Следующий технический документ китайского аналитического центра описывает состояние искусственного интеллекта в Китае и мире. Он делит свое внимание на инновации и прорывы в области ИИ, инженерные разработки и другие практические применения ИИ, а также инициативы по управлению ИИ в области надежности и безопасности.

Название
Белая книга по искусственному интеллекту (2022)
人工智能白皮书（2022年）

Автор
Китайская академия информационных и коммуникационных технологий (CAICT; 中国信息通信研究院; 中 国信通院). CAICT является аналитическим центром при Министерстве промышленности и информационных технологий КНР (MIIT; 工业和信息化部; 工信部).

Источник
Веб-сайт CAICT, 12 апреля 2022 года.

Исходный текст на китайском языке доступен в Интернете по адресу:
http://www.caict.ac.cn/kxyj/qwfb/bps/202204/P020220412613255124271.pdf
Архивная версия исходного текста на китайском языке доступна в Интернете по адресу:
https://perma.cc/SAC8-GZ5D *1 доллар США ~ 6,7 китайского юаня юаньминби (RMB), по состоянию на 6 июня 2022 года.*

Дата перевода
6 июня 2022

Translator **Редактор - переводчик**
Языковая группа Etcetera, Inc. Бен Мерфи, менеджер по переводу CSET

CAICT

№ 202205

Белая книга по искусственному интеллекту

(2022)

Китайская академия информационных и коммуникационных технологий
(CAICT)

Апрель 2022 года

Заявление об авторских правах

Авторские права на этот технический документ принадлежат Китайской академии информационных и коммуникационных технологий (CAICT) и защищены законом. "Источник: Китайская академия информационных и коммуникационных технологий" должно быть указано при любом использовании текста или мнений данного технического документа путем воспроизведения, извлечения или другими способами. CAICT будет стремиться привлечь нарушителей вышеуказанного Заявления к юридической ответственности.

Предисловие

На фоне новой технологической и научной революции и промышленных преобразований глубокая интеграция искусственного интеллекта (ИИ) и промышленности является неизбежным выбором для высвобождения мультипликативного эффекта цифровизации, ускорения развития стратегических новых отраслей и создания всеобъемлющих конкурентных преимуществ. В настоящее время быстрое проникновение ИИ в различные отрасли промышленности способствует сквозной интеграции и развитию между развивающимися отраслями, между развивающимися отраслями и традиционными отраслями, а также между технологиями и обществом. В начале "14-го пятилетнего плана" всестороннее изучение тенденций развития ИИ имеет большое значение и актуальность.

В этом техническом документе основное внимание уделяется описанию аспектов политики, технологий, приложений и управления в области искусственного интеллекта. На **политическом уровне**стратегическое положение ИИ постоянно укреплялось в Китае и за рубежом, чтобы способствовать высвобождению дивидендов от ИИ. На **техническом и прикладном уровне**технология искусственного интеллекта, представленная глубоким обучением, быстро развивалась, и новые технологии начали изучаться и применяться; инженерные возможности (工程化能力) были расширены.
постоянно совершенствуется и продолжает активно применяться в таких областях, как здравоохранение, производство и автономное вождение; а надежная технология искусственного интеллекта привлекает широкое внимание общества. В то же время **governance-level** работе на уровне руководства также уделяется большое внимание со стороны всего мира. Процесс регулирования продолжает ускоряться в различных странах, и промышленная практика, основанная на надежном ИИ, продолжает углубляться.

В целом, в этой Белой книге утверждается, что ИИ постепенно вступил в новую стадию, и направление развития на следующем этапе будет определяться и определяться "трехмерными" (3D) координатами **технологических инноваций, инженерных практик** (工程实践)**, надежности и безопасности** (可信安全). В частности, первый
измерение подчеркивает инновации. Инновации, связанные с алгоритмами и вычислительной мощностью (compute), будут продолжать постоянно появляться. Второе измерение выделяет инженерное дело. Инженерные возможности постепенно стали ключевым фактором, позволяющим ИИ расширять возможности тысяч отраслей промышленности в больших масштабах. Третье измерение подчеркивает надежность. Разработка ответственного и заслуживающего доверия ИИ стала консенсусом, и в центре внимания станет внедрение абстрактных принципов управления на протяжении всего жизненного цикла ИИ.

В связи с быстрым развитием ИИ, его широким спектром влияния и беспрецедентной степенью влияния нам необходимо еще больше углубить наше понимание ИИ. Мы приветствуем ваши критические замечания и исправления по поводу любых недостатков в этой Белой книге.

Содержание

Список цифр

I. Обзор развития искусственного интеллекта

ИИ - это новая стратегическая технология, которая определит будущее и станет важной движущей силой нового витка научно-технической революции и промышленных преобразований. Много раз Генеральный секретарь Си Цзиньпин давал важные указания, подчеркивая, что "мы должны глубоко понять особенности разработки ИИ нового поколения, усилить интеграцию ИИ и развитие промышленности и придать новый импульс высококачественному развитию". В последние годы технологии, связанные с искусственным интеллектом, продолжали развиваться, процесс индустриализации и коммерциализации продолжал ускоряться, и в настоящее время ускоряется их глубокая интеграция с тысячами отраслей промышленности. Находясь на особом этапе начала "14-го пятилетнего плана", мы твердо верим, что всесторонний обзор тенденций развития политики, технологий, приложений и управления ИИ может помочь достичь консенсуса в отрасли и способствовать устойчивому и здоровому развитию ИИ.

(i) Все страны постоянно совершенствуют стратегии искусственного интеллекта и одну за другой используют важные возможности для развития

Искусственный интеллект стал ключевой областью технологических инноваций и важной опорой эпохи цифровой экономики. С 2016 года более 40 стран и регионов подняли развитие искусственного интеллекта до уровня национальной стратегии. За последние два года, особенно под воздействием пандемии COVID-19, все больше и больше стран признали, что ИИ играет ключевую роль в повышении глобальной конкурентоспособности, и одну за другой углубляли свои стратегии в области ИИ. **ЕС** выпустил *Цифровой компас 2030 года: европейский путь к цифровому десятилетию* и *Обновление промышленной стратегии 2020года,* которые призваны всесторонне изменить глобальное влияние цифровой эпохи. В этих документах содействие развитию искусственного интеллекта указано в качестве важной задачи. **Соединенные Штаты** последовательно создали Национальное управление инициативы в области искусственного интеллекта, Национальную целевую группу по исследовательским ресурсам в области искусственного интеллекта и другие учреждения. Различные ведомства активно разрабатывают ряд стратегий, поднимающих ИИ до уровня "индустрии будущего" и "технологии будущего", постоянно укрепляя и повышая глобальную конкурентоспособность Соединенных Штатов в области ИИ и обеспечивая их статус "лидера". После разработки своей *Комплексной инновационной стратегии 2020года* **Япония** выпустила *Стратегию ИИ 2021* в июне 2021 года, которая направлена на продвижение инноваций и планов создания в области ИИ и всестороннее построение оцифрованного правительства. В сентябре 2021 года **Соединенное Королевство** опубликовало новую десятилетнюю стратегию национального ИИ, которая является еще одной важной стратегией, запущенной после 2016 года и направленной на изменение влияния сферы ИИ. **China's** *Предложение Центрального комитета Коммунистической партии Китая по*

Составление 14-го пятилетнего плана национального экономического и социального развития

и долгосрочных целей на 2035[1] год указывает на то, что мы должны ориентироваться на передовые области, такие как искусственный интеллект, реализовать ряд перспективных и стратегических крупных научно-технических проектов и содействовать здоровому развитию цифровой экономики. экономика.

Инвестиции в удовлетворение инновационных потребностей в области искусственного интеллекта продолжают расти.

Содействие развитию ИИ с помощью программ стимулирования и прямых инвестиционных проектов уже является широко распространенной практикой крупных экономик. ЕС продолжает увеличивать финансовую поддержку индустрии искусственного интеллекта и энергично продвигать цифровую трансформацию в Европе. Крупнейший в истории проект ЕС по поддержке исследований и разработок и инноваций, программа "Horizon Europe", имеет общий объем инвестиций в размере 95,5 млрд евро и явно включает ИИ в сферу финансовой поддержки. В апреле 2021 года в форме нормативных актов ЕС использовал "Программу цифровой Европы" для инвестирования в проекты, включая искусственный интеллект, на общую сумму 7,59 миллиарда евро. **Соединенные Штаты считают сохранение своих лидирующих позиций своей стратегической целью и продолжают увеличивать инвестиции в область искусственного интеллекта.** Невоенный бюджет США на ИИ в 2021 году увеличился примерно на 30% и составил в общей сложности 1,5 миллиарда долларов США. Кроме того, в Законе *Соединенных Штатов об инновациях и конкуренции,* Искусственный интеллект, квантовые вычисления и другие области перечислены в качестве приоритетных в бюджете США на исследования и разработки на 2022 финансовый год. В будущем в исследования и разработки в различных областях, включая искусственный интеллект, будет инвестировано в общей сложности 100 миллиардов долларов США. **Великобритания рассматривает инвестиции и планирование экосистемы искусственного интеллекта как долгосрочную стратегию**, запуск национальной программы исследований и инноваций в области искусственного интеллекта и поддержку передовых исследований в области искусственного интеллекта. Согласно статистике, в период с 2014 по 2021 год ее инвестиции в искусственный интеллект превысили 2,3 миллиарда фунтов стерлингов.

Использование приложений для руководства и продвижения внедрения технологий искусственного интеллекта стало консенсусом различных стран. Соединенные Штаты руководят инновациями и комплексным применением технологий искусственного интеллекта в отраслях и секторах. В июле 2021 года Национальный научный фонд США в партнерстве с различными департаментами и известными предприятиями создал 11 новых национальных исследовательских институтов ИИ, охватывающих взаимодействие человека и компьютера, оптимизацию ИИ, динамические системы, обучение с подкреплением и другие

1 Translator's note: For an English translation of the CCP Central Committee Proposal on the 14th Five-Year Plan, see: https://cset.georgetown.edu/publication/proposal-of-the-central-committee-of-the- chinese-communist-party-on-drawing-up-the-14th-five-year-plan-for-national-economic-and-social- development-and-long-range-objectives-for-2030/. For an English translation of the final, authoritative version of China's 14th Five-Year Plan, see: https://cset.georgetown.edu/publication/china-14th-five- year-plan/.

