

Искусственный интеллект: проблемы и перспективы развития

М. Макушин¹

УДК 621.37 | ВАК 05.27.01

Современный этап развития человечества связан с переходом в цифровую эру, формированием цифровой экономики. Основой этого перехода являются достижения в области микроэлектроники, вычислительной техники, средств / систем / сетей связи, программного обеспечения и искусственного интеллекта (ИИ), поддерживаемые прогрессом во многих других областях человеческой деятельности. Все эти направления являются критическими для перехода мировой экономики в новое состояние, и эффективность этого состояния будет определяться «самым слабым звеном». В свою очередь, развитие средств и технологий ИИ также обусловлено прогрессом в сфере вышеупомянутых направлений.

Первое определение ИИ было дано еще в 1956 году, при этом под ИИ понимались только вычислительные составляющие. Единого, общепринятого определения искусственного интеллекта нет до сих пор – все зависит от позиции, с которой дается определение ИИ. Но почти во всех определениях есть общие моменты: аппаратно-программное моделирование интеллектуальной деятельности человека; способность созданной человеком системы правильно интерпретировать внешние данные (структурированные и / или неструктурированные) и самообучаться на их основе; возможность формулировать логические выводы; работа в физической или цифровой среде с использованием возможностей, предоставляемых алгоритмами оптимального или неоптимального выбора из широкого множества возможностей. Развитие ИИ тесно связано с прогрессом в области микроэлектроники, вычислительной техники, программного обеспечения и ряда других дисциплин и технологий. Можно сказать, что предпосылки для этого развития полностью сформировались только сейчас – в период трансформационного перехода к цифровой экономике. Соответственно основными факторами развития ИИ становятся экономические – рентабельность систем ИИ, экономический эффект от их использования и т. д.

ГЛАВНЫЕ ТЕНДЕНЦИИ РАЗВИТИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПО ЦИКЛУ ЗРЕЛОСТИ КОРПОРАЦИИ GARTNER²

Цикл зрелости (понятие, введенное корпорацией Gartner) показывает, как в различных организациях осуществляется

внедрение ИИ. По данным Gartner, в период 2018–2019 годов процент организаций, использующих ИИ, увеличился с 4 до 14%. Внедрение ИИ на данном этапе отличается от того, которое было несколько лет назад, когда не существовало альтернативы созданию собственных решений в области машинного обучения (МО). Сейчас наибольший спрос предъявляется на автоматизированное МО машин и интеллектуальные приложения. Также популярны и другие подходы, такие как «платформа как услуга»³ или облачные сервисы ИИ.

Разговорный ИИ⁴ (Conversational AI) по-прежнему остается в центре внимания исследователей, что объясняется всемирным успехом Amazon Alexa, Google Assistant и других подобных приложений. Тем временем продолжают появляться новые технологии, такие как расширенный интеллект и объясняемый ИИ⁵. Цикл зрелости корпорации Gartner по ИИ рассматривает инновации и определяет тенденции развития ИИ. В 2019 году в графике технологической зрелости корпорации Gartner было представлено много новых технологий (рис. 1), но лишь

¹ Термин был введен в оборот аналитической компанией Gartner (Стэмфорд, штат Коннектикут, США).

² PaaS (Platform as a Service) – платформа как услуга, модель предоставления облачных вычислений, когда потребитель получает доступ к использованию информационно-технологических платформ, таких как ОС, СУБД, связующему ПО, средствам тестирования и разработки облачного провайдера.

³ Conversational AI – разговорный ИИ относится к использованию приложений для обмена сообщениями, речевых помощников и чат-ботов для автоматизации общения и для создания персонализированного взаимодействия с клиентами.

⁴ Explainable AI (XAI) – концепция «объясняемого ИИ» заключается в том, что технологии не только решают задачи, но и способны аргументировать свои действия.

¹ ЦНИИ «Электроника», главный специалист, mmackushin@gmail.com.

² Нуре cycle (Gartner Hype cycle) – цикл зрелости технологий, графическое отображение распространения, адаптации и социального влияния на специалистов и общество специфических технологий.

немногие имеют ценность и, вероятно, еще меньшее число этих технологий получит широкое распространение. Чтобы оценить ценность и риски ИИ, необходимо установить реалистичные ожидания в отношении ИИ.

Дополненный ИИ

Дополненный ИИ – это ориентированная на человека модель партнерства людей и ИИ, работающих вместе для повышения когнитивных способностей. Основное внимание уделяется вспомогательной роли ИИ в развитии человеческого потенциала.

ИИ, взаимодействуя с людьми и улучшая то, что они уже знают, уменьшает количество ошибок, рутинную работу и может улучшить взаимодействие с клиентами, обслуживание граждан и уход за пациентами. Цель дополненного ИИ – повысить эффективность автоматизации, дополняя ее человеческим подходом и здравым смыслом для управления рисками автоматизации принятия решений.

Чат-боты

Чат-боты – это «лицо» ИИ, они влияют на все области, где существует взаимодействие с людьми, например,

у производителя автомобилей KIA есть чат-бот, который общается со 115 тыс. пользователей в неделю. Чат-боты могут быть текстовыми или голосовыми, или их комбинацией. Подобные приложения существуют в отделах кадров, ИТ-службах поддержки и самообслуживания, но на первом месте – обслуживание клиентов, где чат-боты уже оказывают наибольшее влияние, изменяя способ взаимодействия с клиентами. Переход от принципа «пользователь изучает интерфейс» к подходу «чат-робот изучает то, что хочет пользователь» приобретает все большее значение для адаптации, повышения производительности и обучения на рабочем месте.

