

АНАТОМИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Черепанова Т. В., Корняков К. В.
Департамент разработки ПО искусственного интеллекта
YADRO
Нижний Новгород, Россия
E-mail: t.cherepanova@yadro.com, kirill.korniakov@yadro.com

Системы искусственного интеллекта – крайне обширная область знаний, стремительно развивающаяся в настоящее время, и на наших глазах меняющая мир вокруг нас, проникая во все технологические области. В ходе выступления мы попробуем заглянуть «под капот» систем, использующих искусственный интеллект, разобратся из каких программных и аппаратных компонент они собираются, и с какими сложностями сталкиваются их разработчики. Также, рассмотрим основные тренды в технологиях искусственного интеллекта, и какую роль в будущем индустрии играют производители платформ.

ВВЕДЕНИЕ

Посмотрим на определение искусственного интеллекта – это область компьютерных наук, которая занимается разработкой и созданием систем и программ, способных имитировать когнитивные способности человека, а именно, анализ информации, обучение из опыта, решение задач и принятие решений. Искусственный интеллект может включать в себя такие методы и технологии, как машинное обучение, нейронные сети, обработка естественного языка, компьютерное зрение, робототехника и другие. Он может применяться в различных сферах, включая медицину, робототехнику, экономику, финансы, игровую индустрию, автомобильную промышленность и другие.

Искусственный интеллект имеет потенциал изменить мир и повлиять на многие аспекты нашей жизни, повысив эффективность и точность многих процессов. Однако существуют и некоторые этические и социальные вопросы, связанные с применением искусственного интеллекта, которые требуют внимательного рассмотрения и регулирования. Искусственный интеллект, может быть, и источником определенных опасностей в случае неправильного обучения или использования, или привести к развитию зависимости, которая может привести к утрате человеческих навыков и способностей.

Несмотря на широкую распространённость технологий сегодня, область искусственного интеллекта продолжает активно развиваться, появляются новые модели использования, создаются все более сложные алгоритмы, которые в свою очередь требуют более мощных вычислительных ресурсов.

I. ОБЛАСТИ ПРИМЕНЕНИЯ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Искусственный интеллект имеет широкий спектр применений в различных областях. Некоторые из них включают:

- Медицина: ИИ используется для диагностики и прогнозирования заболеваний, создания индивидуализированных лечебных пла-

нов и разработки новых лекарственных препаратов.

- Финансы: ИИ применяется для анализа больших объемов финансовых данных, определения трендов и паттернов, прогнозирования рынка и принятия инвестиционных решений.
- Производство: ИИ помогает автоматизировать процессы производства, решать проблемы сборки и контроля качества, оптимизировать время и ресурсы.
- Коммуникации: ИИ используется для разработки систем автоматического распознавания речи, машинного перевода, создания виртуальных помощников и ботов для общения с клиентами.
- Безопасность: ИИ помогает обнаруживать аномальное поведение в системах безопасности, распознавать лица и голоса, анализировать видео материалы для предотвращения преступлений.

Это лишь некоторые области, применение искусственного интеллекта уже стало стандартом де-факто, уже изменило нашу жизнь, и потенциал развития огромный, и его применение будет продолжать расти и проникать во все области информационных технологий, сферы жизни и бизнеса. Предметная область носит стремительный, подрывной характер технологических инноваций, и открытые подходы и стандарты поддерживают это масштабное развитие. Например, 6 лет назад не было активного распространения технологий глубокого обучения и нейронных сетей, а использовали традиционное компьютерное зрение.

II. ЛАНДШАФТ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Для создания решений с использованием искусственного интеллекта требуются не только алгоритмические библиотеки и модели, но и множество технологий вокруг, обеспечивающих работу с данными, эффективное использование вычислительных ресурсов и операционное управление. Группа аналитиков регулярно выпускает отчет Machine Learning, Artificial Intelligence and

Data (MAD) Landscape, где в виде таблицы-карты собраны все ключевые компании и технологии, которые функционируют в области искусственного интеллекта, таблица огромна и стремительно меняется. Ландшафт включает в себя следующие ключевые направления:

- Инфраструктура: Различные системы хранения, доступа и защиты данных, оркестрация, мониторинг, системы безопасности.
- Аналитика: Платформы анализа и визуализации данных
- Машинное обучение и искусственный интеллект: универсальные инструменты для разработки, тестирования и развертывания моделей, подготовки данных и специализированные решения для определённых технологий.
- Приложения: множество готовых решения для широкого спектра задач.

Большинство решений базируется на фундаменте открытых (open-source) библиотек, моделей, технологий и стандартов, и тем самым открытый подход поддерживает бурный рост и развитие.

Первые шаги в создании искусственного интеллекта были сделаны еще в 1940-1950 годах. В это время ученые искали способы моделирования человеческого интеллекта в машинах. Одним из ранних прорывов стало создание теста Тьюринга в 1950 году, разработанного Аланом Тьюрингом. Активное развитие технологий началось в 1990-ые годы, когда появились более мощные компьютеры. В 1997 году компьютер Deep Blue созданный ИВМ, победил чемпиона мира по шахматам Гарри Каспарова. Рассмотрим стремительность эволюционного развития на примере языковых моделей. Первым широкоиспользуемым алгоритмом, основанным на принципах языковой модели, в начале 2010-х стал продвинутый «Т9», работающих на телефоне. В средах разработки появились подсказки ввода. В 2013 году появилось векторное представление слова. В 2018 году появилась первая GPT (Generative Pre-trained Transformer) модель, способная генерировать текст. GPT-2 состояла из 1.5 миллиарда параметров и 40 Gb данных тренировки и считается канонической моделью на которой основано дальнейшее развитие. Сегодня доступны открытые модели размером до 100 миллиардов параметров, а количество разных моделей на популярном ресурсе Hugging Face, поддерживающих диалог, превышает 2500.