направления исследований. исследовательские проекты охватывают множество областей, таких как строительство, здравоохранение, биология, геология, электричество, образование и энергетика. **Великобритания поддерживает индустриализацию ИИ,** запустив совместный план Управления по искусственному интеллекту и исследований и инноваций Великобритании для обеспечения того, чтобы ИИ приносил пользу всем отраслям и регионам и способствовал широкому применению ИИ. **Япония уделяет особое внимание строительству инфраструктуры и приложениям ИИ,** предлагая ускорить строительство соответствующей инфраструктуры, делая упор на межотраслевые связи. платформы передачи данных и стандарты, связанные с ИИ, всестороннее содействие применению ИИ в различных отраслях, таких как здравоохранение, сельское хозяйство, транспорт и логистика, умные города и производство, а также усиление поддержки малых и средних предприятий (МСП). В набросках 14-го пятилетнего плана Китая[2] четко указано, что основными направлениями будут энергичное развитие индустрии искусственного интеллекта, создание кластеров индустрии искусственного интеллекта и всестороннее расширение возможностей традиционных отраслей. В апреле 2021 года Министерство промышленности и информационных технологий поддержало создание второй партии национальных пилотных зон инноваций и применения искусственного интеллекта в Пекине, Тяньцзине (новый район Биньхай), Ханчжоу, Гуанчжоу и Чэнду и постоянно усиливает руководящую роль приложений. Министерство науки и технологий поддержит строительство ряда пилотных зон инноваций и развития ИИ и последовательно утвердит 15 национальных пилотных зон инноваций и развития ИИ нового поколения в Пекине, Шанхае, Тяньцзине, Шэньчжэне, Ханчжоу и других регионах.

(ii) Искусственный интеллект вступил в новую стадию, когда в центре внимания становится устойчивое и здоровое развитие

С момента рождения искусственного интеллекта в 1956 году связанные с ним теории и технологии продолжали развиваться. Только в последнее десятилетие искусственный интеллект смог по-настоящему перейти от лабораторных исследований к крупномасштабной промышленной практике. Это было связано с прорывами в области глубокого обучения и других алгоритмов, постоянным увеличением вычислительной мощности и непрерывным накоплением массивных данных(海量数据). В процесс промышленного развития и расширения прав и возможностей, с одной стороны, в большом количестве практических сценариев можно увидеть путь развития от "удобного в использовании" к "простому в использовании". Это неотделимо от непрерывного совершенствования самой технологии, непрерывной оптимизации инженерных реализаций, а также поддержки и гарантий, предоставляемых системами управления. С другой стороны, в связи с выявлением различных рисков и проблем в приложениях ИИ и постоянным углублением понимания людьми ИИ, управление ИИ стало темой, привлекающей повышенное внимание со стороны всех секторов по всему миру, и голоса, призывающие к надежности и безопасности, продолжают расти.

2 Translator's note: CSET's English translation of China's 14th Five-Year Plan Outline is available online at: https://cset.georgetown.edu/publication/china-14th-five-year-plan/.

Помимо придания большого значения технологическим инновациям в области искусственного интеллекта в будущем, мы также должны уделять больше внимания инженерным практикам, надежности и безопасности. Это также представляет собой новую "3D" разработку

координирует, выводя индустрию технологий искусственного интеллекта на новый этап. Фактически, усилия отрасли в различных областях уже начались и никогда не прекращались, но сегодня инженерные практики, надежность и безопасность стали более важными. 3D-координаты не являются полностью независимыми, но взаимосвязаны и взаимно поддерживают друг друга. На рисунке 1 показана схема эволюции текущей волны ИИ по различным направлениям и описан контекст развития по каждой координате.

Технологические инновации, инженерные практики, надежность и безопасность стали новыми координатами для "3D" разработки ИИ

Постоянное совершенствование автоматизированных систем

I Автоматическое управление, ориентированное на искусственный интеллект, и
Системы ввода-вывода и управления, такие как MLOP, становятся все более

Устойчивое улучшение-вычислительной мощности в

Г NVIDIA A100, Cambrian SiYuan 370, [1] и другие чипы в 2-3 раза мощнее, чем продукты предыдущего поколения

Постепенное принятие концепции доверия

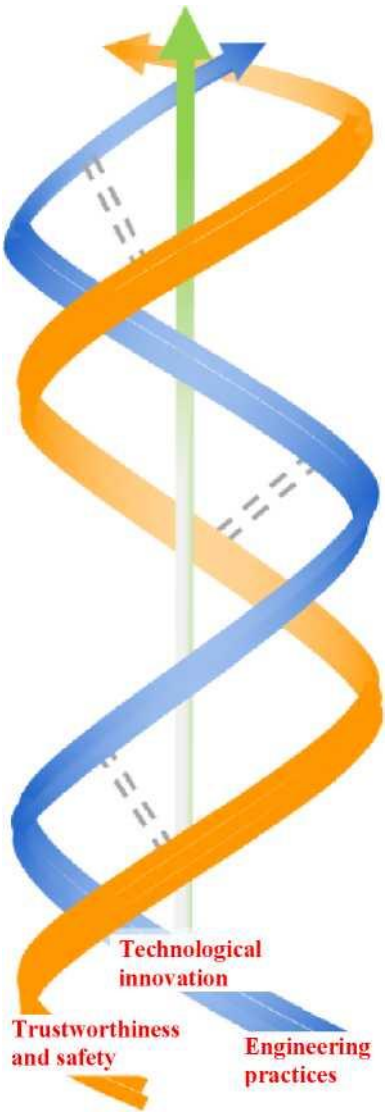
Впервые предложено академическими кругами. затем эта концепция была широко принята международными организациями и правительствами, которые привержены делу содействия развитию надежного ИИ.

Постоянное совершенствование

I Программные фреймворки для глубокого обучения | I различных предприятий были открыты один за другим, с j синергетический эффект между

Взрыв технологии глубокого обучения

I Технические характеристики компьютера
I зрение, разумная речь, естественная Обработка J-языка и другие области



Непрерывное появление гипермасштабных

I Начиная с GPT-3, последовательность гипермасштабных |
Я предварительно обучил модели, такие как Пангу и Я

Отрасли промышленности

Промышленные круги продвигают | I создание заслуживающего доверия ИИ стандарты и компании по всему миру j] мир запустил надежные инструменты, связанные с искусственным

Набор инструментов полного цикла постепенно

I Набор инструментов с полным жизненным циклом становится [1] все более полный, с акцентом на | таких этапах, как маркировка данных, очистка, I разработка и обучение модели, [1]

Диверсификация вычислительных

[От графического процессора до ASIC, FPGA и ! "исследование других чипов, от чипов DNN до нейроморфных чипов J, продолжается.

Стремление к технологическим инновациям

Сосредоточьтесь на инженерных практиках

Обратите внимание на надежность и безопасность

Рисунок 1 Схема трех измерений эволюции ИИ

Стремление к технологическим инновациям в конкретных сценариях всегда было целью и движущей силой развития искусственного интеллекта. Взрыв алгоритмов, представленных глубоким обучением, приоткрыл завесу над приливной волной искусственного интеллекта. Искусственный интеллект в настоящее время широко используется в компьютерном зрении, интеллектуальной речи, обработке естественного языка и других областях, где он последовательно превзошел уровень человеческого распознавания. Диверсификация вычислительных мощностей ИИ и постоянное совершенствование вычислительных мощностей в одной точке(单点算力)обеспечили мощную поддержку развитию ИИ.
В последнее время в Китае и за рубежом часто появляются гипермасштабируемые предварительно обученные модели, постоянно обновляющие рейтинговый список областей применения. В будущем изменения в алгоритмах и вычислительных мощностях будут продолжаться, закладывая фундамент для эры более высокого интеллекта.

Инженерные практики и возможности все чаще становятся важной поддержкой для высвобождения дивидендов от технологии искусственного интеллекта. Усилия в области инженерных практик можно проследить до появления фреймворков с открытым исходным кодом, таких как Caffe, TensorFlow и PaddlePaddle. Защищая детали базового оборудования и операционной системы, эти фреймворки значительно снижают сложность разработки и развертывания моделей, эффективно способствуя распространению технологий искусственного интеллекта. В настоящее время интеграция ИИ с поддерживающими технологиями, такими как облачные вычисления и большие данные, продолжает углубляться, и цепочки инструментов, сосредоточенные на различных связях, таких как обработка данных, обучение модели, развертывание и операции, а также мониторинг безопасности, постоянно расширяются. Система управления исследованиями и разработками в области искусственного интеллекта становится все более совершенной, а технологии автоматизированных операций и технического обслуживания (O & M), представленной MLOps, уделяется все больше и больше внимания. С постоянным совершенствованием инженерных практик и возможностей метод расширения возможностей "малых мастерских и проектной системы"(“小作坊、项目制”)становится все более популярным.
прошлое. В будущем реализация приложений с искусственным интеллектом и доставка продуктов станут более удобными и эффективными.

Надежность и безопасность постепенно стали незаменимыми гарантиями в процессе расширения возможностей ИИ. Заслуживающий доверия ИИ был впервые предложен академическими кругами. В последние годы исследования заслуживающего доверия ИИ, сосредоточенные на безопасности, стабильности, объяснимости, защите конфиденциальности и справедливости, продолжают набирать обороты. Концепция заслуживающего доверия искусственного интеллекта получила широкое внимание со стороны международных организаций. "Принципы ИИ G20", предложенные Группой двадцати (G20) в июне 2019 года, четко предлагают содействовать развитию инноваций в области заслуживающего доверия ИИ, что стало важным консенсусом. Концепция надежного ИИ постепенно внедрялась на протяжении всего жизненного цикла ИИ, а промышленные практики постоянно обогащались. Это превратилось в важную

методологию для реализации соответствующих требований управления ИИ.

В целом, ИИ вступает в **новую стадию "развития инноваций, углубления приложений, разработки норм".** С точки зрения индустриализации самого ИИ, итеративные технологические обновления являются источником развития. В настоящее время ИИ несовершенен, путь интеллектуализации(智能化)все еще изучается, и инновационный характер технологий поможет открыть новые возможности для развития. С точки зрения расширения возможностей традиционных отраслей промышленности с помощью искусственного интеллекта, особенно после пандемии, преобразования в области цифровизации и интеллектуализации были ускорение, вывод приложений искусственного интеллекта на ускоренный путь, в то время как связанные приложения продолжают развиваться. С точки зрения управления, технологии и промышленное развитие обязательно опережают нормативные акты и системы. Проблемы управления становятся все более и более серьезными, и обеспечение здорового развития ИИ стало глобальной проблемой. Здесь происходят как постепенные изменения, так и структурные и даже направленные корректировки. Необходимо всесторонне и систематически улучшать возможности во всех отношениях, чтобы способствовать устойчивому и здоровому развитию ИИ.