Машинное обучение

МО может решить различные проблемы бизнеса, такие как персонализированное обслуживание клиентов, рекомендации по цепочке поставок, динамическое ценообразование, медицинская диагностика или легализация денежных средств. Для извлечения знаний и моделей из данных МО использует математические модели. Освоение технологий МО расширяется, поскольку организации сталкиваются с экспоненциальным ростом объемов данных и достижений в вычислительной инфраструктуре.

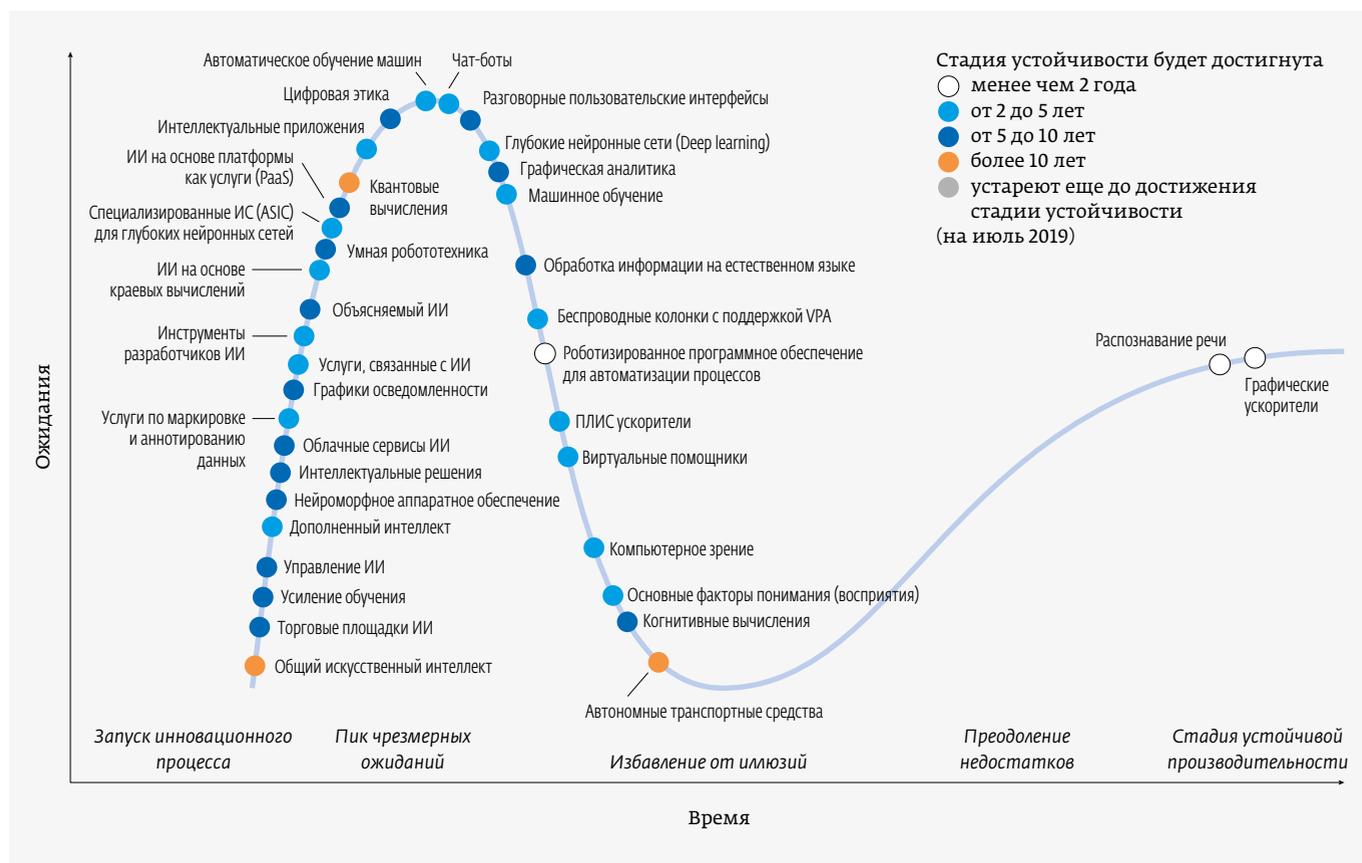


Рис. 1. Основные тенденции развития искусственного интеллекта в соответствии с циклом зрелости технологии, введенным корпорацией Gartner. Источник: Gartner

В настоящее время МО используется во многих областях и отраслях, чтобы стимулировать улучшения и находить новые решения для бизнес-задач. Например, American Express использует аналитику данных и алгоритмы МО, чтобы обнаружить мошенничество практически в реальном масштабе времени и избежать потенциальных многомиллионных убытков. Корпорация Volvo уже использует аналитические данные, чтобы предсказывать, когда детали могут выйти из строя или когда транспортные средства будут нуждаться в профилактическом или текущем обслуживании, что повышает безопасность автомобиля.

Этические и правовые нормы использования ИИ

Организации, осваивающие ИИ, должны уметь им управлять. Они должны знать о потенциальных нормативных и репутационных рисках. Вопрос управления ИИ – это процесс создания определенной политики.

Для развития управления ИИ требуется сосредоточиться на трех областях: доверие, прозрачность и разнообразие. Чтобы обеспечить успешное освоение ИИ необходимо сосредоточиться на решении вопроса о доверии к источникам данных и результатам применения ИИ. Также нужно определить требования прозрачности для источников данных и алгоритмов, чтобы снизить риски и повысить доверие к ИИ. Наконец, для достижения этики и точности ИИ, следует обеспечивать получение необходимых данных, алгоритмы и разнообразие точек зрения.

