УДК 004.89

СОЗДАНИЕ ЭКОСИСТЕМЫ ДОВЕРЕННЫХ КОМПОНЕНТОВ ДЛЯ УСТОЙЧИВОГО МАШИННОГО ОБУЧЕНИЯ

Зайцев С.В., студент гр. 53, I курс
Томский государственный университет, г. Томск

В условиях стремительного роста применения искусственного интеллекта и машинного обучения (ML) в критически значимых сферах – от здравоохранения и образования до национальной безопасности и финансового регулирования – проблема устойчивости, воспроизводимости и доверия к ИИ-системам становится не просто исследовательским вызовом, а социально-политическим и экономическим императивом. Ключевым элементом решения этой задачи является не только разработка надёжных моделей, но и формирование экосистемы доверенных компонентов, которые могут быть многократно использованы, верифицированы, оценены и сертифицированы. Такая экосистема обеспечивает структурную устойчивость, транспарентность и нормативную совместимость решений машинного обучения на всех уровнях их жизненного цикла. В данной статье предпринимается попытка систематизировать понятие доверенного ML-компонента, проанализировать условия его функционирования в рамках экосистемы и определить архитектурные, нормативные и организационные принципы устойчивого построения таких систем [1].

Под доверенным компонентом машинного обучения следует понимать программный или логический модуль, который выполняет конкретную функцию (например, препроцессинг, обучение, инференс, мониторинг, логгирование, защита приватности) и обладает зафиксированными характеристиками: проверенной реализацией, воспроизводимым поведением, задокументированными ограничениями и доказанным соответствием требованиям безопасности, этики и закона. Такие компоненты должны быть взаимозаменяемыми, версионируемыми и независимо верифицируемыми. Их наличие делает возможным проектирование сложных ИИ-систем по принципу безопасной модульной сборки, в которой архитектура легко поддаётся анализу, а риски локализованы.

Формирование полноценной экосистемы доверенных компонентов предполагает наличие нескольких взаимосвязанных слоёв:

1. Технический слой, в который входят библиотеки и модули, реализующие стандартные функции обработки и анализа данных, модели

обучения, средства визуализации и предсказания, а также утилиты для взаимодействия с окружением (сетями, API, базами данных). Доверенные компоненты здесь должны быть совместимы с промышленными стандартами (например, ONNX, MLIR, PMML), сопровождаться спецификациями входов/выходов, единиц измерения и ограничений, а также содержать встроенные self-test-механизмы и средства контроля версий [2].

2. Метаданные и паспортизация, в которых каждый компонент сопровождается машиночитаемыми описаниями (например, в формате SPDX, JSON-LD, RDF), включающими сведения об авторстве, происхождении, лицензировании, целевом применении, пройденных проверках и уровне зрелости. Эти метаданные составляют основу для формирования доверенных репозиториев, каталогов и индексов, к которым могут обращаться как пользователи, так и автоматические системы проектирования.

3. Верификационная инфраструктура, обеспечивающая автоматическую или полуавтоматическую проверку соответствия компонентов требованиям. Это включает статический и динамический анализ, формальную верификацию, тесты на устойчивость, интерпретируемость, отказоустойчивость, защищённость от атак, а также проверку на наличие предвзятости и дискриминации. Результаты таких проверок фиксируются в общедоступных регистрах или снабжаются цифровыми подписями от верифицирующих организаций.

4. Нормативный и юридический слой, включающий лицензии (например, Apache 2.0, MIT, Copyleft, Open Use), модели этической сертификации (например, AI Ethics Label), аттестацию на соответствие стандартам (ISO/IEC 23894, AI Act, GDPR), а также процедуры оценки соответствия в регулируемых отраслях (медицине, банковском деле, энергетике). Он позволяет формировать правовую определённость при использовании доверенных компонентов и обеспечивает юридическую защищённость всех участников цепочки поставок ИИ [3].

5. Организационно-сетевой слой, в который входят институции (фондовые платформы, стандартизирующие организации, научные консорциумы), платформы сотрудничества (например, LF AI & Data, OpenSSF, GAIA-X, AI Verify), а также механизмы координации между поставщиками компонентов, пользователями, разработчиками и аудиторами. Этот слой обеспечивает коллективное управление рисками, распределённую экспертизу и устойчивость всей экосистемы к точечным сбоям или умышленным саботажам.

Сильной стороной подхода к проектированию экосистемы доверенных компонентов является встроенная поддержка устойчивости, как технической, так и социальной. Во-первых, компоненты с известной спецификацией легче

подвергать автоматическому тестированию, заменять и масштабировать. Во-вторых, их формализованная структура облегчает интеграцию в CI/CD пайплайны и процессы управления жизненным циклом модели (MLOps), включая повторное обучение, валидацию и мониторинг в реальном времени [4].

В-третьих, коллективная разработка и прозрачность архитектуры способствуют воспроизводимости результатов, что особенно важно для научного сообщества и регулируемых отраслей.

Однако построение такой экосистемы сталкивается с рядом вызовов. Необходимо выработать общие стандарты описания и взаимодействия компонентов, чтобы обеспечить совместимость и повторное использование. Следует решить проблему доверия к источникам компонентов: даже открытый код может содержать вредоносные вставки или быть скомпрометирован. Возникает необходимость в разработке реестров доверенных компонентов с централизованной или федеративной системой репутации, а также в институтах сертификации, которым можно доверять. Кроме того, требуется продуманная модель экономической мотивации для разработчиков, поддерживающих доверенные компоненты, особенно в условиях открытого кода и распределённой разработки.

Будущее устойчивого и доверенного машинного обучения невозможно без комплексной экосистемы компонентов, в которой каждый элемент можно будет проверить, заменить, объяснить и верифицировать. Такая система приближает область ИИ к зрелым отраслям инженерии – таким как авиация, электроэнергетика или фармацевтика, – где сертифицированные компоненты и стандартизованные интерфейсы давно являются нормой. Создание экосистемы доверенных компонентов – это путь к цифровому ИИ-строительству нового поколения, в котором устойчивость, транспарентность и этическая ответственность встроены в саму логику архитектурных решений [5].

Список литературы:

1. National Telecommunications and Information Administration. The Minimum Elements for a Software Bill of Materials (SBOM). – Washington : NTIA, 2021. – 18 p.

2. Sigstore. Sigstore Documentation [Электронный ресурс]. – URL: https://docs.sigstore.dev/ (дата обращения: 09.06.2025).

3. OpenSSF. Supply-chain Levels for Software Artifacts (SLSA), Version 1.0 [Электронный ресурс]. – URL: https://slsa.dev/spec/v1.0/ (дата обращения: 09.06.2025).

4. Pashchenko, I. A Qualitative Study of Dependency Management and Its Security Implications / I. Pashchenko, D.-L. Vu, F. Massacci // Proceedings of the

ACM SIGSAC Conference on Computer and Communications Security. – 2020. – P. 1513–1531. – DOI 10.1145/3372297.3423361.

5. Zimmermann, M. Small World with High Risks: A Study of Security Threats in the npm Ecosystem / M. Zimmermann, C.-A. Staicu, C. Tenny, M. Pradel // USENIX Security Symposium. – 2019. – P. 995–1010.