II. Технологии и приложения искусственного интеллекта продолжают развиваться по трем направлениям: "Инновации, инженерия и надежность".

В новых условиях технология искусственного интеллекта также должна адаптироваться к новым изменениям. В этой главе рассматриваются тенденции развития технологий и приложений искусственного интеллекта в соответствии с новыми 3D-координатами. Технологические инновации, сосредоточенные вокруг алгоритмов, вычислений и данных, всегда являются доминирующей темой прогресса. Соответствующие технологии в инженерных практиках начали охватывать весь процесс ИИ, ускоряя крупномасштабные приложения ИИ. Надежная технология искусственного интеллекта является важной поддержкой в решении проблем управления и привлекает все большее внимание со стороны всех секторов.

(i) Искусственный интеллект продолжает совершать прорывы в стремлении к экстремальным инновациям

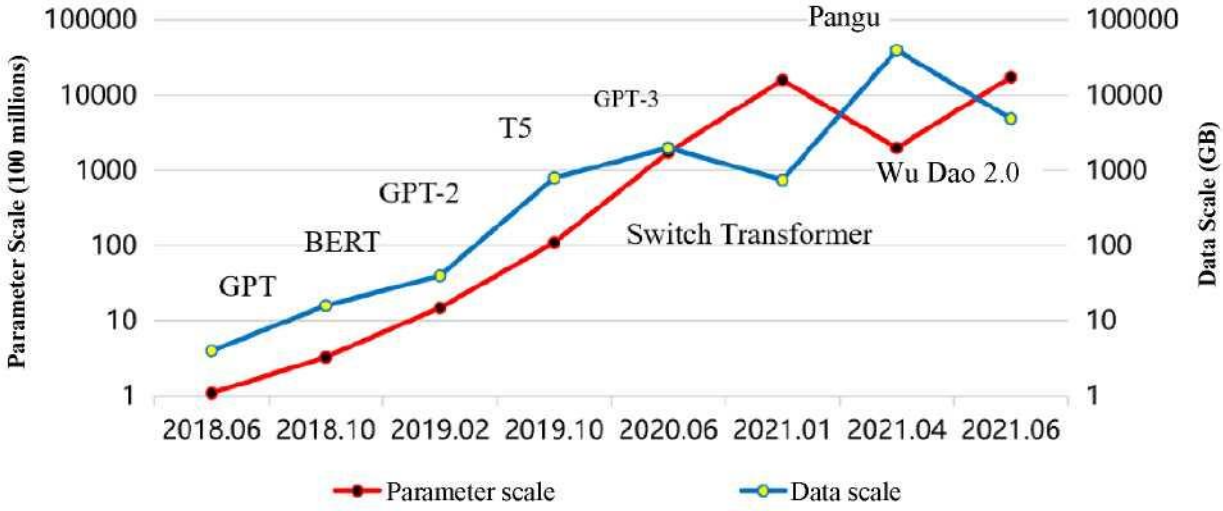
Алгоритмы, вычисления и данные всегда считались тремя лошадьми, которые тянут развитие ИИ, и важной основой для продвижения развития ИИ. **На уровне** алгоритмов гипермасштабные предварительно обученные модели стали одной из самых горячих тем за последние два года, и списки ранжирования в различных областях постоянно обновляются. Исследования в области искусственного интеллекта, основанного на знаниях, и других направлениях стали важным средством изучения улучшения когнитивных способностей. Интеграция и инновации искусственного интеллекта и различных областей научных исследований привлекают все большее внимание, и искусственный интеллект стал важным инструментом в фундаментальных научных исследованиях. **На уровне базовой вычислительной мощности** количество одноточечных вычислений продолжает расти, а кастомизация и диверсификация вычислительной мощности стали важной тенденцией развития. Вычислительные технологии, сосредоточенные вокруг трех основных возможностей обработки данных, хранения данных и взаимодействия с данными, развивались и совершенствовались, продолжая исследования в таких

областях, как нейроморфные чипы и квантовые вычисления. **На уровне данных** технология искусственного интеллекта, представленная глубоким обучением, требует большого объема помеченных данных. Это также породило специализированные технологии и даже сервисы. С постоянным повышением специфичности и глубины ориентации на проблемы службы передачи данных стали более совершенными и индивидуальными. Кроме того, поскольку важность знаний в ИИ широко обсуждается, создание и использование наборов знаний(知识集) продолжает увеличиваться.

1. Постоянно появляются новые алгоритмы, и интеграция технологий представляет собой важную тенденцию

Гипермасштабные предварительно обученные модели способствуют постоянному улучшению технических характеристик и продолжают развиваться в направлении увеличения масштабов и расширения возможностей. С тех пор как OpenAI запустил GPT-3 в 2020 году, такие компании и исследовательские институты, как Google, Huawei, Пекинская академия искусственного интеллекта (BAAI), Китайская академия наук (CAS) и Alibaba, последовательно запустили гипермасштабируемые предварительно обученные модели, в том числе Switch Transformer, DALL-E, MT-NLG, Pangu, Wu Dao 2.0, Zidong Taichu(紫东太初)и M6, постоянно обновляющие различные рейтинговые списки. Общий общий балл оценки понимания языка (GLUE) модели ERNIE 3.0 от Baidu превысил 90% по заданиям на понимание естественного языка. По сравнению с моделью CLIP от OpenAI оценка набора данных изображений и текста из нескольких источников (RUC-CAS-wenlan) модели Wu Dao - Wen Lan от BAAI показала значительное улучшение на 37,0%. В настоящее время количество предварительно обученных параметров модели и масштаб обучающих данных увеличиваются со скоростью 300x / год, продолжая краткосрочное направление эволюции увеличения размера моделей и увеличения масштаба обучающих данных. Кроме того, кросс-модальная предварительная подготовка больших моделей становится все более распространенной. Это привело к переходу от изучения только текстовых данных в первые дни к совместному изучению текста и изображений, и теперь стало возможным обрабатывать трехмерные данные с текстом, изображениями и речью. В будущем появятся предварительно обученные модели, которые используют больше кодировок изображений, больше языков и больше типов данных. Это также будет полезно для изучения обобщения искусственного интеллекта.

Быстрый рост количества параметров и масштаба обучающих данных больших моделей



Источник: CAICT

Рисунок 2 Схема роста больших параметров модели и масштаба обучающих данных

Легкая технология глубокого обучения продолжает изучаться, и эффективность вычислений значительно возросла. Сложные модели глубокого обучения часто потребляют большой объем дискового пространства и вычислительных ресурсов, и их трудно применять в условиях ограниченных ресурсов, таких как терминалы и пограничные вычисления. Технологии с преимуществами малой памяти и малыми вычислительными затратами стали отраслевой необходимостью. Облегченное глубокое обучение стало важной технологией для решения этой проблемы. Это включает в себя такие направления исследований, как проектирование более компактных и эффективных структур нейронных сетей, обрезка больших моделей (то есть "обрезка" части структуры модели) и квантование параметров сети для уменьшения объема вычислений. Например, MobileNet, предложенные Google, и ShuffleNet, предложенные Megvii, стали типичными представителями компактных моделей. Облегченная модель PaddleOCR, выпущенная Baidu, была уменьшена в размере до 2,8 МБ и стала популярной после того, как была доступна с открытым исходным кодом на GitHub.

Технология "Генеративного ИИ" продолжает развиваться, и в будущем способности слушать, говорить, читать и писать будут органично сочетаться. В настоящее время технология "генеративного ИИ" широко используется в интеллектуальном письме, генерации кода, чтении аудио, новостных трансляциях, голосовой навигации, восстановлении изображений и других областях. Автоматизированный синтез текста, речи, изображений, видео и других данных с помощью машин является движущей силой революции в производстве цифрового контента в Интернете. Органичное сочетание навыков аудирования, говорения, чтения, письма и других способностей стало тенденцией развития будущего. Например, CCTV, Информационное агентство Синьхуа и Gmw.cn запустили digital human anchors, которые поддерживают создание видео в один клик из аудио / текстового контента и могут обеспечить быстрое и автоматическое создание программного контента. Соответствующие цифровые ведущие и цифровые репортеры-люди широко использовались для крупномасштабных репортажей и программ, таких как две сессии[3] и

3 Translator's note: The "two sessions" (两会) are the annual full sessions of the National People's

гала-концерт Весеннего фестиваля.

Вычисление знаний стало важной областью исследований для содействия преобразованию ИИ из перцептивного интеллекта в когнитивный интеллект. Знания концентрируют человеческую мудрость, а двойные движущие силы знаний и данных помогают решить проблему вывода и принятия решений в условиях неполной информации, неопределенности и динамичной среды. Это может повысить уровень интеллекта систем искусственного интеллекта. В настоящее время сосредоточены на приобретении знаний, моделировании знаний, управлении знаниями, применении знаний и других процессов, технологий, охватывающих графы знаний, базы знаний и графовые вычисления, уже сформированы. Эта система, охватывающая возможности представления знаний, вычисления знаний, вывода знаний и принятия решений, может обеспечить управление знаниями и их использование. Как академические, так и промышленные круги начали запускать основанные на знаниях прикладные платформы или решения для искусственного интеллекта. Например, Университет Цинхуа, Чжэцзянский университет, Huawei Cloud, BAAI, Baidu, Emotibot и Gridsum запустили такие решения, как механизмы вычисления знаний, промежуточные площадки знаний (知识中台), платформы разработки знаний и платформы для анализа знаний. В дальнейшем knowledge computing будет сосредоточена на внедрении предварительных знаний в алгоритмы глубокого обучения для построения объяснимых моделей. Таким образом, знания могут быть глубоко задействованы в решении моделей, что еще больше повышает эффективность и качество, а также надежность, объяснимость и возможность передачи ИИ.