Интеллектуальные приложения

Большинство организаций предпочитают применять ИИ с его широкими возможностями в корпоративных приложениях. Интеллектуальные приложения – это корпоративные приложения со встроенными или интегрированными технологиями искусственного интеллекта, предназначенные для поддержки или замены человеческой деятельности с помощью интеллектуальной автоматизации, анализа данных и рекомендаций, направленных на повышение производительности и принятия решений.

Сегодня поставщики корпоративных приложений встраивают технологии ИИ в свои предложения, а также внедряют возможности платформы ИИ – от планирования корпоративных ресурсов до управления взаимоотношениями с клиентами, управления человеческим капиталом и повышения производительности труда.

Заказчикам готового программного обеспечения потребуется добиваться от своих поставщиков, чтобы они (поставщики) информировали о включении ИИ в маршрутные карты развития своей продукции с целью повышения ценности бизнеса в форме расширенной аналитики, интеллектуальных процессов и улучшенного пользовательского опыта [1].

НЕКОТОРЫЕ АСПЕКТЫ РАЗВИТИЯ ТЕХНОЛОГИЙ ИИ

Наступление эры ИИ реального масштаба времени

По мере того, как оказание услуг в реальном масштабе времени становится частью повседневной жизни, вычислительная инфраструктура претерпевает серьезные изменения. Появилась и развивается обширная экосистема – от интеллектуальных личных помощников, мгновенно предоставляющих информацию на естественном языке, до предприятий розничной торговли, генерирующих информацию о динамике покупательских предпочтений с помощью аналитики в магазинах. Подобные услуги, предоставляемые в реальном масштабе времени, открывают огромные рыночные возможности.

В целях извлечения выгод из подобных услуг, данные и аналитические данные должны быть мгновенно доступны. Они в значительной степени будут управляться сервисами с поддержкой ИИ. Облачные гиганты, такие как Amazon Web Services (AWS), Microsoft, Alibaba и SK Telecom, активно разрабатывают вычислительную инфраструктуру для предоставления этих услуг.

Операторы центров обработки данных (ЦОД) теперь должны оптимизировать вычисления для удовлетворения требований к реагированию в реальном масштабе времени. Следовательно, ИТ-архитектуры также должны учитывать изменяющиеся и быстро развивающиеся рабочие нагрузки и алгоритмы (в значительной степени управляемые ИИ) наряду с возрастающей интеграцией вычислений в системы хранения и сети.

Следующая волна вычислений должна быть адаптивной, когда ПО и аппаратное обеспечение объединяются, а также программируются для достижения необходимой производительности в реальном масштабе времени, максимальной пропускной способности, низкой задержки и высокой эффективности энергопотребления. По мере роста числа решений в области вычислений в реальном масштабе времени с применением ИИ, усложнения рабочих нагрузок и увеличения объемов неструктурированных данных в ЦОДах происходит сдвиг в сторону адаптивного ускорения вычислений, разработки, создания и эксплуатации сетей [2].

Перспективы развития нейроморфных вычислений

В своем последнем отчете «Нейроморфное восприятие и вычисления 2019» (Neuromorphic Sensing and Computing 2019) специалисты Yole Développement подсчитали, что рынок нейроморфных вычислений может вырасти с 69 млн долл. в 2024 году до 5 млрд долл. в 2029 году и 21,3 млрд долл. в 2034 году (рис. 2). Изменения, в первую очередь, затронут такие приложения, как смартфоны, робототехника и «умные» дома.

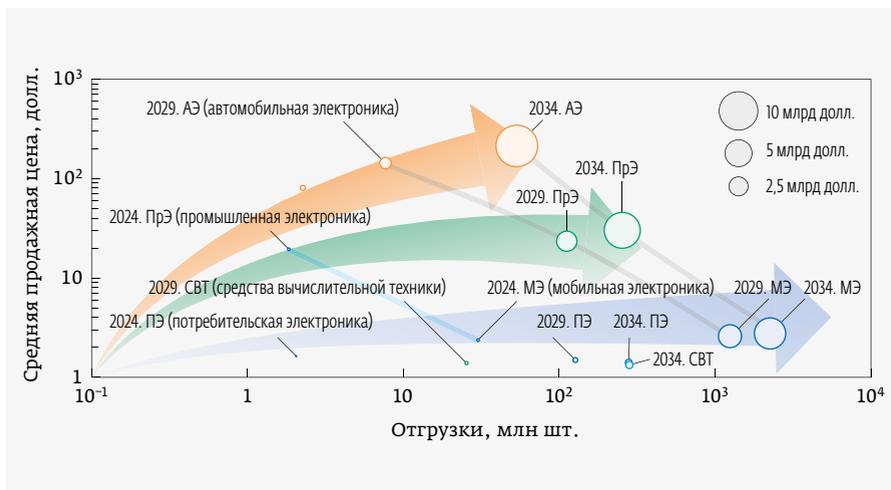


Рис. 2. Прогнозы продаж нейроморфных сенсорных и вычислительных устройств на 2024, 2029 и 2034 годы по сегментам рынка. *Источник: Yole Développement*

«перцептрона», которые используют значения с плавающей запятой и являются в настоящее время эталонной моделью нейронов в индустрии ИИ. Основное различие между «фон-неймановскими» вычислениями и нейроморфными вычислениями состоит в том, что первые, в основном, представляют собой математическую модель, которая следует аналитическим правилам, и время (как параметр) не оказывает на нее влияние. Введение времени, которое входит в понятие «импульса» и использования воспоминаний в связях, то есть синапсов, делает нейроморфный подход новой парадигмой, которая выходит за рамки современных ограничений фон Неймана [3].

