕРЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

В. В. Сергеева

Белорусский государственный университет

Минск, Беларусь

e-mail: SergeevaVV@bsu.by

Статья посвящена выявлению специфики аббревиатур сферы искусственного интеллекта. Исследуемые аббревиатуры были проанализированы с точки зрения их семантики и структуры. Установлены основные семантические классы, в которых применяются сокращенные наименования, а также доминирующие структурные типы аббревиатур сферы искусственного интеллекта.

Ключевые слова: искусственный интеллект; нейросетевые модели; аббревиатура, инициализм, термин; терминология.

SPECIFICS OF ABBREVIATIONS IN THE ENGLISH-LANGUAGE NAMES OF THE FIELD OF ARTIFICIAL INTELLIGENCE

V. V. Sergeeva

Belarus State University

Minsk, Belarus

e-mail: SergeevaVV@bsu.by

The article dwells on the consideration of the specifics of abbreviations in the field of artificial intelligence. The studied abbreviations were analyzed in terms of their semantics and structure. The main semantic classes in which abbreviated names are used, as well as the dominant structural types of abbreviations in the field of artificial intelligence are established.

Keywords: artificial intelligence; neural network models; abbreviation; initialism; term; terminology.

Искусственный интеллект (ИИ) все активнее внедряется во все сферы жизнедеятельности современного общества. С каждым днем расширяются границы его применения в таких областях как медицина, образование, наука, экономика, бизнес-процессы, банковская сфера, логистика, искусство и др. Интеллектуальные компьютерные программы уже стремятся превзойти способности человека в обучении, в решении проблем технического характера, способности мыслить, коммуницировать, рассуждать, принимать решения и даже создавать шедевры музыки, архитектуры, живописи и т.д. Для контроля данной сферы государства разрабатывают и принимают национальные стратегии и стандарты применения и развития искусственного интеллекта.

Терминология в сфере искусственного интеллекта (AI – Artificial Intelligence) начала формироваться в 60-х годах XX века и окончательно

закрепилась в начале 80-х годов. Тогда же начали издаваться энциклопедические справочники, терминологические и толковые словари по искусственному интеллекту. Большинство терминов по своему происхождению являются англоязычными, что связано с тем фактом, что интенсивное исследование в сфере искусственного интеллекта проводилось именно в Соединенных Штатах Америки.

Необходимо отметить, что так как данная сфера является междисциплинарной, то в ее терминологии встречаются термины логики, математики, кибернетики, программирования, лингвистики, психологии и др. (*abstraction* ‘абстракция’, *algorithm* ‘алгоритм’, *behaviourism* ‘бихейвиоризм’, *common sense logic* ‘логика здравого смысла’, *computer linguistics* ‘компьютерная лингвистика’, *decision tree* ‘дерево решений’, *hyper-event* ‘гиперсобытие’, *implication* ‘импликация’, *induction* ‘индукция’ и др.).

Стремительное развитие в области цифровых технологий и искусственного интеллекта ведет к созданию большого количества новых, в том числе сокращенных терминов.

Материал исследования составляют аббревиатуры, отобранные методом сплошной выборки из англоязычных глоссариев терминов сферы искусственного интеллекта: AI-Glossary produced by Cyber Gear [1] и Ultimate Artificial Intelligence Glossary 2021 [2].

В данном исследовании вслед за Е. С. Кубряковой, А. Н. Елдышевым, Е. А. Дюжиковой, И. В. Варфоломеевой и др. мы считаем аббревиацией любые процессы сокращения лексических единиц и выделяем такие типы аббревиатур, как слоговые (аферезис, апокопа, стяжение, синкопа); сложнослоговые, которые также включают сращения, вставки, накладки и частично сокращенные сложные слова; инициальные аббревиатуры (буквенные, акронимы, инициально-буквенные, инициально-слоговые и др.) [3, с. 8]. Кроме вышеназванных, выделяют также смешанный тип аббревиатур.