III. СЛОЖНОСТИ ПОСТРОЕНИЯ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Фундаментальные принципы создания систем с использованием технологий искусственного интеллекта опираются на алгоритмы (или модели) и качественные входные данные, на которых эта модель будет обучена. Также немаловажным является интеграция искусственной

системы в другие технологические процессы, обеспечивая качественную работу и эффективное доверительное взаимодействие с человеком. Если заглянуть внутрь технологического стека, то его можно условно разделить на три уровня:

1. приложения и задачи, которые приносят непосредственно пользу;
2. программное обеспечение среднего слоя (framework или middleware) где реализованы сами алгоритмы;
3. аппаратное обеспечение.

Все три составляющих подбираются под конкретную задачу, объем данных и доступные вычислительные ресурсы. Рассмотрим три основные проблемы, которые возникают внутри технологического стека, которые могут быть не видны разработкам конечных приложений, но над их решением трудятся тысячи инженеров.

Проблема переносимости ПО и фрагментация

Существует несколько вычислительных аппаратных платформ доступных на рынке от разных производителей и каждый предлагает свое собственное решение, включая программный стек. Приложение разработанное под одну платформу не будет работать на другой, возможности переноса ограничены. Некоторые библиотеки и фреймворки могут не доступны на некоторых платформах. Наблюдаются тренды на унификацию и стандартизацию начиная от языков программирования до API алгоритмических библиотек. При разработке решения рекомендуется внимательно анализировать доступные библиотеки, опираться на уровень абстракции, когда переносимость будет бесшовной или с минимальными изменениями.

Комбинаторный взрыв вариантов реализации

Обеспечения функциональной переносимости в большинстве случаев будет недостаточно. Как правило алгоритмы искусственного интеллекта требуют значительных вычислительных ресурсов для исполнения и обучения. Внутри реализации алгоритма необходима оптимизация под определенную платформу, и здесь количество вариаций многократно увеличивается. Для уменьшения количества вариаций алгоритмы реализуются на языках, поддерживающих параллельные вычисления, таких как OpenCL, CUDA или SYCL и производители платформ предоставляют современные компиляторы основанные на LLVM технологии с IR представлением. К сожалению, не всегда можно достичь максимальной производительности с языками и компиляторные подходы выходят на доменный уровень, такие технологии как MLIR и TVM, где оптимизируется операции

высокого уровня и как кубики собираются в вычислительный конвейер. Закрытые экосистемы производителей накладывают ограничения на оптимизации, по сути их приходится разрабатывать отдельно под каждую экосистему. Современный тренд на открытые аппаратные платформы такие как RISC-V будет драйвером объединения экосистем и ускорит развитие технологий в целом.

Большие данные

Объем данных непрерывно растет и вычислительный ресурс одной машины оказывается недостаточными для решения множества задач искусственного интеллекта. Существуют подходы и библиотеки распределённых вычислений как низкого уровня – MPI, так и уровня приложения – Spark, которые предоставляют возможность выполнения одной задачи на нескольких узлах одновременно. Однако, использование данных методов усложняет процесс не только алгоритмического программирования, но и требует заблаговременного обдумывания масштабов данных и нагрузок на вычислительные ресурсы еще на стадии проектирования системы.

IV. ТРЕНДЫ В ТЕХНОЛОГИЯХ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Технологии продолжают свое неуклонное развитие, которое описывается циклом зрелости (Hype cycle) предложенным компанией Gartner. На это кривой видно какие из технологий уже подтвердили свою эффективность и активно используются, какие на стадии активного анализа их возможностей, а какие только исследуются. Сегодня в зоне внимания исследователей, экспертов и бизнеса находятся генеративные модели искусственного интеллекта, которые вышли на пик ожиданий. Вслед за ними выстраивается еще десяток захватывающих технологий, такие как

умные роботы, нейроморфные вычисления, ответственный искусственный интеллект. Каждая из них обладает своим уникальным потенциалом и обещает принести революцию в нашу жизнь. Таким образом, научное сообщество стоит на пороге новых достижений в области искусственного интеллекта, которые предоставят не только новые возможности, но и вызовы, требующие от нас как ученых и разработчиков сотрудничества и ответственного использования этих прогрессивных технологий.

V. Выводы

Непрерывно развивающиеся технологии искусственного интеллекта предоставляют и будут предоставлять научный и практический интерес в нескольких направлениях:

- Применение новейших технологий в практической деятельности.
- Исследования новых алгоритмов и подходов в области искусственного интеллекта направлены на поиск эффективных способов решения сложных проблем и улучшения качества жизни людей.
- Погружение в технологические стеки и решение фундаментальных проблем системного и аппаратного уровня, что позволит создавать более мощные и эффективные интеллектуальные системы, способные обрабатывать и анализировать огромные объемы данных.

СПИСОК ЛИТЕРАТУРЫ

1. The 2023 MAD (ML/AI/Data) Landscape [Электронный ресурс] / – Режим доступа: <https://mad.firstmark.com/>. – Дата доступа: 18.10.2023.
2. What's New in Artificial Intelligence from the 2023 Gartner Hype Cycle [Электронный ресурс] / – Режим доступа: <https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2023-gartner-hype-cycle/>. – Дата доступа: 25.10.2023.