Глубокое обучение является технологией с очень большим потенциалом. Вычислительные потребности глубокого обучения стали в настоящее время одним из существенных факторов успеха на рынке таких фирм, как Intel и NVIDIA. Также динамично развиваются многие стартапы, предлагающие новые способы улучшения вычислений. Большинство из этих компаний используют «фон-неймановскую» (принстонскую) архитектуру. Однако на рынке появляется новый тип разработчиков вычислительных средств, которые предлагают вырваться за рамки существующих ограничений ИИ. Одними из первопроходцев на этом направлении, отстаивающими новую парадигму ИИ, являются фирмы, специализирующиеся на средствах нейроморфного восприятия и вычислений.

Нейроморфные вычисления основаны на том, как мозг обрабатывает информацию, используя нейроны, синапсы и форматы данных, которые называются «пиковыми» (импульсными). Использование пиковых нейронов лучше имитирует работу мозга, чем нейроны в стиле

Искусственный интеллект и увеличение интереса к краевым вычислениям

Традиционные архитектуры ЦОД обладают централизованной структурой. С одной стороны, это обеспечивает существенные преимущества, а с другой – оказывается недостатком. По мере развертывания средств ИИ и других приложений, требующих минимального времени ожидания, подчас невозможного при использовании традиционных архитектур ЦОД, все большее внимание привлекают технологии краевых вычислений.

Проблема традиционных архитектур заключается в их централизованной структуре. Данные часто перемещаются на сотни километров от края сети до компьютеров ЦОД, а затем обратно. Это удобно, когда речь идет об электронной почте, Google, Facebook и других приложениями, предоставляемых через облачные структуры. Человеческий мозг не способен воспринимать время ожидания между, скажем, кликом мышкой по сообщению e-mail в браузере и открытием



Акционерное общество
ЭРКОН
Натурно-производственное объединение

603104, г. Нижний Новгород, ул. Мартова, д. 6,
помещение ПЗ, этаж 2, офис 204

тел. (831)464-50-23 [прямая], (831)202-25-52 отдел продаж

Разработка и производство постоянных непроволочных резисторов, СВЧ резисторов, поглотителей и чип-индуктивностей.

- Современная производственная база.
- Высокое качество.
- Индивидуальный подход к потребителю.

Новинки:

Аттенуатор (поглотитель) ПР1-25 (150 Вт, от 3 до 30 ДБ)
Резистор сверхвысокочастотный Р1-160 (до 20 ГГц)
Мощный резистор Р1-170 (до 1000 Вт)

www.erkon-nn.ru
e-mail: info@erkon-nn.com

Таблица 1. Движущие факторы развития краевых вычислений

Фактор	Причина использования
Объем создаваемых вне облачных ЦОД данных	Увеличение числа подключенных к различным сетям датчиков разных типов – до 1 трлн шт. и более к 2024 году
Издержки применения	В случае использования краевых вычислений отсутствует плата провайдерам интернет-услуг за передачу данных (за пропускную способность) и плата поставщикам облачных услуг за двустороннее перемещение данных и их обработку облачными ресурсами
Доступность услуг	Облачные вычисления могут быть недоступны в отдаленных и труднодоступных районах
Требования безопасности данных	Запрет на передачу в облако чувствительных к несанкционированному разглашению данных

сообщения на экране ПК пользователя. Однако ИИ и другие развивающиеся приложения, такие как Интернет вещей (IoT), облачные игры, виртуальная реальность, требуют гораздо меньшего времени отклика сети или времени ожидания. Это означает, что обработка данных в ЦОД должна перемещаться ближе к краю сети. Краевые (пограничные) вычисления^{*} могут выполняться в небольших ЦОД, размером примерно с транспортный контейнер, а не в зданиях размером со склад, в которых в настоящее время осуществляется поддержка облачных вычислений.

Самое главное в краевых вычислениях, что считается движущей силой их развития, – это объемы данных, создаваемых вне (облачных) ЦОД (табл. 1). Время ожидания – не единственная проблема, требующая для своего решения применения пограничных вычислений. Кроме того, различные организации нуждаются в краевых вычислениях там, где они не могут получить доступ к облаку: на корабле, в шахте или на нефтяном месторождении. Одновременно расширяющийся перечень правил конфиденциальности требует в некоторых приложениях обработки данных на месте, особенно в здравоохранении. Например, в больницах обычно категорически запрещено отправлять какие-либо данные в облако. Но даже если бы это было разрешено, то затраты на пропускную способность сделали бы перемещение большей части этих данных, особенно диагностических изображений, непомерно дорогим.

* Edge computing – краевые вычисления, метод оптимизации облачных вычислительных систем путем дополнения обработки данных на границе сети вблизи источника данных. Данный подход позволяет снизить интенсивность информационного обмена между датчиками и центром обработки данных, выполняя генерирование данных в источнике или вблизи него. Требует использования ресурсов, не подключенных к сети постоянно (ноутбуки, смартфоны, планшетные ПК, датчики и т. п.).

Расширение применения ИИ стимулирует рост спроса на краевые вычисления. Приложения ИИ представлены двумя основными рабочими нагрузками: обучением и формированием логических выводов. Обучение – это обучение модели ИИ методом решения какой-либо проблемы. Этот процесс часто включает в себя организацию больших (вплоть до петабайтов – 1 П байт = $1,13 \cdot 10^{15}$ байт) объемов данных. Обучение проходит на энергозатратных графических процессорах, причем каждая полностью загруженная стойка этих приборов потребляет от 30 до 40 кВт. Обучение, как правило, должно выполняться в больших ЦОД, чтобы удовлетворить требования к питанию, а также проблемы конфиденциальности и (законодательного) регулирования некоторых приложений.

