ВЫБОР УРОВНЯ АВТОНОМНОСТИ АГЕНТНОГО ИИ В БИЗНЕС–ПРОЦЕССАХ

Шадт М.А., магистрант,

научный руководитель

Королев В.В., доцент,

ГАОУ ВО ЛО «ЛГУ им. А.С. Пушкина», г. Санкт-Петербург, Россия

Аннотация. На основе анализа восьми публикаций систематизированы факторы выбора автономности агентного ИИ в бизнес-процессах. Обоснованы четыре режима, различающиеся полномочиями системы и участием человека. Предложенный подход связывает допустимость передачи операций с последствиями ошибок, условиями контроля и экономической эффективностью.

Ключевые слова: агентный ИИ, автономность, бизнес-процессы, контроль, риски, эффективность.

Агентные системы используют языковые модели для выбора и выполнения действий через внешние инструменты, учитывая полученную обратную связь. Однако техническая способность выполнить задачу ещё не определяет допустимые полномочия. При подготовке документа агентом сотруднику необходимо проверить, какие сведения использованы и насколько обосновано содержание. Пока документ

остаётся проектом, ошибки можно исправить до передачи получателю. Отправка создаёт внешние последствия: получатель может принять решение на основании ошибочных сведений. Поэтому в рассматриваемом случае агенту поручается подготовка проекта, а отправка допускается после проверки и одобрения сотрудником. Цель исследования - разработать и теоретически обосновать подход к выбору автономности агентного ИИ с учётом экономической эффективности, последствий ошибок и человеческого контроля.

В работе сопоставлены восемь научных и аналитических публикаций об агентном ИИ. Анализ посвящён условиям, от которых зависит передача системе права выполнять операции: требованиям к данным, надёжности результата, участию человека и затратам. На основе этого сопоставления сформирован подход к выбору автономности для отдельной операции или участка бизнес-процесса. Под автономностью понимается степень самостоятельности агента в выборе и выполнении действий. Она определяется тем, какие операции система вправе выполнять без согласования, какие требуют предварительного одобрения и когда задача должна быть передана сотруднику.

Рассмотренные публикации раскрывают разные стороны применения агентного ИИ. Re Act
описывает получение внешней информации в ходе выполнения задачи [6], а обзор L
.
Wang
и соавторов - архитектуры агентов и способы оценки их работы [1].
F
.
Bousetouane
рассматривает использование знаний предметной области [3]. Эти работы помогают оценить возможности системы, однако для определения её полномочий необходимо учитывать также надёжность действий. В материалах
Anthropic
обсуждаются накопление ошибок и условия остановки [4], а в

τ
-
bench
проверяются достижение требуемого результата и повторяемость успешного выполнения [8]. Поэтому при выборе автономности следует оценивать последствия ошибок и возможность их своевременного обнаружения.

Организационные условия дополняют эту оценку. Материалы Dynatrace отражают применение проверок качества данных и мониторинга [5],
Deloitte
- необходимость подготовки данных и пересмотра процессов [7]. Работа
AI
Agents
That
Matter
показывает значение совместной оценки качества и стоимости выполнения [2]. На этой основе в предлагаемый подход включены требования к данным, возможности человеческого контроля и совокупные затраты. При расчёте последних учитываются также расходы организации на проверку и исправление результатов.

Обратимость действий выделена в работе как самостоятельный фактор, поскольку последствия ошибки зависят и от возможности её исправить. Например, ошибочную запись во внутренней системе можно скорректировать, но сведения из уже отправленного документа получатель может использовать до получения исправлений. Поэтому возможность отмены действия следует оценивать вместе со временем обнаружения ошибки и последствиями, возникающими до её устранения. Элемент научной новизны состоит в систематизации этих условий для выбора полномочий агента. В предлагаемом подходе сначала определяются допустимые по последствиям ошибок и условиям контроля режимы, а затем сравниваются затраты на их применение.

В основу разграничения четырёх режимов положено распределение решений между сотрудником и агентом (таблица 1). В первом режиме человек принимает решение и выполняет операцию, используя подготовленные системой рекомендации. Во втором исполнение передаётся агенту, но каждое действие требует одобрения. Третий режим допускает выполнение заранее разрешённых операций без отдельного согласования. Четвёртый предусматривает передачу дополнительного полномочия - выбора последовательности действий на заданном участке процесса. Поэтому различие двух последних режимов заключается в том, кто определяет следующий шаг: человек или установленный порядок выполнения либо сам агент. Во всех случаях самостоятельность ограничена предоставленными полномочиями. Предложенные режимы описывают варианты организации работы, а не стадии технического развития системы.

