УДК 004.8

Захаров А.С.

магистрант 2 курса

Санкт-Петербургский государственный университет

аэрокосмического приборостроения

Российская Федерация, г. Санкт-Петербург

ГИБРИДНАЯ МОДЕЛЬ ПЕРСОНАЛЬНЫХ ВЫЧИСЛЕНИЙ: ВЗАИМОДЕЙСТВИЕ ЛОКАЛЬНЫХ УСТРОЙСТВ, ОБЛАЧНЫХ СЕРВИСОВ И ИИ-АГЕНТОВ

Аннотация. *В статье рассматриваются перспективы развития персональных вычислений на основе взаимодействия локальных устройств, облачной инфраструктуры и интеллектуальных агентов. Проанализированы особенности распределения вычислительных задач, обмена данными и управления действиями в гибридной среде. На основе существующих научных подходов предложена концептуальная модель организации такого взаимодействия. Определены её потенциальные преимущества и ограничения, связанные с производительностью, надёжностью и защитой информации.*

Ключевые слова: *искусственный интеллект; персональные вычисления; облачные технологии; периферийные вычисления; ИИ-агенты; распределённые системы; гибридная архитектура.*

Zakharov A.S.

second-year master's student

Saint Petersburg State University of Aerospace Instrumentation

Russian Federation, Saint Petersburg

HYBRID PERSONAL COMPUTING MODEL: INTERACTION BETWEEN LOCAL DEVICES, CLOUD SERVICES AND AI AGENTS

Abstract. *The article examines the prospects for personal computing based on the interaction of local devices, cloud infrastructure, and intelligent agents. The distribution of computational tasks, data exchange, and coordination of actions in hybrid environments are analyzed. Based on existing scientific approaches, a conceptual model of such interaction is proposed. Its potential advantages and limitations related to performance, reliability, and information security are identified.*

Keywords: *artificial intelligence; personal computing; cloud technologies; edge computing; AI agents; distributed systems; hybrid architecture.*

Введение. Развитие персональных вычислений постепенно изменяет привычную модель использования цифровых устройств. Ранее основная часть пользовательских задач выполнялась непосредственно на персональном компьютере, а удалённые сервисы применялись преимущественно для обмена информацией и хранения данных. В настоящее время вычислительные возможности распределяются между смартфонами, компьютерами, домашними серверами и облачной инфраструктурой. Дополнительные изменения связаны с распространением искусственного интеллекта (ИИ), позволяющего обрабатывать информацию и автоматизировать отдельные действия пользователя. Особый интерес представляет объединение перечисленных компонентов в единую систему, где вычислительные задачи могут выполняться на разных устройствах в зависимости от их сложности и требований к обработке данных.

Цель настоящей статьи заключается в рассмотрении принципов взаимодействия локальных устройств, облачной инфраструктуры и ИИ-агентов, а также в разработке концептуальной модели распределения задач между ними. Основное внимание уделяется организации совместной

работы компонентов, условиям передачи информации и сохранению пользовательского контроля.

Методы исследования. Исследование выполнено с применением сравнительного анализа научных публикаций, посвящённых периферийным вычислениям, взаимодействию локальных и облачных языковых моделей и архитектуре интеллектуальных агентов. Под периферийными вычислениями понимается обработка информации на устройствах, расположенных вблизи источника данных, включая персональные компьютеры и другие пользовательские устройства. Для разработки концептуальной модели использован системный подход, предполагающий рассмотрение вычислительных компонентов, каналов связи и правил распределения задач как элементов единой архитектуры.

В качестве основных критериев сопоставления выбраны потребность в вычислительных ресурсах, чувствительность обрабатываемой информации, требования к задержке, доступность сети и возможность самостоятельного выполнения задачи. Предлагаемая модель представляет собой результат теоретического обобщения. Её экспериментальная реализация и количественная оценка производительности в рамках настоящей работы не проводились.