После завершения обучения моделей наступает этап формирования логических выводов. Это процесс, в котором модель применяет полученные в процессе обучения алгоритмы и данные в своем производственном приложении. Формирование логического вывода требует гораздо меньшего объема данных и может выполняться на границе сети – в смартфоне, электромобиле, мини-ЦОД, на предприятии, магазине розничной сети или на бензоколонке.

К другим приложениям, требующим краевых вычислений (и часто использующие ИИ), относятся игры, IoT, интеллектуальные заводы, сфера доставки и логистики. Кроме того, розничные технологии требуют использования краевых вычислений для обеспечения необходимой оперативности реагирования.

Особый спрос на краевые ЦОД отмечается в розничной торговле. Такой «магазин будущего», как Amazon Go, имеет сотни камер, и точно так же сеть универсамов Walmart использует видео для наблюдения за торговыми залами. Потребности таких предприятий сетевой розничной торговли буквально разгоняют спрос на средства краевых вычислений.

Градостроители также стремятся использовать ИИ и краевые приложения для укрепления здоровья и безопасности горожан, отслеживания потребностей в обслуживании инфраструктуры и управления трафиком средств/сетей связи.

Устойчивый спрос на ИИ и краевые вычисления также предъявляют транспорт (в том числе автономные транспортные средства), перспективные производства и визуальный контроль продукции, энергетика (особенно на средства дистанционного контроля и управления) [4].

Перспективы развития микроконтроллеров с краевым ИИ

Объединение ИИ с IoT порождает «Интернет вещей на основе ИИ» (AIoT), который, помимо прочего, образует огромную новую область применения микроконтроллеров. Это, в свою очередь, будет означать, что машинное обучение больше не ограничивается миром суперкомпьютеров. В наши дни прикладные процессоры смартфонов помогают расширять его возможности, улучшая обработку аудио, изображений и голоса, прогнозирование человеческой деятельности, обработку и понимание простого языка, ускорение результатов поиска в базе данных и прочее.

В январе 2020 года в Кремниевой Долине состоялся 2-й Саммит разработчиков технологии микромашиного обучения (TinyML Summit), который показал, что число сторонников новой технологии будет быстро расти. Специалисты полагают, что первые микроконтроллеры с блоками машинного обучения появятся в течение 2–3 лет, а «захватчики рынка» – в ближайшие 3–5 лет. На первом саммите весной прошлого года корпорация Google продемонстрировала первую подобную структуру (TensorFlow Lite for Microcontrollers), обладавшую памятью емкостью всего 22 Кбайт, на основе ядра ARM Cortex M3, которая была предназначена для реализации модели выявления основных голосовых команд. Этот прибор был пригоден только для формулирования логических выводов, но не для обучения. Сейчас в данной области много стартапов, наблюдается рост венчурных инвестиций и начало процесса слияний/поглощений.

К разработкам сообщества TinyML большой интерес проявили ведущие поставщики микроконтроллеров, такие как NXP, Renesas и STMicroelectronics. Отмечается, что по мере того, как модели нейронных сетей миниатюризируются, объемы их возможностей расширяются. Большинство известных поставщиков микроконтроллеров считают, что для реализации возможностей TinyML больше всего подходят процессорные ядра семейства Cortex-M. Недавно фирма ARM представила новое ядро Cortex-M55, разработанное специально для целей машинного обучения, особенно

при совместном использовании с ускорителем ИИ Ethos-U55. Оба этих прибора созданы для сред с ограниченными ресурсами.

Доведение такого рода возможностей до масштабов микроконтроллера открывает обширные перспективы. Например, создание слухового аппарата, использующего ИИ для фильтрации фонового шума от разговоров. Другой вариант – умные бытовые приборы, распознающие пользователя по лицу и переключающиеся на его персональные настройки (в случае, если прибор используется несколькими пользователями). Наконец, узлы датчиков с поддержкой ИИ, способные работать в течение многих лет на самых крошечных батарейках. Помимо прочего, обработка данных в оконечной точке обеспечивает требования по времени ожидания, безопасности и конфиденциальности данных (за счет того, что они никуда не передаются), а это нельзя игнорировать [5].

Человеко-машинный интерфейс «умных» домов и краевой искусственный интеллект

Реальных примеров использования краевого ИИ в «умном» доме предостаточно. Развлекательные устройства, которые могут автоматически обнаруживать и масштабировать видеопотоки с низким разрешением до более высокого разрешения с превосходным качеством восприятия, что позволяет лучше использовать телевизионные дисплеи с высоким разрешением. Даже знакомые, а теперь почти повсеместные приложения для видеоконференций можно улучшить, добавив видео и аудио более высокого качества, и сделать их доступными на недорогих устройствах.

Такие решения могут сочетать видеосенсоры, датчики технического зрения и речевые сенсоры с возможностями ИИ обработки, обеспечивающие расширенную функциональность нового поколения знакомых устройств, таких как интеллектуальные дисплеи и звуковые панели, телевизионные приставки, бытовая техника и камеры видеонаблюдения.