Интеграция искусственного интеллекта и научных исследований продолжает углубляться, что начало "подрывать" традиционную исследовательскую парадигму. В последние годы способность ИИ анализировать огромные объемы данных освободила исследователей от необходимости ограничиваться традиционными исследованиями в стиле "теоремы вывода" (“推导定理式”). Вместо этого они могут найти
соответствующая информация, основанная на многомерных данных, для ускорения процесса исследования. В 2020 году DeepMind предложила AlphaFold2, которая получила первую премию в номинации "Критическая оценка прогнозирования структуры белка" (CASP). Он может точно предсказывать трехмерную структуру белков с точностью, сравнимой с трехмерной структурой, решаемой с помощью экспериментальных методов, таких как электронная криомикроскопия. Китайско-американская исследовательская группа использовала методы искусственного интеллекта для увеличения предела молекулярной динамики [моделирования] на несколько порядков, обеспечивая при этом высокую точность "ab initio". По сравнению с аналогичной работой в прошлом масштаб вычислительного пространства был увеличен в 100 раз, а скорость вычислений была ускорена в 1000 раз. Команда выиграла премию ACM Гордона Белла 2020 года. Что еще более удивительно, продолжают появляться интегрированные исследования ИИ с механикой, химией, материаловедением, биологией и даже инженерными областями, а глубина и широта применения ИИ будут продолжать расширяться в будущем.

Congress (NPC; 全国人民代表大会; 全国人大), China's parliament, and the National Committee of the Chinese People's Political Consultative Conference (CPPCC; 中国人民政治协商会议全国委员会; 全国政协), an advisory body, which are held concurrently each year in March.

2. Продолжаются прорывы в области одноточечных вычислительных мощностей, а новые технологии все еще находятся на стадии исследования

В настоящее время продолжаются прорывы в области вычислительной мощности искусственного интеллекта, а чипы для обучения и вывода по-прежнему быстро развиваются. В основном это обусловлено спросом на вычислительные ресурсы. С одной стороны, это отражается на этапе обучения модели. Согласно данным OpenAI, темпы роста вычислительной мощности моделей намного превышают темпы роста вычислительной мощности аппаратного обеспечения искусственного интеллекта с разрывом в 10 000 раз между ними. С другой стороны, из-за повсеместного распространения вывода спрос на вычислительные мощности для вывода продолжает расти. В то же время продолжаются исследования по новым

архитектуры вычислительной мощности. Нейроморфные чипы, вычисления в памяти и квантовые вычисления привлекли много внимания, но они, как правило, находятся на стадии исследования.

Инновации в обучающих чипах ускорились, и чипы вывода развиваются в направлении специальной настройки. Количество обучающих чипов на базе графических процессоров продолжает расти, и компании, ориентированные на инновации в области графических процессоров, начали использовать свои возможности. Появилась группа стартапов, специализирующихся на графическом процессоре, в том числе Moore Threads, Iluvatar CoreX и Biren Technology. **Возможности облачных обучающих чипов, основанных на ASIC и других архитектурах, значительно улучшились.** SiYuan 370 от Cambricon, DTU 2.0 i от Enflame Technology(遨 思 2.0) и Kunlun II от Baidu увеличили вычислительную мощность в 3-4 раза по сравнению с предыдущим поколением. **Специализированные пользовательские чипы сквозного вывода появляются повсюду, и смарт-чипы для приложений для мобильных телефонов стали изюминкой.** В январе 2021 года MediaTek запустила высокопроизводительный чип для мобильных телефонов Dimensity 1200, который может обрабатывать данные 5G, AI и изображения на границе. В августе Google запустила свой первый смартфон с чипом Tensor, эксклюзивно предназначенный для линейки телефонов Pixel.

Нейроморфные чипы, вычисления в памяти и квантовые вычисления по-прежнему остаются ключевыми направлениями исследований. Нейроморфные чипы, вычисления в памяти, квантовые вычисления и другие технологии могут обеспечить преимущества высокой вычислительной мощности и низкого энергопотребления на теоретическом уровне. Хотя был достигнут определенный прогресс, в целом их нынешняя технологическая зрелость относительно низка. Центр чипов, вдохновленных мозгом, Пекинского университета объявил о таких достижениях, как "Интеллектуальный IoT-чип со сверхнизким энергопотреблением (AIoT)" на ISSCC в 2021 году. Новые чипы искусственного интеллекта пользуются благосклонностью инвестиционных фондов. С 2021 года многие компании завершили раунд финансирования или раунд A + на сотни миллионов китайских юаней юаней (юаней), в том числе производитель чипов 3D vision Aivatech, Reexen, который специализируется на исследованиях и разработках интегрированных чипов для хранения и вычисления нейроморфных датчиков, а также на исследованиях и разработках чипов AI vision компания Axera (爱芯科技).

3. Продолжающееся увеличение масштаба данных создает горячую точку для

интеграции знаний предметной области

Стремительное развитие искусственного интеллекта способствует постоянному увеличению объема данных. По оценкам IDC, глобальный объем данных достигнет 163 ЗБ в 2025 году, при этом на долю неструктурированных данных придется 80-90%. Службы передачи данных вступили в стадию глубокой настройки. Baidu, Alibaba, JD и другие компании запустили сервисы настройки данных, основанные на различных сценариях и потребностях. Наборы данных, необходимые предприятиям, переходят от простых сценариев общего назначения к персонализированным сложным сценариям, таким как прогрессирование наборов данных распознавания речи от

Мандарин на менее распространенные языки, диалекты и другие сценарии, а также развитие наборов данных интеллектуального диалога от таких сценариев, как вопросы и ответы с кратким ответом и голосовое управление, до сценариев приложений и бизнес-вопросов и ответов.

Все стороны активно изучают возможность создания высококачественных наборов знаний для поддержки будущей разработки приложений ИИ, основанных на знаниях. Набор знаний содержит традиционные данные, такие как голос, изображение и текст, а также определения, правила, логические взаимосвязи и другие данные. Это представление знаний, основанное на данных. Хорошо известные наборы знаний в промышленности включают Wordnet и Hownet. Например, Alibaba и Гонконгский политехнический университет разработали набор знаний FashionAI, основанный на знаниях в области дизайна одежды, что ускорило применение ИИ в индустрии дизайна одежды.

(ii) Набор инструментов искусственного интеллекта стал основой инженерных практик и возможностей

Благодаря непрерывному развитию технологий искусственного интеллекта в последние годы число инженерных приложений ускорилось. В финансовой сфере технологии искусственного интеллекта начали глубоко проникать во все процессы фронт-, мидл- и бэк-офисов. Медицинский ИИ начал вступать в стадию маркетизации. По состоянию на конец августа 2021 года в общей сложности 28 продуктов были одобрены для получения сертификатов регистрации медицинских изделий класса III. В связи с быстрым развитием искусственного интеллекта в производственной сфере "Делойт" прогнозирует, что среднегодовые темпы роста Китая в течение следующих пяти лет будут превышать 40%. В настоящее время применение ИИ предприятиями демонстрирует переход от предварительной разведки к крупномасштабному применению. Вообще говоря, постоянное совершенствование инженерных методов и возможностей стало ключом к будущим приложениям.

Разработка искусственного интеллекта начала становиться центром внимания всех секторов. В академических кругах Институт разработки программного обеспечения Университета Карнеги-Меллона в последние годы начал исследования в области искусственного интеллекта и провел национальную исследовательскую программу, финансируемую официальными учреждениями США совместно с университетами и промышленностью. Всемирно известные эксперты в области искусственного интеллекта Майкл И. Джордан и Эрик Син (刑波) считают,

что разработка ИИ - это развивающаяся инженерная дисциплина и тенденция в развитии

ИИ от теоретической дисциплины к инженерной дисциплине. В отраслевых кругах Gartner уже два года подряд называет разработку искусственного интеллекта одним из своих ежегодных стратегических технологических трендов. Alibaba Cloud и другие предприятия рассматривают разработку искусственного интеллекта как ключ к превращению искусственного интеллекта в производительность предприятия.

AI engineering фокусируется на эффективном сочетании процесса полного жизненного цикла инструментальных систем, процессов разработки и управления моделями. На уровне инструментальной системы систематизация и открытость стали развитием

характеристики технологической цепочки инструментов платформы исследований и разработок. Первоначально была создана относительно полная инструментальная система, ориентированная на такие технологии, как машинное обучение и глубокое обучение. Эта система значительно упрощает обработку данных, разработку и развертывание моделей, а также управление и эксплуатацию. Ключевые программные фреймворки в основном используют фреймворки с открытым исходным кодом, такие как TensorFlow, PyTorch, Paddle, MindSpore и OneFlow. На уровне процесса разработки инжиниринг фокусируется на процессе жизненного цикла разработки модели ИИ, обеспечивает эффективное и стандартизированное непрерывное производство, непрерывную доставку и непрерывное развертывание и, наконец, отправляет лучшую модель на уровень приложений для создания ценности для бизнеса. Например, MLOps устанавливает стандартизированный процесс разработки, развертывания и ввода-вывода модели для соединения команды построения модели, бизнес-команды и команды O & M. На уровне управления моделями, с постепенным углублением интеллектуальных приложений для предприятий, типы и количество моделей значительно увеличились. Предприятиям необходимо создать механизм управления жизненным циклом модели и внедрить стандартизированное управление и управление для истории версий модели, производительности, атрибутов, соответствующих данных и производных файлов модели.

Технология автоматизированного машинного обучения (AutoML) является важной возможностью для улучшения инженерных возможностей. Автоматизированное машинное обучение относится к автоматизации всего или части всего процесса разработки и применения машинного обучения. Это может эффективно уменьшить проблемы на текущем этапе развития ИИ, такие как высокий порог для развития ИИ и нехватка технических талантов. Эта технология в основном включает автоматическую предварительную обработку данных, автоматизированное проектирование объектов, автоматизированный поиск гиперпараметров, автоматизированное проектирование структуры сети моделей и автоматизированное развертывание моделей. Такие технологии, как разработка с низким уровнем кодирования и предварительно обученные модели, также тесно связаны с автоматизированным машинным обучением и демонстрируют тенденцию к комплексному развитию. В настоящее время ведущие интернет-компании и инновационные компании начали активно внедрять технологии и инструменты AutoML. Однако, ограниченные зрелостью этой технологии, сценарии применения AutoML по-прежнему ограничены этапами процесса разработки (такими как разработка функций) или некоторыми конкретными техническими областями (такими как распознавание речи, обнаружение объектов или интеллектуальный диалог).

