

УДК 004.8

Дербеденев Дмитрий Игоревич,

студент,

Кубанский государственный аграрный университет имени И.Т. Трубилина

Тапехина Татьяна Евгеньевна,

старший преподаватель кафедры иностранных языков,

Кубанский государственный аграрный университет имени И.Т. Трубилина

ЭВОЛЮЦИЯ МОДЕЛЕЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: ОТ АНГЛОЦЕНТРИЗМА К МУЛЬТИЯЗЫЧНОСТИ

В данной статье исследуется историческое развитие технологий искусственного интеллекта и появление первых языковых моделей. Ставятся задачи по анализу эволюции архитектур от экспертных систем до трансформерных моделей. Рассматривается формирование англоцентризма в разработке ИИ-моделей, обусловленного доминированием английского языка в научной коммуникации, веб-контенте и инвестиционной среде. На основе анализа современных рыночных, технологических и научных трендов делается вывод о том, что моноязычная парадигма исчерпала свой потенциал, а переход к мультязычным ИИ-моделям является стратегической необходимостью для глобальной экспансии и предотвращения возможных кризисов индустрии искусственного интеллекта.

Abstract

This article examines the historical development of artificial intelligence technologies and the emergence of the first language models. The tasks are to analyze the evolution of architectures from expert systems to transformer models. The formation of anglocentrism in AI model development, driven by the dominance of English in scientific communication, web content, and the investment environment, is examined. Based on an analysis of current market, technological, and scientific trends, the conclusion is drawn that the monolingual paradigm has exhausted its potential, and the transition to multilingual AI models is a strategic

necessity for global expansion and the prevention of potential crises in the artificial intelligence industry.

Ключевые слова: искусственный интеллект, моноязычные модели, англоцентризм, мультязычность, обработка естественного языка, языковая модель, большие данные.

Keywords: artificial intelligence, monolingual models, anglocentrism, multilingualism, natural language processing, language model, big data.

Антропоцентрическая цивилизация демонстрирует устойчивую тенденцию к технологической оптимизации жизненных процессов. Если XX столетие охарактеризовалось эрой механизации, то текущее столетие правомерно определяется как эпоха когнитивных машин. Интеграция искусственного интеллекта (ИИ) в повседневные практики произошла с высокой скоростью, трансформировав его из предмета академических исследований в повсеместный технологический инструментарий. Современные ИИ-модели стали неотъемлемым компонентом профессиональной, образовательной и бытовой сфер, функционально замещая поисковые системы, прикладное программное обеспечение и элементы социальной коммуникации. Данный феномен актуализирует исследовательский вопрос о предпосылках и факторах, обусловивших столь стремительную и латентную технологическую трансформацию.

Для анализа данного вопроса необходимо обратиться к историческим истокам. В период 1930–1950-х годов в междисциплинарном пространстве, синтезировавшем кибернетику, нейробиологию и математическую логику, были заложены концептуальные основы будущих вычислительных моделей. Фундаментальный вклад внёс британский математик Алан Тьюринг, формализовавший в 1936 году абстрактную «машину Тьюринга», ставшую теоретической моделью компьютера общего назначения [1]. В 1943 году нейрофизиолог Уоррен Мак-Каллок и математик Уолтер Питтс опубликовали работу «Логическое исчисление идей, относящихся к нервной активности», где предложили математическую модель искусственного нейрона, способного

при сетевом объединении выполнять логические операции [2]. Эти работы сформировали концептуальный образ «мыслящей машины», наделив вычислительный инструмент способностью к алгоритмическому синтезу информации.

Институционализация искусственного интеллекта как самостоятельной научной дисциплины связана с инициативой Джона Маккарти, предложившего данный термин на семинаре в Дартмутском колледже в 1956 году [3]. Последующая разработка Маккарти языка программирования LISP (1958) создала на десятилетия вперед инструментальную основу для исследований в области ИИ, оказав значительное влияние на такие языки, как Python и C [4]. Данный период ознаменовал собой переломный момент для машинного обучения, обозначив начало новой эры в обработке данных.