Общим для этих приложений является потребность в решениях на основе краевого ИИ, специально разработанных для «умного» дома, а не для приложений

ООО
СМП



ИНТЕРНЕТ-МАГАЗИН
www.SMD.ru

электронные
для поверхностного
компоненты
монтажа

НОВОЕ В ПРОГРАММЕ ПОСТАВОК

- Разборные металлические EMI SMD экраны
- Кварцевые генераторы Q532 на частоты до 125 МГц

Москва, Ленинградский пр., 60 к. 32; e-mail: sale@smd.ru
Тел: (499) 158-7396, (495) 943-6244, (499) 943-6780

Таблица 2. Уникальность предъявляемых к краевому ИИ «умного» дома требований. Источник: EE Times

	«Умные» дома	Автомобили	Мобильные устройства
Изделия	Телеприставки, телевизоры, «умные» дисплеи, камеры, динамики / колонки и многое другое	Автомобили	Смартфон
Цена	Доллары – десятки долларов	Тысячи долларов и более	Сотни долларов
Размер дисплея	Крупноформатные	Нет данных	Малоформатные
Основная ценность краевого ИИ	Развлечения, безопасность	ADAS*	Качество изображения
Ключевые ИС	Обработка в реальном масштабе времени речевых и видеоданных, а также данных систем технического зрения, с помощью ИИ	Параллельная обработка данных с нескольких камер	Срок службы батареи

* ADAS (Advanced Driver Assistance Systems) – перспективные системы помощи водителю.

смартфонов или автомобилей (табл. 2). Чтобы сделать краевой ИИ еще более демократичным, такие решения должны обладать:

- возможностью поддерживать многорежимный пользовательский интерфейс с улучшенным ИИ, объединяющий речевые и видеоданные, а также данные систем технического зрения в эффективную систему;
- ИИ, более доступным с помощью стандартных инструментов для широкого круга разработчиков и новаторов;
- возможностью обеспечивать соответствие предлагаемых мер безопасности и конфиденциальности ожиданиям потребителей [6].

БЛИЖАЙШИЕ ПЕРСПЕКТИВЫ РЫНКА ИИ Ожидаемый вклад технологий ИИ в ВВП

Инвестиции в технологии ИИ растут в различных областях, включая национальную безопасность, финансы, здравоохранение, транспорт и «умные» города. По данным PricewaterhouseCoopers (PWC), потенциальный вклад технологий ИИ в мировую экономику в 2030 году достигнет 15,8 трлн долл. (табл. 3). Наибольшие экономические преимущества от увеличения ВВП за счет использования ИИ (14,5 и 26,1% соответственно) получают С. Америка и КНР – это 70% мирового экономического эффекта освоения ИИ.

Подобные перспективы вдохновляют разработчиков технологий ИИ, машинного и глубокого обучения. Это также открывает двери для поставщиков инструментальных средств, таких как поставщики САПР, в плане автоматизации некоторых этапов контроля безопасности, в частности в процессе верификации. Многие фирмы уже заявили о работах в этом направлении [7].

Расходы на ИИ к 2024 году удвоятся

По прогнозу корпорации IDC, мировые расходы на ИИ в течение следующих четырех лет удвоятся – с 50,1 млрд долл. в 2020 году до более 110 млрд долл. в 2024 году. Расходы на системы ИИ будут быстро увеличиваться в течение следующих нескольких лет по мере того, как организации будут внедрять средства ИИ в рамках своих усилий по цифровой трансформации и сохранению собственной конкурентоспособности в цифровой экономике. Среднегодовые темпы прироста в сложных процентах

Таблица 3. Выигрыш различных стран и регионов от увеличения ВВП в 2030 году за счет использования технологий искусственного интеллекта

Страна / регион	Вклад ИИ в рост ВВП в 2030 году	
	доля, %	величина, трлн долл.
КНР	26,1	7,0
Латинская Америка	5,4	0,5
Промышленно-развитые страны Азии	10,4	0,9
С. Америка	14,5	3,7
С. Европа	9,9	1,8
Ю. Европа	11,5	0,7
Прочие страны	5,6	1,2
Итого	13,6	15,8

(CAGR) затрат на эти цели в период 2019–2024 годов составят 20,1%.

Два ведущих фактора внедрения ИИ – это улучшение качества обслуживания клиентов и помощь сотрудникам в улучшении их работы. Две отрасли, которые будут тратить больше всего на решения ИИ на протяжении всего прогнозируемого периода – это розничная торговля и банковское дело. На долю ПО и услуг в 2020 году будет приходиться чуть более 1/3 всех расходов на ИИ, а остальное – на аппаратное обеспечение.

В региональном разрезе более половины всех расходов на ИИ в течение всего прогнозируемого периода придется на США, второе место займет 3. Европа, а КНР – третье. Наибольшие CAGR за период 2020–2024 годов будет наблюдаться в Японии (32,1%) и Латинской Америке (25,1%) [8].

Рост рынка чипсетов ИИ – в 8,75 раза за семь лет

Согласно данным исследовательской организации GMI, мировой рынок чипсетов ускорителей ИИ вырастет с 8 млрд долл. в 2019 году до 70 млрд долл. в 2026 году (табл. 4). CAGR за прогнозируемый период составят 35%.

Такая динамика отчасти объясняется ростом популярности платформ электронной торговли, использующих механизмы рекомендаций на основе ИИ, размещенные в облаке. Растущее признание облачных виртуальных помощников, использующих ИИ для понимания речи, также стимулирует спрос. Компании, оказывающие финансовые услуги, тоже активно внедряют ИИ для совершенствования своих возможностей по обнаружению мошенничества и лучшего расчета кредитных рисков.

Ожидается, что европейский рынок столкнется с резким ростом спроса благодаря тому, что многие фирмы региона разрабатывают системы автономного вождения. В автономных транспортных средствах ИС ИИ используются для интерпретации данных с камер и объединения их с информацией от других датчиков, таких как лидары и УЗ-датчики для формирования полной картины

Таблица 4. Прогноз развития рынка чипсетов искусственного интеллекта

Сектор/показатель	Значение
Среднегодовые темпы прироста в сложных процентах (CAGR) за период 2020–2026 гг., %	
Специализированные ИС (ASIC)	45
Краевые вычисления	>40
Технология обработки данных на естественном языке (NLP)	35
Облачные вычисления	30
Рынок С. Америки	38
Рынок АТР	>40
Рынок в целом	>35
Объем мирового рынка, трлн долл.	
2019	>8
2026	>70

окружающей среды. Ожидается, что за счет развития в Европе индустрии автономных транспортных средств CAGR продаж чипсетов ИИ в период 2020–2026 годов составит 33% (рис. 3).