Технические требования к совместному управлению облачными пограничными

терминалами постепенно становятся все более значимыми, и процесс миграции в облако искусственного интеллекта продолжает ускоряться. С глубокой интеграцией искусственного интеллекта в различные отрасли промышленности пограничные и терминальные устройства искусственного интеллекта будут получать все более широкое распространение. В то же время разработчики также столкнутся с проблемами сложной и сложной адаптации периферийных устройств, а также сложного управления и управления. С одной стороны, платформа реализует адаптацию и развертывание модели для периферийных устройств с помощью таких технологий, как сжатие модели и адаптивная генерация модели. С другой стороны, благодаря разработке и настройке оптимизации компиляции и промежуточного представления можно реализовать совместное управление и O & M для облачных, пограничных и терминальных устройств.

Безопасная и заслуживающая доверия технология искусственного интеллекта развивается в направлении интеграции

При постоянном внимании всех секторов к вопросу доверия к ИИ безопасные и надежные технологии ИИ стали горячей областью исследований. Основное внимание в исследованиях уделяется улучшению стабильности, объяснимости, защите конфиденциальности и справедливости систем искусственного интеллекта. Эти технологии составляют основные вспомогательные возможности надежного ИИ.

Техническая направленность стабильности системы искусственного интеллекта постепенно расширилась с цифровой области на физическую. Системы искусственного интеллекта сталкиваются с уникальными атаками, такими как атаки с отравлением, враждебные атаки и бэкдор-атаки, которые усиливают проблемы безопасности. Эти методы атаки могут существовать независимо или одновременно. Например, печатая состязательные образцы очков, злоумышленники могут напрямую создавать физические помехи системам распознавания лиц. Или злоумышленник может вставить на дорожный знак образец враждебного образца возмущения, что заставляет автономную систему вождения ошибочно распознавать знак "стоп" как знак "ограничение скорости". Технология тестирования стабильности, ориентированная на системы искусственного интеллекта, также стала ключевой. Huawei, Baidu и другие компании запустили соответствующие технологии тестирования, основанные на нечеткой логике, и стремятся исследовать и улучшать стабильность систем искусственного интеллекта.

Технология улучшения объяснимости ИИ все еще находится в зачаточном состоянии, и различные пути продолжают изучаться. Повышение объяснимости систем ИИ стало популярной областью работы, и основные пути включают в себя создание соответствующих механизмов визуализации для оценки и интерпретации промежуточных состояний модели; анализ влияния обучающих данных на конечную конвергентную модель ИИ с помощью функции влияния; использование методов для анализа того, какие данные характеризуют модель ИИ используется для прогнозирования; и исследование объяснимости моделей черного ящика путем использования простых интерпретируемых моделей для локальной аппроксимации сложных моделей черного ящика.

Вычислительная технология, обеспечивающая конфиденциальность, способствует безопасному и надежному взаимодействию с данными ИИ. Системы искусственного интеллекта должны полагаться на большой объем данных, но поток данных и сама модель искусственного

интеллекта могут привести к утечке конфиденциальных личных данных. Сочетание искусственного интеллекта и вычислительной технологии, сохраняющей конфиденциальность, может обеспечить подлинность и достоверность необработанных данных из источника данных. Используя вычислительную технологию, обеспечивающую конфиденциальность, данные становятся "доступными и невидимыми", формируя логически централизованное представление физически распределенных многомерных данных. Это может гарантировать, что модели искусственного интеллекта будут иметь достаточные и достоверные данные.

Ключ к повышению справедливости ИИ заключается в том, чтобы начать как с данных, так и с технологий. С широким применением систем искусственного интеллекта такие проблемы, как несправедливое поведение при принятии решений и дискриминация в отношении некоторых групп, становятся все более заметными. Основные причины таких искажений при принятии решений заключаются в следующем: ограниченные условиями сбора данных, веса разных групп в данных несбалансированы; когда модель ИИ обучается на несбалансированном наборе данных, принятие решений моделью становится несправедливым. Чтобы обеспечить справедливость принятия решений в системах искусственного интеллекта, на уровне данных основным методом является построение полностью разнородных наборов данных, чтобы свести к минимуму присущую им дискриминацию и предвзятость в данных; затем наборы данных периодически проверяются для обеспечения высокого качества данных. На технологическом уровне существуют также алгоритмы, которые используют количественные показатели справедливого принятия решений для уменьшения или устранения предвзятости при принятии решений и потенциальной дискриминации.

Систематическое продвижение надежности ИИ и технологий безопасности станет важной тенденцией. С одной стороны, большая часть текущих актуальных исследований проводится с точки зрения одного измерения, такого как стабильность, конфиденциальность или честность. Существующая исследовательская работа показала, что различные требования, такие как стабильность, справедливость и объяснимость, являются взаимно синергетическими или ограничительными. Если учитывается только один аспект требования, это может привести к конфликтам с другими требованиями. Ключевым стало то, как построить системную исследовательскую структуру для поддержания оптимального динамического баланса между различными характерными элементами. С другой стороны, необходимо провести исследование надежности и безопасности на системном уровне. Эта проблема существует не только на уровне алгоритмов искусственного интеллекта. Это также касается всей системы, такой как проблемы безопасности операционной системы, программного обеспечения, сторонних библиотек и аппаратного обеспечения, используемого для запуска ИИ. Необходимо обеспечить надежность и безопасность для полного жизненного цикла ИИ и его цепочки.

Мир очень внимательно относится к управлению ИИ, а безопасность и надежность ИИ стали в центре внимания

Чем больше возможностей для развития ИИ, чем глубже его влияние и чем с большим количеством проблем он сталкивается, тем важнее и срочнее управлять им. В настоящее время в мире сформировалась модель управления, предусматривающая участие различных организаций и скоординированное совместное управление. Страны и организации внедрили ряд принципов управления, был достигнут существенный прогресс в законодательном процессе, а отраслевые

организации и корпоративные структуры активно изучают надежные методы внедрения.

(i) Риски ИИ продолжают возрастать, и на начальном этапе создается глобальный механизм управления

Углубленное расширение возможностей ИИ порождает проблемы

Риски и проблемы, связанные с ИИ, многообразны. В дополнение к естественным недостаткам самой технологии ИИ, в отличие от чисто технических рисков, источником рисков ИИ является влияние приложений систем ИИ на существующие нормативные системы, этику и социальный порядок.

Технические риски, присущие ИИ, продолжают расти. Технология искусственного интеллекта с глубоким обучением в основе постоянно выявляет скрытые риски, вытекающие из ее собственных характеристик. Во-первых, модели глубокого обучения имеют недостатки хрупкости и уязвимости, что затрудняет получение достаточной уверенности в надежности систем искусственного интеллекта. Во-вторых, модели "черного ящика" отличаются высокой степенью сложности и неопределенности, что может легко привести к непредсказуемым рискам. В-третьих, результаты, генерируемые алгоритмами искусственного интеллекта, чрезмерно зависят от обучающих данных. Если в данных обучения присутствует предвзятость и дискриминация, это приведет к несправедливости при принятии разумных решений.

Вызовы существующим правовым и нормативным системам продолжают расширяться. ИИ по-разному влиял на правовые и нормативные системы: с точки зрения определения квалификации юридического лица предоставление Саудовской Аравией гражданства роботу Софии вызвало глобальные споры. Кроме того, также возникли такие вопросы, как может ли ИИ стать изобретателем патента. Например, в июле 2021 года Федеральный суд Австралии постановил, что системы искусственного интеллекта могут быть указаны в качестве изобретателей в патентных заявках, что совершенно противоположно позиции Соединенных Штатов и Соединенного Королевства. С точки зрения защиты конфиденциальности, разработка ИИ сопровождается нарушением личной неприкосновенности частной жизни. Партия Центрального телевидения Китая (CCTV) "15 марта" была разоблачена, и большое количество компаний незаконно собирали информацию о лицах клиентов для использования в коммерческих целях. Что касается разделения обязанностей, то в 2015 году первая роботизированная операция в Великобритании привела к смерти, а инцидент с "потерей контроля над дверью" Tesla поставил под сомнение автоматизированную систему помощи при вождении.

Воздействие на этику и общественный порядок становится все более серьезным. ИИ рискует повлиять на права человека. ИИ вызывает дискриминацию, предлагает новые правила поведения людей и вносит изменения в рабочую силу. В августе 2021 года российский сервис онлайн-платежей Xsolla использовал алгоритм для оценки того, были ли сотрудники "разобщенными и неэффективными", и уволил 147 сотрудников, что составляло треть рабочей силы компании. ИИ прямо или косвенно наносит вред людям и влияет на общественный порядок. В ноябре 2020 года в сообщениях СМИ утверждалось, что иранский ученый-ядерщик был убит оружием, управляемым "искусственным интеллектом". В 2019 году умные колонки Amazon посоветовали кому-то совершить самоубийство.

Возникла глобальная волна управления искусственным интеллектом

В настоящее время, перед лицом различных рисков и проблем, возникающих в результате углубленного расширения возможностей ИИ Страны по всему миру уделяют все больше и больше внимания управлению ИИ. Управление искусственным интеллектом - это сложный системный проект. Согласно Белой книге по управлению искусственным интеллектом,, система управления искусственным интеллектом состоит из совместного участия и сотрудничества нескольких субъектов, таких как правительство, отраслевые организации, предприятия и общественность. Это формирует метод управления, сочетающий "мягкие законы", такие как этические принципы, и "жесткие законы", такие как законы и правила, который направлен на реализацию общей цели и видения науки и техники во благо и на пользу человечеству, а также на содействие здоровому и упорядоченному развитию ИИ. На рисунке 3 показана схема механизмов управления ИИ.