Следующий прорыв сопоставимой значимости произошёл спустя три десятилетия с появлением в 1990-х годах глобальной сети Интернет и сопутствующих больших массивов информации. В промежуточный период область ИИ переживала циклические колебания, периоды роста интереса и финансирования сменялись «зимами ИИ», вызванными неоправданными ожиданиями. Однако параллельно велись прагматичные разработки, такие как созданные в Стэнфорде экспертные системы Dendral (для определения молекулярной структуры органических соединений) и MYCIN (для диагностики заболеваний крови) [5]. MYCIN, в частности, использовала продукционную модель представления знаний (правила вида «ЕСЛИ – ТО»), что позволяло системе пояснять процесс генерации вывода [6]. Каждая подобная система стимулировала развитие смежных научных областей, однако для качественного скачка требовалась технологическая революция, которая стала возможной в 2010-х годах с утверждением парадигмы глубокого обучения.

Для непрофессионала различие между эрой интернета 2000-х и технологическим прорывом 2010-х может быть неочевидным, поскольку большие объёмы данных для обучения ИИ стали доступны уже в 1990-е.

Ключевая проблема заключалась в отсутствии адекватной модели для комплексной обработки информации. Преобладавшие модели были ориентированы на математические операции с числовыми данными. Переломным стал 2017 год с публикацией архитектуры Transformer, представляющей собой модель обработки естественного языка нового поколения [7]. Её структура, включающая варьируемое количество скрытых слоёв и механизм внимания, позволяет нейронным сетям моделировать сложные взаимосвязи в данных. На основе данной архитектуры были созданы наиболее производительные современные ИИ-модели, такие как GPT (Generative Pre-trained Transformer) от OpenAI и BERT (Bidirectional Encoder Representations from Transformers) от Google, которые легли в основу множества чат-ботов, включая DeepSeek.

Появление новой формальной технологии обработки естественного языка трансформировало общедоступные интернет-данные в ключевой ресурс для обучения. Последующее развитие моделей привело к интеграции мультимодальности, включив обработку изображений, аудио и видео. Именно в этот период в полной мере проявилась значимость английского языка как лингвистической основы машинного обучения.

Согласно различным оценкам, от 55% до 60% контента во Всемирной паутине представлено на английском языке. Организации типа Common Crawl формируют масштабные архивы веб-данных, где доминирует англоязычный текст. Эти архивы стали первичным источником данных для обучения моделей GPT и BERT [8]. Возникает эффект положительной обратной связи: поскольку наиболее мощные модели изначально обучались на английском, их наивысшая эффективность достигается именно с этим языком. Это стимулирует разработчиков создавать продукты на английском для использования передовых инструментов, что, в свою очередь, усиливает его лингвистическую доминанту.

Большинство научных публикаций и конференций осуществляется на английском языке, что отражает концентрацию научного знания в западном

академическом сообществе. Крупные платформы, такие как GitHub, объединяют разработчиков на англоязычной основе, а синтаксис большинства языков программирования также восходит к английскому.

С лингвистической точки зрения английский язык обладает рядом синтаксических особенностей, упрощающих автоматическую обработку: относительно фиксированный порядок слов, отсутствие развитой падежной системы и высокая стандартизированность грамматики снижают степень неоднозначности и ускоряют процесс машинного обучения.

Существенным фактором развития являются инвестиции и размер целевой аудитории. Исторически первые ИИ-модели возникли на западном рынке, который остаётся одним из наиболее ёмких и платёжеспособных. Инвестиции технологических гигантов, таких как Google и Microsoft, в ИИ-инфраструктуру исчисляются десятками миллиардов долларов, а общий объём венчурных инвестиций в ИИ-стартапы с начала 2025 года превышает 190 миллиардов долларов. Несмотря на финансовое доминирование, западные компании, действующие в условиях рыночной конкуренции, вынуждены осваивать новые рынки, что создаёт новые вызовы.

В современной экономике успех продукта коррелирует с широтой охвата аудитории. Инвесторы фокусируются на отраслях с потенциалом роста пользовательской базы. Если ранее ориентация на англоязычную аудиторию была оптимальной стратегией, то сегодня, с ростом азиатских рынков, сохранение моноязычной модели равносильно стратегическому поражению. Это может привести к негативному циклу, где сокращение аудитории повлечёт снижение инвестиций и технологическое отставание.