В странах АТР рост CAGR за тот же период времени составит 40%. Один из факторов – реализация в КНР с 2017 года «Плана развития средств искусственного интеллекта нового поколения» (新一代人工智能发展计划) с общим бюджетом 150 млрд долл.

Рынок чипсетов ИИ С. Америки в прогнозируемом периоде до 2026 года также будет демонстрировать ускоренное развитие сектора краевых вычислений (рис. 4).

Отмечается, что 25% всего рынка чипсетов ИИ в 2019 году было использовано для понимания языка с помощью алгоритмов обработки данных на естественном языке (natural language processing, NLP). Продажи



Рис. 3. Прогноз структуры продаж чипсетов ИИ в Европе за период 2016–2026 годов. Источник: Global Market Insights

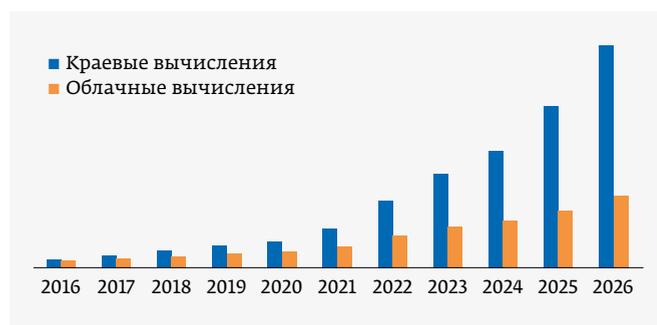


Рис. 4. Рынок чипсетов ИИ С. Америки в период до 2026 года. Источник: Global Market Insights

чипсетов в данном секторе, используемых в интеллектуальных динамиках, автомобильных информационно-развлекательных системах, устройствах с голосовым управлением и приборах IoT в период до 2026 года будет демонстрировать CAGR на уровне 35% [9].

Развитие сектора ИС краевого искусственного интеллекта

По прогнозу Deloitte Insights, в 2020 году в мире продано более 750 млн ИС краевого ИИ – кристаллов ИС или их частей, предназначенных для выполнения и/или ускорения задач машинного обучения на окончательном приборе, а не в удаленном центре обработки данных (ЦОД). Суммарная стоимость их продаж составит 2,6 млрд долл. Кроме того, рынок ИС краевого ИИ будет расти намного быстрее, чем рынок ИС в целом. В 2024 году отгрузки ИС краевого ИИ превысят (возможно намного) 1,5 млрд шт. Соответственно CAGR за прогнозируемый период составят не менее 20% по сравнению с 9% по всей полупроводниковой промышленности.

Подавляющее большинство ИС краевого ИИ (более 70%) будет использовано в старших моделях смартфонов, то есть в 2020 году и в течение следующих нескольких лет рост отгрузок ИС краевого ИИ будет определяться главным образом смартфонами. Предполагается, что в 2020 году более трети из отгруженных 1,56 млрд смартфонов будет оснащено подобными микросхемами.

Смартфоны не единственные устройства, в которых используются ИС краевого ИИ – они также содержатся в планшетных ПК, носимой электронике (wearables), «умных» динамиках и т. п. В краткосрочной перспективе подобные приборы, скорее всего, окажут гораздо меньшее влияние на продажи ИС краевого ИИ, чем смартфоны. Это обусловлено тем, что их рынок не растет (планшетные ПК), или он слишком мал, чтобы существенно изменить ситуацию (например, в 2020 году отгрузки «умных» динамиков и носимой электроники со встроенными ИС краевого ИИ составят всего 125 млн шт. – рис. 5) [10].



Рис. 5. Прогноз структуры рынка ИС краевого ИИ в 2020 и 2024 годах (млн шт.).

Источник: Deloitte Insights

АМЕРИКАНО-КИТАЙСКОЕ ПРОТИВОСТОЯНИЕ В ОБЛАСТИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Идущая в настоящее время Холодная технологическая война между США и КНР напоминает прежнюю Холодную войну между США и СССР. Только если раньше в центре внимания была космическая гонка, то сейчас одним из основных полей сражений стала область ИИ. Действительно, средоточием стратегической конкуренции между Китаем и Западом, наряду с базовыми полупроводниковыми технологиями, стали ИИ и алгоритмы, порожденные технологиями двойного применения.

Недавно на одном из мероприятий Управления перспективного планирования оборонных научно-исследовательских работ (DARPA) МО США, посвященном ходу реализации его «Инициативы по возрождению электроники» (Electronics Resurgence Initiative), отмечалось: «...лидерство США в микроэлектронике имеет жизненно важное значение для лидерства США в области искусственного интеллекта».