Е. С. Кубрякова определяет аббревиатуру как «процесс создания единиц вторичной номинации со статусом слова, который состоит в усечении любых линейных частей источника мотивации и который приводит в результате к появлению такого слова, которое в своей форме отражает какую-либо часть или части компонентов исходной единицы» [4, с. 71].

Семантический анализ исследуемого материала показал, что наибольшее количество аббревиатур являются наименованиями различных направлений сферы искусственного интеллекта, типов нейросетей, видов языковых моделей и обработки естественного языка, а также видов машинного обучения и машинного перевода.

Одним из ключевых понятий искусственного интеллекта является понятие *нейросеть*, которая представляет собой программное или аппаратное воплощение математической модели, созданной по образцу сетей нервных клеток живого организма (биологических нейронных сетей). Простейшие искусственные нейронные сети (*ANN* (Artificial Neural Network)) способны к обучению и могут находить простые взаимосвязи в массиве данных. Более сложные модели способны к глубинному обучению и выполнению сложных логических преобразований благодаря наличию в их структуре нескольких скрытых слоев нейронов. Среди сокращенных наименований различных видов нейросетей отметим, например, следующие: *CNN* (Convolutional Neural Networks) ‘сверточная нейронная сеть’, *GANs* (Generative Adversarial Networks) ‘генеративно-сопоставительная сеть’, *RNN* (Recurrent Neural Networks) ‘рекуррентная нейронная сеть’ и др.

Целый ряд аббревиатур можно встретить в таком направлении сферы искусственного интеллекта, как Large Language Models (LLM) ‘большие языковые модели’. Этот сегмент объединяет нейросетевые модели, которые используют различные алгоритмы машинного обучения, для обобщения, прогнозирования и генерирования языков на основе больших наборов текстовых данных. LLM активно применяются в чат-ботах, поисковых системах, для перевода текстов, при написании маркетинговых текстов, электронных писем и т.д. К таким нейросетевым моделям относятся, например, *GPT* (Generative Pre-trained Transformer) ‘генеративные предобученные трансформеры’, *LaMDA* (Language Model for Dialogue Applications) ‘языковая модель для диалоговых приложений’, *BERT* (Bidirectional Encoder Representation from Transformers) ‘двунаправленные презентации кодировщика для трансформеров’, *BLOOM* (BigScience Large Open-science Open-access Multilingual Language Model) и др.

В области машинного перевода (Machine Translation (MT)) как еще одного важного направления в сфере искусственного интеллекта присутствуют сокращенные наименования, обозначающие разные типы данного перевода в зависимости от подходов и стратегий перевода, вида лингвистического анализа текста и др.: *CBMT* (Corpus-based Machine Translation) ‘машинный перевод основанный на корпусах текстов’, *EBMT* (Example-based Machine Translation) ‘машинный перевод на базе примеров’, *HMT* (Hybrid Machine Translation) ‘гибридный машинный перевод’ (одновременное применение нескольких подходов машинного перевода), *NBMT* (Neural-based Machine Translation) ‘машинный перевод на базе нейронных сетей’, *RBMT* (Rules-based Machine Translation) ‘ма-

шинный перевод на основе правил’, *SBMT* (Statistical-based Machine Translation) ‘машинный перевод, основанный на статистике’.

Обработка естественного языка (*NLP* (Natural Language Processing)) представляет направление в машинном обучении, которое посвящено распознаванию или пониманию естественного языка (*NLU* (Natural Language Understanding)), его генерации (*NLG* (Natural Language Generation)) и обработке естественного языка (*NLP* (Natural Language Processing)). В этой области встречаются также такие аббревиатуры как *ASR* (Automatic Speech Recognition) ‘автоматическое распознавание речи’, *NLQ* (Natural Language Query) ‘запрос на естественном языке’, *NLT* (Natural Language Technology) ‘технология естественного языка’ и др.