Результаты исследования. Одним из основных условий формирования гибридной среды является развитие локальных вычислений на устройствах пользователя. Современные компьютеры и смартфоны могут выполнять отдельные операции искусственного интеллекта без постоянного обращения к удалённому серверу. Такие задачи могут включать распознавание речи, классификацию информации и обработку небольших текстовых запросов. В обзоре Ю. Чжэна и соавторов рассмотрены подходы к выполнению больших языковых моделей на устройствах с ограниченными ресурсами, включая оптимизацию памяти и сокращение вычислительной нагрузки [1]. Однако возможности локальных

систем ограничиваются характеристиками оборудования, поэтому выполнение сложных задач не всегда оказывается практически целесообразным.

Облачная инфраструктура в гибридной архитектуре может использоваться для обработки запросов, требующих значительных вычислительных ресурсов. При этом важным становится не простое разделение задач на локальные и удалённые, а возможность выбирать место их выполнения в зависимости от текущих условий. Подобный подход представлен в исследовании Х. Цзиня и Я. У, посвящённом системе CE-CoLLM. Авторы рассмотрели взаимодействие периферийных устройств и облачных вычислительных ресурсов при выполнении языковых моделей, предложив механизмы адаптивного распределения вычислений [2]. Исследование показывает техническую возможность совместной обработки запросов, хотя эффективность конкретной системы зависит от используемого оборудования и характеристик сети.

Следующим компонентом гибридной среды могут выступать ИИ-агенты. В отличие от обычных интеллектуальных помощников, ограниченных формированием текстового ответа, агентные системы способны последовательно выбирать действия и обращаться к доступным программным инструментам. В исследовании Ш. Яо и соавторов представлен подход ReAct, объединяющий формирование рассуждений с выполнением действий во внешней среде [3]. Для персональных вычислений подобный принцип представляет интерес как основа взаимодействия с приложениями, файлами и удалёнными сервисами. Однако наличие агентных возможностей не означает, что система должна самостоятельно принимать любые решения без участия пользователя.

На основе рассмотренных подходов предлагается гибридная модель, включающая три взаимосвязанных уровня. Первый уровень образуют персональные устройства пользователя, на которых находятся приложения,

исходные данные и средства выполнения относительно простых операций. Второй уровень представлен облачной инфраструктурой, предоставляющей дополнительные вычислительные ресурсы и специализированные сервисы. Третий уровень составляет агентная система координации, определяющая последовательность действий, выбирающая подходящие инструменты и организующая обмен результатами между компонентами.

Принципиально важно, что агентная система в предложенной модели рассматривается не как отдельное обязательное физическое устройство, а как программный компонент. Она может размещаться непосредственно на компьютере пользователя либо распределяться между несколькими вычислительными узлами. Такой подход позволяет отделить управление задачами от непосредственного выполнения вычислений. Н. Агравал и Р. Майер в исследовании периферийных интеллектуальных агентов выделяют различные варианты размещения агентных систем и рассматривают проблемы автономности, масштабирования и взаимодействия компонентов [4].

Практическую организацию гибридного взаимодействия можно рассмотреть на условном примере подготовки аналитического документа. Пользователь обращается к агенту с просьбой обработать несколько файлов, сопоставить содержащиеся сведения и сформировать итоговый материал. Первоначальный поиск документов и извлечение необходимой информации могут выполняться локально. Если задача требует более сложной обработки, система может предложить обращение к облачной модели, предварительно определив, какие данные допустимо передавать. Полученный результат возвращается на устройство пользователя, где сохраняется в выбранном формате и проходит окончательную проверку. Данный пример иллюстрирует предлагаемую архитектуру, а не результаты реализованного программного эксперимента.

Существенным преимуществом такой модели может стать согласованное взаимодействие нескольких личных устройств. Например, задача формулируется на смартфоне, часть вычислений выполняется на домашнем компьютере, а результат становится доступен на ноутбуке. Для реализации подобного сценария необходимы механизмы идентификации устройств, синхронизации состояния задач и разграничения доступа. Кроме того, система должна учитывать ситуации, когда отдельный компьютер выключен или временно недоступен. Поэтому устойчивость взаимодействия становится не менее важным требованием, чем производительность вычислений.

Особое значение приобретает распределение задач с учётом чувствительности данных. Локальная обработка может быть предпочтительна при работе с личными документами и информацией, передачу которой необходимо ограничивать. Однако сохранение вычислений внутри устройства само по себе не обеспечивает полной защиты. Требуются контроль разрешений, безопасное хранение информации и проверка действий программных компонентов. Дополнительную проблему создаёт возможность воздействия на агента через недостоверные инструкции во внешних документах. В исследовании Э. Дебенедетти и соавторов, посвящённом системе AgentDojo, продемонстрированы риски атак, при которых содержимое обрабатываемых данных способно влиять на действия интеллектуального агента [5].