Данные, собранные Китайской академией информационных и коммуникационных технологий (CAICT)

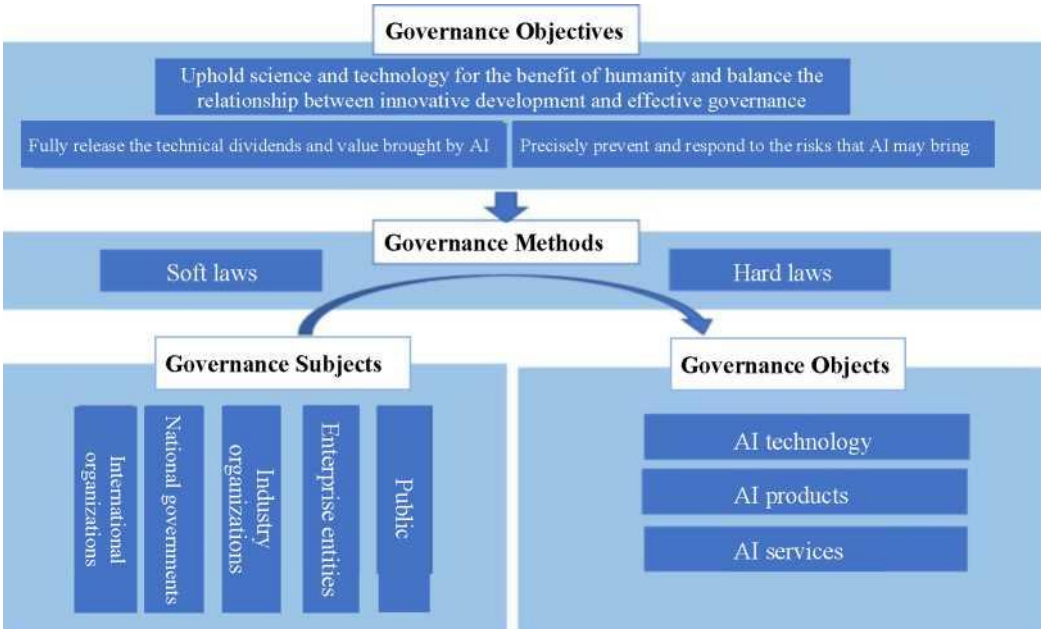


Рисунок 3 Схема механизмов управления ИИ

Крупнейшие экономики мира сосредоточены на обсуждении ключевых вопросов управления искусственным интеллектом, и межправительственные международные организации стали важными голосами. Организация Объединенных Наций, G20, ОЭСР и G7 стали важными организациями, руководящими глобальным управлением ИИ. Соответствующие результаты исследований ОЭСР сыграли важную роль в продвижении глобального управления ИИ. Они являются важным ориентиром для резолюций G7, касающихся ИИ, а также Принципов ИИ G20.

Организация **Объединенных Наций** (ООН) активно продвигает процесс этического управления ИИ. Организация Объединенных Наций по вопросам образования, науки и культуры (ЮНЕСКО) опубликовала свою *Рекомендацию об этике искусственного интеллекта* 25 ноября 2021 года. Это первая в мире нормативная база, разработанная для этики ИИ, и самый широкий консенсус, достигнутый на правительственном уровне в мире на сегодняшний день. В то же время это возлагает на каждую страну ответственность за применение рамочной

на соответствующем уровне. Всемирная организация здравоохранения (ВОЗ) опубликовала свое первое руководство по использованию ИИ в здравоохранении, *этике и управлению искусственным интеллектом в интересах здравоохранения,* 28 июня 2021 года. Это гарантирует, что технология искусственного интеллекта может служить общественным интересам всех стран по всему миру.

В июне 2019 года **Группа двадцати** (G20), ссылаясь на Принципы *ОЭСР по искусственному интеллекту,* одобрила *Принципы ИИ G20,* которые выступают за использование и развитие ИИ "с уважением правовых принципов, прав человека и демократических ценностей". Это стало первым международным межправительственным консенсусом по управлению искусственным интеллектом и установило концепцию развития, ориентированную на людей. Китай поддерживает укрепление диалога, сосредоточенного вокруг ИИ, реализацию принципов ИИ G20 и содействие здоровому развитию ИИ во всем мире.

22 мая 2019 года **Организация экономического сотрудничества и развития** (ОЭСР) опубликовала первые в мире межправительственные руководящие принципы политики в области ИИ, сформировав Принципы *ОЭСР по искусственному интеллекту.* В нем были установлены пять принципов ответственного управления надежным ИИ. ОЭСР учредила свою Обсерваторию политики в области искусственного интеллекта (OECD.AI) в феврале 2020 года для обмена лучшими примерами политики и практики в области ИИ, содействия международному сотрудничеству и оказания помощи государствам-членам в создании надежных систем ИИ на благо общества в целом.

The **Группа семи** (G7) начала исследование консенсуса в отношении управления ИИ среди стран с развитой экономикой по всему миру. Саммит G7 в январе 2021 года указывает на то, что государства-члены будут сотрудничать в разработке международных стандартов ИИ; в сентябре на заседании органов по защите данных и конфиденциальности G7 они заявили, что защита данных и надзор за конфиденциальностью станут основной задачей управления ИИ в будущем, и призвали промышленность разрабатывать продукты ИИ, отвечающие требованиям защиты данных требования.

(ii) Управление ИИ вступило в новую стадию координации мягкого и жесткого законодательства и регулирования сценариев

С момента публикации Принципов *ИИ Asilomar* в 2017 году произошел глобальный всплеск в изучении и формулировании этических принципов ИИ. В настоящее время Принципы ИИ G20 широко признаны международным сообществом, межправительственные организации стали важной силой, определяющей направление управления ИИ, а страны по всему миру ускоряют совершенствование своих соответствующих систем правил управления ИИ. Первый проект Закона ЕС *об искусственном интеллекте* в 2021 году ознаменовал ускорение перехода управления ИИ от принципиальных ограничений "мягкого закона" к более существенному регулированию "жесткого закона". В то же время, с углублением интеграции между ИИ и реальной экономикой, управление ИИ становится все более ориентированным на конкретные сценарии.

1. Процесс материализации управления ИИ ускорился

В настоящее время фокус управления ИИ в разных странах разный, но в целом он демонстрирует тенденцию к ускоренной эволюции. Иными словами, с самого раннего этапа построения системы социальных норм, основанной на "мягком праве", она начинает двигаться к системе предотвращения рисков и контроля, гарантированной "жестким законом".

ЕС неуклонно продвигается от этики к регулированию и желает взять на себя ведущую роль в глобальных правилах регулирования ИИ. 21 апреля 2021 года Европейский союз опубликовал проект Закона об *искусственном интеллекте.* Это первый в мире закон, систематически регулирующий ИИ. В нем уточняется четырехуровневая структура рисков ИИ, основное внимание уделяется регулированию систем с высоким уровнем риска и предлагаются относительно полные меры нормативной поддержки. Это еще один важный шаг ЕС после публикации Руководства *по этике для заслуживающего доверия ИИ* (2018) и *Белой книги по искусственному интеллекту: европейский подход к совершенству и доверию* (2020). Это знаменует переход глобального управления ИИ от мягких ограничений, таких как этические принципы, к этапу всеобъемлющих и действенных правовых норм.

Соединенные Штаты подчеркивают важность пруденциального надзора для содействия инновационному развитию. Исполнительный указ 2019 *года о сохранении американского лидерства в области искусственного интеллекта* установил, что общий тон Соединенных Штатов в управлении искусственным интеллектом был сосредоточен на укреплении их глобального лидерства. В 2019 году Соединенные Штаты предложили Закон об *алгоритмической подотчетности*, требующий оценки воздействия на автоматизированные системы принятия решений "высокого риска". В 2020 году Сенат США представил *Национальный закон о конфиденциальности биометрической информации,* который обеспечивает защиту конфиденциальности биометрической идентификации с поддержкой искусственного интеллекта на основе защиты персональных данных. В мае 2021 года Закон США об *алгоритмическом правосудии и прозрачности онлайн-платформ* предложил обязательства и требования к прозрачности алгоритмов с точки зрения трех субъектов: пользователей, регулирующих органов и общественности. В июле 2021 года Управление подотчетности правительства США выпустило систему подотчетности ИИ для обеспечения справедливости, надежности, отслеживаемости и управления системами ИИ.

В Китае принимаются во внимание как мягкие, так и жесткие законы, и оба они используются для содействия управлению искусственным интеллектом. На уровне принципов и этики Национальный комитет специалистов по управлению искусственным интеллектом нового поколения, выпустив *в июне 2019 года Принципы управления искусственным интеллектом нового поколения: развитие ответственного искусственного интеллекта*, опубликовал *Этические нормы для искусственного интеллекта нового поколения,* целью которых является интеграция этики в полный жизненный цикл ИИ и активное руководство общество в целом должно ответственно проводить исследования и разработки в области искусственного интеллекта и прикладные мероприятия. **Что касается юридического процесса,** Китай еще не издал единый закон, касающийся ИИ, но *Закон о защите личной информации*,

официально введенный в действие в ноябре 2021 года, вместе с Законом о *кибербезопасности* и Законом*обезопасности данных* образуют прочную правовую систему для регулирования основополагающих элементов ИИ. Кроме того, на местном уровне ведутся активные исследования. *В июле 2021 года в Шэньчжэне были изданы Правила продвижения индустрии искусственного интеллекта Особой экономической зоны Шэньчжэня (проект)*, призванные содействовать здоровому развитию индустрии искусственного интеллекта.

В то же время Великобритания, Франция, Япония, Южная Корея и другие страны также провели работу, связанную с управлением ИИ. Соединенное Королевство уделяет особое внимание разработке норм ИИ и поощряет образование в области ИИ и подготовку талантов. *Искусственный интеллект: возможности и последствия для будущего принятия решений* (2016), *ИИ в Великобритании: готов, желает и способен?* (2018), Хартия *новых технологий* (2021) и многие другие документы и отчеты призывают к созданию кодекса ИИ и этических рамок на национальном уровне. Франция углубила свое понимание этических проблем ИИ с помощью семинаров экспертов и академических дебатов. Япония, Южная Корея и другие страны внимательно относятся к этике ИИ с точки зрения развития интеллектуализированной трансформации производства и применения новых технологий.

2. В типичном управлении, основанном на сценариях, каждый из них ускоряет реализацию со своей собственной направленностью

Сложность управления искусственным интеллектом также отражается в диверсификации и дифференциации сценариев его применения. В разных сценариях глубина применения и влияние технологии искусственного интеллекта различаются. Управление типичными сценариями стало предметом работы в разных странах, особенно в таких областях, как автономное вождение, интеллектуальное здравоохранение и распознавание лиц.

В области автономного вождения Германия взяла на себя ведущую роль в разработке этических принципов и рамочных законов, а различные страны активизировали внедрение дифференцированного и категоризированного分级分类надзора ().

Германия представила свой *Этический кодекс автоматизированного и подключенного вождения* в 2017 году и приняла проект *Закона об автономном вождении* в мае 2021 года. В 2021 году Великобритания обсудила и внесла *поправки в Закон* о автомобильных дорогах, чтобы ввести новые положения о безопасном использовании автономных транспортных средств на автомагистралях. В Китае в мае 2021 года Администрация киберпространства Китая совместно с соответствующими ведомствами разработала *несколько положений об управлении безопасностью автомобильных данных (проект для комментариев)*, чтобы запросить общественное мнение.