Несмотря на огромные суммы, затрачиваемые в США государственным и частным секторами на развитие ИИ, некоторые отраслевые аналитики утверждают, что для того, чтобы идти в ногу с Китаем, необходимо еще больше инвестиций – с учетом объемов средств, затрачиваемых на эти же цели «Поднебесной» в рамках общенационального плана «Сделано в Китае 2025» (中国制造2025). Например, аналитики корпорации Lux Research считают, что недавно выделенные средства в размере примерно 1 млрд долл. на работы специализированных на ИИ и квантовых вычислениях институтов не смотрятся на фоне 50 млрд долл., зарезервированных Пекином на эти цели в рамках плана работ «Инвестиционного фонда микроэлектронной промышленности КНР» (中国集成电路产业投资基金). Имеются в виду вновь созданные и финансируемые Национальным научным фондом США (National Science Foundation, NSF) пять институтов, работа которых начнется в следующем году. На их поддержку американские корпорации Accenture, Amazon, Google и Intel, со своей стороны, выделили более 160 млн долл.

Кроме того, увеличение инвестиций на ИИ предусмотрено проектом «Закона о создании полезных инициатив по стимулированию производства полупроводниковых приборов в Америке» (Creating Helpful Incentives to Produce Semiconductors for America Act, CHIPS for America Act). В своей нынешней форме этот закон предусматривает выделение на исследования в области ИИ в США около 1,157 млрд долл.

Указывается, что китайские технологии в области ИИ быстро

совершенствуются, и КНР уже стала «игроком высшего уровня» по большинству направлений развития ИИ. Более того, цели китайского правительства заключаются не только в том, чтобы догнать ведущие страны мира, но и в том, чтобы возглавить мир. Холодная технологическая война между США и КНР делает некоторые американские технологии недоступными для китайских фирм (например, Huawei сталкивается с ограничениями доступа к американским микроэлектронным технологиям), что вынуждает Китай ускорять и расширять свои внутренние программы НИОКР для разработки технологий, сопоставимых или даже более перспективных, чем современные «глобальные» технологии.

Доклад корпорации Lux Research, представленный на недавней конференции по нейронным системам обработки информации (Conference on Neural Information Processing Systems), показал, что нынешнее лидерство США в области ИИ во многом обусловлено способностью американских компаний и университетов привлекать и удерживать ведущих китайских исследователей [11].

Существует ли превосходство КНР в области искусственного интеллекта?

В недавно опубликованном исследовании «Конкуренция в области ИС искусственного интеллекта: вызов Китая в условиях технологической войны» (Competing in Artificial Intelligence Chips: China's Challenge amid Technology War) утверждается, что опасения по поводу того, что Китай может угрожать американскому доминированию в сфере ИИ, «не основаны на реальности».

Основные выводы доклада:

- Китай очень опоздал с началом НИОКР в области ИИ. Соединенные Штаты начали работы 60 лет назад, сосредоточившись «с самого начала на фундаментальных прорывных исследованиях». В КНР, напротив, исследования в области ИИ до 1980-х годов не проводились;
- деятельность в области ИИ в КНР во многом определяется разработкой приложений для ИИ.

Возможно, что один из самых важных выводов этого исследования заключается в том, что технологические ограничения США вынуждают Китай укреплять фундаментальные и прикладные исследования в сфере ИИ, чтобы догнать США по основным фундаментальным технологиям. По иронии судьбы, ограничения на экспорт технологий из США стимулируют реформу китайской инвестиционной и инновационной политики в области высоких технологий. Это может помочь Китаю исправить одну из фундаментальных слабостей своей инновационной системы в области ИИ [12].

* * *

По мере дальнейшего развития ИИ будет происходить его сближение с интеллектом человека, этому будут

способствовать исследования в области нейроморфных вычислений. Инвестиции в технологии ИИ, как ожидается, обеспечат высокий уровень отдачи. Чем они больше, тем выше эффект. По прогнозам именно у КНР и США, тратящих на исследования в области ИИ наибольшие средства, вклад ИИ в рост ВВП к 2030 году будет самым высоким. Нельзя также не отметить важность вклада технологий ИИ в обеспечение обороноспособности и интересов национальной безопасности. Чем сложнее стоящие задачи, тем более высокий уровень ИИ становится необходимым.

ЛИТЕРАТУРА

1. **Goasduff L.** Top Trends on the Gartner Hype Cycle for Artificial Intelligence. 2019. Gartner, 09.12.19.
2. **Scraba A.** Entering the Era of Real-Time AI // EE Times magazine. 05.22.2020.
3. **Levinson R., Tschudi Y.** Neuromorphic computing, a better solution for Artificial Intelligence? – An interview with BrainChip // I-Micronews magazine. 02.27.2020.
4. **Wagner M.** AI Drives Data Centers to the Edge // EE Times magazine. 05.11.2020.
5. **Ward-Foxton S.** Adapting the Microcontroller for AI in the Endpoint // EE Times. 02.19.2020.
6. **Ganju V.** Edge AI Solutions for Smart Homes Can Transform HMI // EE Times magazine. 07.30.2020.
7. **Sperling E.** Security Holes In Machine Learning And AI // Semiconductor Engineering. July 9th. 2018.
8. Worldwide Spending on Artificial Intelligence Is Expected to Double in Four Years, Reaching \$110 Billion in 2024, According to New IDC // Semiconductor Digest. August 28. 2020.
9. **Ward-Foxton S.** AI Chip Market Will Hit \$70B in 2026 // EE Times. 04.23.2020.
10. **Stewart D., Loucks J.** Putting AI into the Edge is a No-Brainer and Here's Why // EE Times. 02.18.2020.
11. **Leopold G.** AI Race Comes Down to Money and Brains // EE Times. 09.10.2020.
12. **Yoshida J.** Debunking the Myth of China's AI Superiority // EE Times magazine. 04.22.2020.

ООО "Руднев-Шиляев"

Разработка и производство:

- платы сбора данных
- измерительные приборы
- виброакустические системы
- инструментальные решения задач заказчика

Москва (495) 787-63-67
(495) 787-63-68

www.rudshel.ru
adc@rudshel.ru