Поэтому в предлагаемой архитектуре целесообразно отделять подготовку действия от разрешения на его выполнение. Например, агент может самостоятельно подобрать файлы, составить план обработки или подготовить текст сообщения, но удаление документов, отправка конфиденциальной информации и изменение критически важных настроек должны требовать дополнительного подтверждения. Такой принцип

позволяет сохранять преимущества автоматизации при ограничении последствий ошибочных решений.

Ещё одним условием является возможность работы при нестабильном интернет-соединении. Если вся вычислительная деятельность зависит от удалённого сервера, потеря связи делает недоступной значительную часть функций. Гибридная архитектура потенциально позволяет сохранять локальную обработку ограниченного набора задач, временно откладывая операции, требующие облачной инфраструктуры. Однако для этого необходимы механизмы сохранения состояния, повторного выполнения запросов и предотвращения конфликтов при синхронизации.

Проведённый анализ позволяет выделить важную особенность предложенной модели: эффективность взаимодействия определяется не максимальным использованием облачных вычислений или локального оборудования, а соответствием выбранного способа выполнения требованиям конкретной задачи. Передача простого действия на удалённый сервер может создавать избыточные задержки, тогда как попытка выполнить сложную операцию на маломощном устройстве способна приводить к неоправданным затратам времени и энергии. Поэтому целесообразность распределения необходимо оценивать с учётом совокупной продолжительности работы, доступности компонентов и рисков обработки информации.

Заключение. Таким образом, перспективным направлением развития персональных вычислений является формирование гибридной среды, объединяющей локальные устройства, облачную инфраструктуру и интеллектуальных агентов. Предложенная концептуальная модель предусматривает распределение вычислительных задач с учётом их сложности, требований к задержке, доступности сети и конфиденциальности данных. ИИ-агенты в такой системе могут выполнять

координирующую функцию, обеспечивая взаимодействие различных устройств и программных сервисов.

При этом реализация подобной архитектуры связана с необходимостью решения вопросов безопасности, синхронизации и контроля автономных действий. Дальнейшее развитие персональных вычислений может быть связано не с вытеснением локальных устройств облачными сервисами, а с более тесным объединением их возможностей. Практическая эффективность предложенной модели требует последующей экспериментальной проверки в условиях реальной пользовательской нагрузки.

Список использованных источников

1. Zheng Y., Chen Y., Qian B., Shi X., Shu Y., Chen J. A Review on Edge Large Language Models: Design, Execution, and Applications // ACM Computing Surveys. — 2025. — Vol. 57, № 8. — Article 209. — P. 1–35. — DOI: 10.1145/3719664.
2. Jin H., Wu Y. CE-CoLLM: Efficient and Adaptive Large Language Models Through Cloud-Edge Collaboration // 2025 IEEE International Conference on Web Services (ICWS). — 2025. — P. 316–323. — DOI: 10.1109/ICWS67624.2025.00046.
3. Yao S., Zhao J., Yu D., Du N., Shafran I., Narasimhan K., Cao Y. ReAct: Synergizing Reasoning and Acting in Language Models // The Eleventh International Conference on Learning Representations. — 2023. — URL: https://openreview.net/forum?id=WE_vluYUL-X (дата обращения: 14.11.2026).
4. Agrawal N., Mayer R. LLM-Based Agents on the Edge: A Survey of Privacy, Scalability, Heterogeneity, and Autonomy // Proceedings of the Thirty-Fifth International Joint Conference on Artificial Intelligence. — 2026. — P. 7722–7730. — DOI: 10.24963/ijcai.2026/858.

5. Debenedetti E., Zhang J., Balunović M., Beurer-Kellner L., Fischer M., Tramèr F. AgentDojo: A Dynamic Environment to Evaluate Prompt Injection Attacks and Defenses for LLM Agents // Advances in Neural Information Processing Systems. — 2024. — Vol. 37. — P. 82895–82920. — DOI: 10.52202/079017-2636.