На текущий момент наблюдается активный рост числа научных публикаций на китайском языке на ведущих международных конференциях (NeurIPS, ICML), а также увеличение объёма разработок в области ИИ в Индии и России. Совокупное население этих стран составляет более трети мирового, а прирост новых пользователей здесь опережает западные рынки. Высокая конкуренция в сфере ИИ стимулировала появление мощных мультязычных

моделей, таких как DeepSeek и GigaChat. Российская компания Yandex, инвестирующая в разработку NLP для русского языка, создала модель YaLM (Yet another Language Model) — одну из первых крупных языковых моделей, обученных преимущественно на русскоязычных данных и демонстрирующую высокое качество обработки русского текста.

Заключение:

Таким образом, английский язык, исторически выступавший основой машинного обучения, вступает в конкурентные отношения с другими языками, включая китайский и русский. Современные ИИ-модели развиваются в парадигме открытости и мультиязычности, что позволяет максимизировать глобальный охват аудитории. Концентрация капитала, человеческих ресурсов и научных исследований в азиатском регионе вынуждает производителей адаптироваться, обеспечивая пользователям комфортное взаимодействие на предпочитаемом языке без семантических потерь. Такой подход также увеличивает объём и разнообразие данных для обучения, что является критическим фактором для совершенствования алгоритмов. Использование мультиязычных моделей представляется необходимым условием для предотвращения стагнации в развитии ИИ («зимы ИИ») и создания оптимальных условий для эволюции пользовательских интерфейсов на основе ИИ.

Список литературы:

1. Тьюринг, А. М. (1937). О вычислимых числах с приложением к проблеме разрешимости. Труды Лондонского математического общества, s2-42(1), 230–265.
2. Мак-Каллок, У. С., & Питтс, У. (1943). Логическое исчисление идей, относящихся к нервной активности. Бюллетень математической биофизики, 5(4), 115–133.
3. Маккарти, Дж., Мински, М. Л., Рочестер, Н., & Шеннон, К. Э. (1955). Предложение по Летнему исследовательскому проекту Дартмутского колледжа по искусственному интеллекту.

4. Маккарти, Дж. (1960). Рекурсивные функции символьных выражений и их вычисление на машине, Часть I. Коммуникации АСМ, 3(4), 184–195.
5. Линдси, Р. К., Бьюкенен, Б. Г., Фейгенбаум, Э. А., & Ледерберг, Дж. (1980). Применения искусственного интеллекта для органической химии: проект DENDRAL. McGraw-Hill.
6. Шортлифф, Э. Х. (1976). Компьютерные медицинские консультации: MYCIN. Elsevier.
7. Вашвани, А., Шейзер, Н., Пармар, Н., Ускоревейт, Дж., Джонс, Л., Гомес, А. Н., Кайзер, Л., & Полосухин, И. (2017). Внимание — это все, что нужно. Достижения в системах обработки нейронной информации, 30.
8. Браун, Т. Б., Манн, Б., Райдер, Н., Суббиа, М., Каплан, Дж., Дхаривал, П., ... & Амодей, Д. (2020). Языковые модели — это Few-Shot Learners. Достижения в системах обработки нейронной информации, 33, 1877-1901.

References:

1. Turing, A. M. (1937). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, s2-42(1), 230–265.
2. McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4), 115–133.
3. McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*.
4. McCarthy, J. (1960). Recursive functions of symbolic expressions and their computation by machine, Part I. *Communications of the ACM*, 3(4), 184–195.
5. Lindsay, R. K., Buchanan, B. G., Feigenbaum, E. A., & Lederberg, J. (1980). *Applications of Artificial Intelligence for Organic Chemistry: The DENDRAL Project*. McGraw-Hill.

6. Shortliffe, E. H. (1976). *Computer-Based Medical Consultations: MYCIN*. Elsevier.
7. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems 30 (NIPS 2017)*.
8. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language Models - are Few-Shot Learners. *Advances in Neural Information Processing Systems, 33*, 1877-1901.