В области интеллектуального здравоохранения постепенно развиваются этические принципы, и нормативный уровень сосредоточен на регулировании доступа к медицинскому оборудованию. Основываясь на *Предлагаемой нормативно-правовой базе 2019 года для внесения изменений в программное обеспечение на основе искусственного интеллекта*

/ машинного обучения (AI / ML) как медицинское устройство (SaMD) (Дискуссионный документ), Управление по САНИТАРНОМУ НАДЗОРУ ЗА КАЧЕСТВОМ ПИЩЕВЫХ ПРОДУКТОВ И медикаментов США выпустило *Искусственный интеллект / машинное обучение -*

Программное обеспечение на основе программного обеспечения как план действий в области медицинского оборудования в январе 2021 года по внедрению нормативных инициатив в отношении программного обеспечения для медицинского оборудования на основе Al. Европейский союз ввел Регламент по медицинскому оборудованию (MDR), требующий, чтобы новые медицинские устройства подавали заявки на получение сертификата соответствия, начиная с мая 2021 года. В июне 2021 года Китай опубликовал *Руководящие принципы регистрации и проверки медицинских устройств с искусственным интеллектом (Проект для обратной связи)* (人工智能医疗器械注册审查指导原则 (征求意见稿)) и продвигал упорядоченное развитие индустрии медицинского оборудования с искусственным интеллектом.

В области распознавания лиц страны по всему миру вступили в эпоху строгого надзора за защитой конфиденциальности и безопасностью информации и данных. Европейский союз включил распознавание лиц в категорию категорий высокого риска в проект *Закона об искусственном интеллекте*, представленный в апреле 2021 года. В октябре Европейский парламент проголосовал за принятие резолюции, призывающей к полному запрету крупномасштабной слежки, основанной на биометрических технологиях искусственного интеллекта. Китай ввел *в действие Закон о защите личной информации* в ноябре 2021 года, а судебное толкование, касающееся распознавания лиц, принятое Верховным народным судом в августе, конкретно регулирует обработку информации о лицах. законодательство США на уровне штатов и на местном уровне запрещает правительственным учреждениям использовать технологию распознавания лиц в общественных местах. Соединенное Королевство опубликовало Хартию *новых технологий* в сентябре 2021 года, указав, что такие технологии, как распознавание лиц, должны использоваться законно и этично.

(iii) Системы обеспечения безопасности ИИ стали ключевым ориентиром для эффективного предотвращения рисков

Для эффективного предотвращения рисков безопасности, связанных с применением технологий искусственного интеллекта, и обеспечения безопасности систем искусственного интеллекта, связанных с национальной безопасностью, экономической жизнью и социальной стабильностью, необходимо срочно предложить систему безопасности для систем искусственного интеллекта и обеспечить эффективное руководство для промышленности, чтобы постепенно улучшите возможности обеспечения безопасности искусственного интеллекта. Структуры безопасности ИИ основаны на потребностях защиты безопасности ИИ и органично интегрируют технологические системы безопасности ИИ и системы управления безопасностью ИИ. Общий системный дизайн и планирование построенной системы безопасности ИИ имеют большое значение для поддержания национальной безопасности ИИ и кибербезопасности.

1. Системы безопасности искусственного интеллекта постепенно обретают

форму

Структура безопасности ИИ должна включать четыре измерения: цели безопасности, возможности безопасности, технологии безопасности и управление безопасностью, как показано на рисунке 4. Эти четыре аспекта защиты помогают предприятиям создавать системы защиты безопасности ИИ, основанные на нисходящем, послойном подходе. В этих усилиях постановка разумных целей в области безопасности является отправной точкой и основой для обеспечения

безопасность приложений ИИ, возможности безопасности обеспечивают эффективную гарантию достижения целей безопасности, а технологии безопасности и управление безопасностью являются основой и воплощением возможностей безопасности.

Система безопасности искусственного интеллекта

AI Security
Управление безопасностью искусственного интеллекта

Системы искусственного интеллекта

Распознавание лиц, распознавание голоса автономного вождения...

Учреждения

Комитеты по безопасности ИИ, специализированные агентства по управлению безопасностью ИИ, межведомственные виртуальные координационные агентства по безопасности ИИ

Требования

Цели безопасности ИИ

Приложения легальны и соответствуют требованиям, функции стабильны и надежны, поведение контролируется вручную
, данные безопасны и надежны, процесс принятия решений прозрачен и справедлив, результаты могут быть
объяснены, инциденты могут быть отслежены

Заверения

Секретные возможности безопасности ИИ

Контратака

Возможности обеспечения безопасности для принятия правовых контрмер против вредоносные злоумышленники с искусственным интеллектом

Legal rights
Защита

Intelligence
Управление разведывательной информацией, поглощение разведывательной

Информация об угрозах

Возможность получать и использовать информацию об угрозах безопасности ИИ для расширения возможностей систем безопасности ИИ, устройств и людей

Активная безопасность

Внедряйте и укрепляйте возможности групп безопасности искусственного интеллекта для реализации динамичных, самоадаптирующихся и саморазвивающихся возможностей безопасности

Непрерывный мониторинг безопасности, анализ инцидентов безопасности
, реагирование на защиту безопасности,
прогнозирование угроз

Пассивная оборона

Развертывание статических, пассивных возможностей безопасности за пределами приложений искусственного теллекта

Оценка безопасности, обнаружение атак безопасности,
защита безопасности алгоритма, защита безопасности данных
protection

Безопасность архитектуры

Способность планировать, проектировать, создавать и использовать приложения искусственного интеллекта с учетом соображений безопасности для повышения их внутренней безопасности

**Оценка соответствия требованиям, механизмы безопасности бизнеса, повышение безопасности алгоритмов
, повышение безопасности данных, проверка
и исправление безопасности модели/структуры**

Поддерживает

Технология безопасности искусственного интеллекта

Безопасность бизнеса

Оценка безопасности бизнеса
Механизмы обеспечения безопасности бизнеса
Оценка безопасности бизнеса
Мониторинг бизнес-рисков
Записи в журнале

Безопасность алгоритма

Состязательность алгоритма оценка
Надежность алгоритма оценка
Надежность алгоритма улучшение
Конфиденциальность алгоритма уверенность
Алгоритм обеспечения безопасности архитектурный дизайн
Оценка объяснимости алгоритма
Конфиденциальность алгоритма оценка
Обеспечение справедливости алгоритма
Объяснимость алгоритма улучшение

Персонал

Персонал по управлению безопасностью ИИ
O & M персонал службы безопасности искусственного интеллекта
Персонал, принимающий решения в системе искусственного интеллекта, сотрудники службы безопасности искусственного интеллекта
Технический персонал службы безопасности ИИ
Персонал аудита безопасности ИИ
Контактные лица службы безопасности ИИ

Безопасность платформы

Тестирование безопасности платформы Framework
Оценка полноты модели
Развертывание безопасности платформы Framework platform
Исправлены ошибки платформы Framework Безопасность цепочки поставок платформы Framework

Безопасность данных

Проверка соответствия данных
Проверка достоверности данных
Отслеживание источников данных
Справедливость данных улучшение
Мониторинг рисков, связанных с данными
Оценка безопасности утечки данных
Сохранение конфиденциальности данных вычисление
Проблема очистки данных
Безопасное удаление данных

Управленческие меры

Технические меры для поддержки мониторинга и обработки инцидентов безопасности
Механизмы сотрудничества в области регулирования
Механизмы отчетности и принятия
Безопасность данных и защита личной информации
Обеспечение прозрачности системы
Оценка рисков безопасности вмешательства человека
Обеспечение прозрачности системы
Отчеты о поведении вредоносного приложения
Сторонние аудиты безопасности

Источник: CAICT

Рисунок 4 Структура безопасности искусственного интеллекта

2. Категоризация и классификация стали новым направлением построения фреймворков

Категоризация и классификация стали новой тенденцией в глобальном управлении искусственным интеллектом. В проекте Закона ЕС *об искусственном интеллекте, Законе,* США *об ответственности за алгоритмы,* Canada's , *Директиве Канады об автоматизированном принятии решений,* и *руководящих мнениях Китая об усилении общего управления алгоритмами информационных служб Интернета,* а также в других законах, нормативных актах и программных документах крупных стран по всему миру предлагается установить требования к категоризированное и градуированное управление системами и алгоритмами искусственного интеллекта. Однако вышеупомянутые законы и нормативные акты предлагают только категоризацию и градуированные требования к управлению или описывают методы категоризации путем перечисления типичных систем искусственного интеллекта. В них отсутствуют принципы категоризации, нет метода или процесса оценки, которым можно было бы следовать, и они не могут быть применены к быстро появляющимся новым приложениям ИИ. Необходимо срочно предложить систему категоризации и классификации ИИ, уточнить принципы категоризации и классификации, а также упростить элементы и методы категоризации и классификации для реальных операций.

В соответствии с идеей категоризированного управления и градуированной защиты, в этом Техническом документе предлагаются следующие рекомендации по категоризации и классификации ИИ. В зависимости от степени автономности системы искусственного интеллекта делятся на три категории: системы искусственного интеллекта с поддержкой человека, человеко-машинные гибридные интеллектуальные системы и полностью автономные интеллектуальные системы. В зависимости от их важности и степени вреда системы искусственного интеллекта делятся на три уровня: интеллектуальные системы с низким уровнем риска, интеллектуальные системы с высоким уровнем риска и интеллектуальные системы со сверхвысоким уровнем риска. Каждый из этих типов систем искусственного интеллекта можно дополнительно разделить на три уровня.

(ii) Заслуживающий доверия ИИ стал важной методологией для реализации требований управления

Согласно *Белой книге по надежному искусственному интеллекту,*[4] столкнувшись с глобальной тревогой, вызванной отсутствием доверия к ИИ, разработка надежного ИИ стала глобальным консенсусом. **Надежный ИИ - это набор методологий для реализации требований к управлению ИИ с точки зрения промышленного измерения и мост между**

4 Translator's note: For an English translation of the CAICT-JD *White Paper on Trustworthy Artificial Intelligence,* see: https://cset.georgetown.edu/publication/white-paper-on-trustworthy-artificial-intelligence/.

управлением ИИ и промышленными практиками. На рисунке 5 показана общая структура надежного ИИ.