Таблица 1 – Уровни автономности агентного ИИ в бизнес-процессах

Уровень	Полномочия системы	Роль человека
Рекомендательный	Анализирует данные, формирует предложения и проекты документов	

Принимает решения и выполняет операции

Выполнение с согласованием

Планирует операции и выполняет действия после подтверждения

Проверяет и разрешает действия до исполнения

Ограниченная автономность

Выполняет разрешённые операции в пределах установленных правил и лимитов

Согласует исключения и выход за пределы полномочий

Автономное выполнение под наблюдением

Самостоятельно выбирает последовательность действий на установленном участке процесса

Контролирует результаты и вмешивается при отклонениях

Для выбора режима необходимо определить, какой результат должна дать операция и какие последствия её ошибочного выполнения организация считает допустимыми. Недостаток исходных данных требует их уточнения независимо от предполагаемой самостоятельности агента. Если действие нуждается в предварительной проверке, исполнение после согласования возможно при условии, что сотрудник способен проверить его содержание и параметры до запуска. Когда такая проверка не позволяет ограничить последствия ошибки, исполнение сохраняется за человеком либо операция пересматривается. Самостоятельное выполнение отдельных операций допускается при заданных лимитах, проверяемом результате и возможности своевременно остановить работу. Передача агенту выбора последовательности действий требует дополнительной проверки всего участка процесса: ошибка промежуточного шага может повлиять на последующие решения. Поэтому наличие наблюдения само по себе недостаточно - сотрудник должен получать информацию об отклонениях и успевать вмешаться до наступления недопустимых последствий.

После определения допустимых режимов необходимо оценить, оправдано ли их применение с экономической точки зрения. Каждый вариант, включая подготовку рекомендаций, сравнивается с существующим способом работы при одинаковом объёме задач и за один период. В расчёт включаются расходы на внедрение и эксплуатацию системы, проверку результатов сотрудниками и исправление ошибок. При этом сокращение времени выполнения не означает экономии: важно, как будет использовано освободившееся рабочее время. Для сравнения режимов предлагается учитывать долю правильно выполненных задач, среднее время обработки, долю задач, потребовавших участия сотрудника, и стоимость одной правильно выполненной задачи. Этот показатель рассчитывается делением общих затрат за период на число таких задач. Если передача дополнительных полномочий агенту не даёт преимуществ перед текущим порядком работы, экономических оснований для увеличения его автономности нет.

Уровень автономности следует выбирать для каждой операции с учётом её особенностей. Поэтому в одной организации агент может выполнять часть действий

самостоятельно, а другие - только после согласования с сотрудником. Успех в отдельных заданиях ещё не означает, что системе можно доверить весь процесс. Предложенная классификация пока имеет теоретический характер: её применение не проверялось на данных организации, а числовые границы допустимого риска не определены. Следующим шагом должно стать сравнение работы человека и агента в разных режимах на одинаковых задачах. Это позволит оценить, когда передача полномочий оправдана с точки зрения качества результата, времени выполнения и общих затрат.

Литература

1. A Survey on Large Language Model based Autonomous Agents / L. Wang, C. Ma, X. Feng [et al.]. 2025. URL: https://arxiv.org/abs/2308.11432v7 (дата обращения: 10.11.2026).

2. AI Agents That Matter / S. Kapoor, B. Stroebl, Z. S. Siegel [et al.]. 2024. URL: https://arxiv.org/abs/2407.01502 (дата обращения: 10.11.2026).

3. Bousetouane F. Agentic Systems: A Guide to Transforming Industries with Vertical AI Agents. 2025. URL: https://arxiv.org/abs/2501.00881 (дата обращения: 10.11.2026).

4. Building effective agents // Anthropic. 2024. 19 Dec. URL: https://www.anthropic.com/engineering/building-effective-agents (дата обращения: 10.11.2026).

5. New global report finds enterprises hitting Agentic AI inflection point // Dynatrace. 2026. 22 Jan
.
URL
:
https
://
www
.
dynatrace
.
com
/
news
/
press
-
release
/
pulse
-
of
-
agentic
-
ai
-2026/ (дата обращения: 10.11.2026).

6. ReAct: Synergizing Reasoning and Acting in Language Models / S. Yao, J. Zhao, D. Yu [et al.]. 2023. URL: https://arxiv.org/abs/2210.03629v3 (дата обращения: 10.11.2026).

7. The agentic reality check: Preparing for a silicon-based workforce // Deloitte Insights. 2025. 10 De
c .
URL
:
https
://
www
.
deloitte
.

com
/
us
/
en
/
insights
/
topics
/
technology
-
management
/
tech
-
trends
/2026/
agentic
-
ai
-
strategy
.
html
(дата обращения: 10.11.2026).

8. τ-bench: A Benchmark for Tool-Agent-User Interaction in Real-World Domains / S. Yao, N. Shinn, P. Razavi, K. Narasimhan. 2024. URL: https://arxiv.org/abs/2406.12045 (дата обращения: 10.11.2026).