Что касается структурных особенностей, то самой репрезентативной группой среди исследуемых аббревиатур в сфере искусственного интеллекта являются инициально-буквенные многокомпонентные сокращения, где каждый компонент исходного наименования представлен начальной буквой: *ANN* (Artificial Neural Network) ‘искусственная нейронная сеть’, *NLP* (Natural language processing) ‘обработка естественного языка’, *PEMT* (Post Edit Machine Translation) ‘постредактирование машинного перевода’, *RPA* (Robotic Process Automation) ‘роботизированная автоматизация процессов’ и др. Встречаются также инициализмы, включающие в свой состав не только начальные буквы (*XGA* – eXtended Graphics Adapter ‘расширенный графический адаптер’, *ONNX* – Open Neural Network eXchange ‘открытый обмен нейронными сетями’), но также целые слоговые отрезки (*CORAN* – Communication ORiented Application aNalysis, *AIBO* – Artificial Intelligence roBOT, *MLOps* – Machine Learning OPerationS). При этом, для признания сокращения инициализмом достаточно, чтобы в его состав входила хотя бы одна начальная буква из числа сокращаемых слов [5].

Среди аббревиатур смешанного характера можно, например, отметить аббревиатуру *GenAI* (Generative Artificial Intelligence) ‘генеративный искусственный интеллект, тип системы искусственного интеллекта, способной генерировать текст, изображения или другие медиаданные в ответ на подсказки’, которая является результатом применения конечного усечения (апокопа) и инициально-буквенной аббревиатуры. Аббревиатура *LangOps* (Language Operations) ‘языковые операции’, которая обозначает рабочие процессы и практики, поддерживающие обучение, создание, тестирование, производственное развертывание и постоянное сопровождение языковых моделей и решений на естественном языке, является примером многокомпонентного усечения, образованного путем сложения фрагментов двух слов развернутого наименования. В материале исследования встречаются также аббревиатуры с графиче-

ским знаком («цифрой-символом»), например, *E2E* (End-to-end Learning) ‘сквозное тестирование; тестирование на всех уровнях’. В данном случае комбинация звуков соответствует произношению цифры, которая их замещает. Частица *to* омонимична звуковой форме цифры 2.

Таким образом, исследуемые аббревиатуры сферы искусственного интеллекта были проанализированы с точки зрения их семантики и структуры. Подводя итог, следует отметить, что в соответствии с их значением наибольшее количество аббревиатур являются наименованиями различных направлений сферы искусственного интеллекта, типов нейросетей, видов языковых моделей и обработки естественного языка, а также видов машинного обучения и машинного перевода. Что касается структурных особенностей аббревиатур в сфере искусственного интеллекта, то было установлено, что самой многочисленной группой являются инициально-буквенные многокомпонентные сокращения, встречаются также инициализмы, включающие в свой состав не только начальные буквы, но также целые слоговые отрезки, аббревиатуры с графическим знаком («цифрой-символом») и некоторые другие аббревиатуры смешанного характера.

Библиографические ссылки

1. AI-Glossary [Electronic resource] Mode of access: <https://cyber-gear.io/AI-Glossary.pdf> (date of access: 05.08.2023).
2. Ultimate Artificial Intelligence Glossary 2021 [Electronic resource] Mode of access: <https://www.dlt.com/sites/default/files/resource-attachments/2021-01/Ultimate-Artificial-Intelligence-Glossary-2021-Ebook.pdf> (date of access: 12.08.2023).
3. Варфоломеева И. В. Аббревиатуры современного английского языка: когнитивно-дискурсивный аспект : автореф. дис. ... канд. филол. наук. М. , 2007. 25 с.
4. Кубрякова Е. С. Типы языковых значений. М. : Наука, 1981. 200 с
5. Криворот В. В. Особенности аббревиации в русском, английском и французском языках // Язык, речь, общение в контексте диалога языков и культур: Сборник научных трудов; отв. ред. О.И. Уланович. Мн. : Изд. центр БГУ, 2012. С. 42–50.