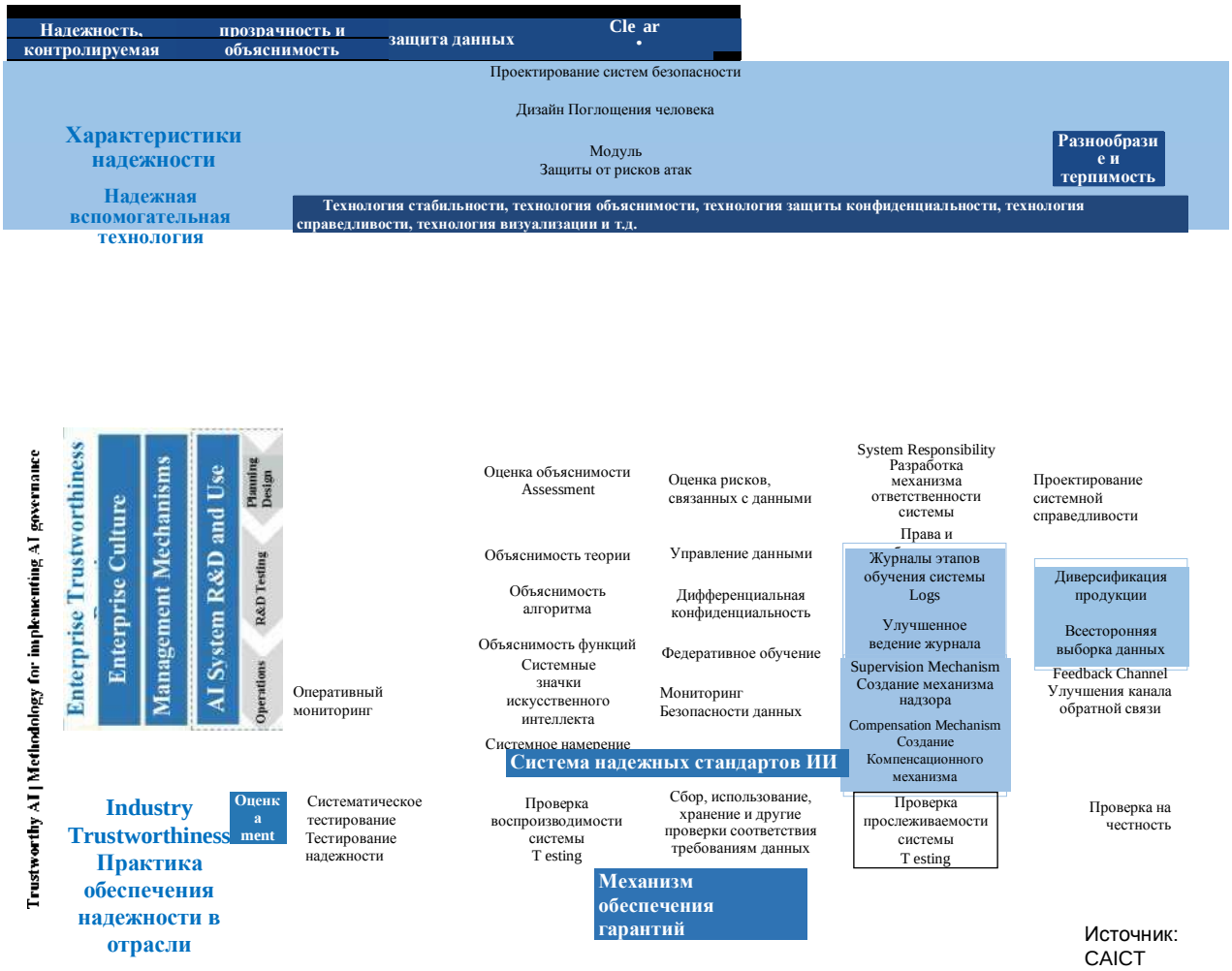


Рисунок 5 Общая структура заслуживающего доверия ИИ

1. Концепция доверия постепенно проникает во весь жизненный цикл ИИ

Заслуживающий доверия ИИ был предложен академическими кругами, активно изучался во многих секторах, а затем начал применяться на практике промышленными кругами. Его значение постепенно становится все богаче и развивается. Определение заслуживающего доверия ИИ больше не ограничивается статусами самих технологий, продуктов и услуг ИИ, а вместо этого постепенно расширилось до набора систематических методологий, включающих все аспекты, связанные с созданием "заслуживающего доверия" ИИ. Это включает в себя внутреннее управление, исследования и разработки, операции и другие аспекты предприятий, а также отраслевую работу по преобразованию соответствующих абстрактных требований в конкретные требования к возможностям для практики, тем самым повышая уровень доверия общества к ИИ.

2. Предприятия стали главной силой в применении надежного искусственного интеллекта Являясь передовой линией исследований и разработок в области технологий искусственного интеллекта и их инновационного применения, предприятиям необходимо решать проблему доверия к ИИ, активно заниматься самодисциплиной и самоуправлением и в полной мере использовать инициативу предприятий по внедрению требований надежности технологий, продуктов и услуг искусственного интеллекта. С 2018 года многие отечественные и зарубежные компании, такие как Google, Microsoft, IBM, Megvii и Tencent, внедрили руководящие принципы корпоративного управления ИИ и сформировали соответствующие департаменты и агентства для содействия внедрению

обязанности руководства. Кроме того, предприятия также активно изучают модели управления искусственным интеллектом, в основе которых лежит принцип надежности. IBM, Microsoft, Huawei, JD и другие китайские и зарубежные предприятия выпустили ряд инструментов обеспечения надежности ИИ, чтобы помочь продуктам ИИ повысить безопасность, надежность, объяснимость, честность и другие надежные возможности в процессе исследований и разработок, а также объединить разработчиков для продвижения концепции надежности посредством экосистема с открытым исходным кодом.

3. Отраслевые организации способствуют созданию безопасной и заслуживающей доверия экосистемы для искусственного интеллекта

Реализация надежного ИИ достигается не только односторонней практикой и усилиями предприятий, но также требует участия и координации нескольких сторон. В конечном счете, должна быть сформирована здоровая экосистема взаимного влияния, взаимной поддержки и взаимозависимости. **На уровне разработки стандартов** с 2017 года ISO / IEC, IEEE, SAC / TC 28 / SC 42 и другие китайские и зарубежные организации по стандартизации взяли на себя ведущую роль в разработке универсальных стандартов для надежного ИИ. В апреле 2021 года требования к безопасности данных китайского национального стандарта *информационной безопасности для распознавания* лиц были открыты для общественности для комментариев. **Что касается самодисциплины в отрасли**, то Китайский альянс индустрии искусственного интеллекта опубликовал *Совместное обязательство по самодисциплине в отрасли искусственного интеллекта* в 2019 году. Впоследствии, в 2020 году, они опубликовали *Руководство по надежным операциям ИИ* и объявили о первой партии результатов оценки надежности коммерческих систем ИИ, в которых приняли участие 16 систем ИИ из 11 компаний. Это послужило важным ориентиром при выборе модели для пользователей. В настоящее время *административные руководящие принципы для заслуживающих доверия исследований и разработок в области искусственного интеллекта* и другие документы составляются совместно с промышленными кругами в целях повышения безопасности и надежности исследований и разработок в области искусственного интеллекта у источника.

II. Резюме и перспективы

Китайские технологии и промышленность искусственного интеллекта добились большого прогресса в развитии. Мы считаем, что в период "14-й пятилетки" [2021-2025] инновации в области технологий искусственного интеллекта будут еще более ускорены, масштабы отрасли будут продолжать расширяться, и появится ряд высококачественных предприятий и промышленных кластеров с большим потенциалом развития, которые станут важным двигателем в продвижении качественное развитие экономики.

Стремление к технологическим инновациям, сосредоточение внимания на инженерных практиках и обеспечение надежности и безопасности постепенно становятся важными направлениями будущего развития ИИ. Оглядываясь назад на развитие искусственного интеллекта за последние десять лет, нетрудно обнаружить, что технологические инновации и инженерные

практики

дополняют друг друга. Прорывы в алгоритмах и вычислительной мощности привели к разработке инструментальных систем, а зрелость инструментов еще больше способствовала применению технологий. В настоящее время ИИ уже широко используется во всех аспектах повседневной работы и жизни людей, а спрос на его безопасность, надежность и качество возрос до беспрецедентного уровня. Содействие надежному и контролируемому развитию ИИ стало глобальным консенсусом. Стоя у истоков "14-го пятилетнего плана", мы с нетерпением ожидаем постоянного совершенствования технологий искусственного интеллекта и энергичного и здорового развития индустрии ИИ и приложений в течение следующих пяти лет.

Во-первых, постоянно исследуя новые технологии, мы должны уделять больше внимания получению технологических дивидендов с помощью инженерных методов и обеспечению безопасности и надежности. Ключевым фактором, определяющим, смогут ли компании с искусственным интеллектом быстро расширить возможности всех отраслей и секторов и реагировать на разнообразные потребности, являются инженерные возможности предприятий. В то же время спрос на безопасные и надежные технологии становится все более и более важным. В настоящее время появилось большое количество компаний, занимающихся вычислительными технологиями, обеспечивающими конфиденциальность, и сосредоточенных на защите данных. В будущем технологии, ориентированные на стабильность и справедливость ИИ, также станут важной силой.

Во-вторых, в процессе промышленной интеллектуализации уровень участия традиционных отраслей будет все более и более глубоким, и они даже будут руководить процессом развития всей отрасли. Фокус промышленного развития уже начал смещаться с "ИИ +" на "+ ИИ". С улучшением процесса оцифровки традиционных отраслей промышленности будут предоставлены огромные объемы данных и богатые сценарии применения, что откроет новые возможности для применения ИИ. В этих традиционных отраслях и областях учреждения с более высоким уровнем проникновения ИИ будут предлагать решения, связанные с ИИ, другим учреждениям в своих областях.

В-третьих, управление искусственным интеллектом будет становиться все более и более важным. Это связано с устойчивым и здоровым развитием ИИ, и координация управления и развития стала необходимостью. Управленческая работа не только практически связана с повседневным применением ИИ, но и стала важной темой международной конкуренции и сотрудничества. Учитывая различия в культурном происхождении и уровнях развития в разных странах и регионах по всему миру, вопрос о том, как эффективно применять методы управления ИИ, является важной задачей. Китайское правительство, отраслевые организации и предприятия взяли на себя ведущую роль в начале изучения управления ИИ, интегрируя концепцию безопасности и надежности в полный жизненный цикл ИИ. В будущем также появится больше практических парадигм.