EDN XKGYNH

BY-SA

Границы применимости ИИ в государственном управлении: Память LLM-агентов vs память человека*

*Вейбер Е. Н.[1], Грудинин М. А.[1], Тулупьева Т. В.[1, 2, *], Вяткин А. А.[1]*

[1] Российская академия наук, Санкт-Петербургский Федеральный исследовательский центр, Санкт-Петербург, Российская Федерация
[2] Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации, Москва, Российская Федерация; *tvt@dscs.pro

РЕФЕРАТ

В последние годы большие языковые модели и агентные системы на их основе рассматриваются как перспективный инструмент цифровой трансформации государственного управления. Однако практическая эффективность таких систем определяется не только качеством текстовой генерации в узком смысле (грамматической корректностью, связностью и общей осведомленностью модели), но и тем, насколько надежно они удерживают контекст, извлекают ранее полученные сведения и воспроизводят процедурные правила в условиях длительных управленческих процессов. Для государственных организаций это, как правило, особенно критично: управленческая деятельность опирается на устойчивую память о нормативных требованиях, организационных регламентах, фактах конкретных дел и их хронологии.

Цель. Целью статьи является определение границ применимости в государственном управлении агентных систем, основанных на больших языковых моделях, на основе сравнительного анализа моделей памяти человека и архитектур памяти таких агентных систем.

Методы. В исследовании используется сравнительно-аналитический подход. Рассматриваются базовые когнитивные модели человеческой памяти и выделяются ее характеристики, значимые для управленческой деятельности. Далее анализируются архитектурные механизмы хранения и извлечения информации в агентных системах, функционирующих на основе больших языковых моделей, как их функциональная аналогия.

Результаты. Установлено, что агентные системы воспроизводят внешние функции памяти человека за счет сочетания кратковременного контекста и внешних хранилищ знаний. Одновременно выявлены принципиальные расхождения, связанные с отсутствием автобиографичности, хронологии опыта, встроенных механизмов ответственности и причинно-следственной верификации. Эти ограничения повышают риск искажения контекста и затрудняют проверку сгенерированных утверждений.

Выводы. Сделан вывод о том, что на сегодняшний день агентные системы, основанные на больших языковых моделях, не могут использоваться для автономного принятия решений в ответственных управленческих процедурах. Вместе с тем они обладают значительным потенциалом как когнитивные ассистенты государственных служащих при обязательном человеческом контроле и сохранении персональной ответственности.

Ключевые слова: искусственный интеллект, агентные системы, контекст, знания, ответственность.

Для цитирования: *Вейбер Е. Н., Грудинин М. А., Тулупьева Т. В., Вяткин А. А.* Границы применимости ИИ в государственном управлении: память LLM-агентов vs память человека // Управленческое консультирование. 2026. № 2. С. 54–67. EDN XKGYNH

Limits of Applicability of Artificial Intelligence in Public Administration: Memory of LLM-Based Agents vs. Human Memory*

*Evgeniya N. Veiber[1], Mikhail A. Grudinin[1], Tatiana V. Tulupyeva[1, 2, *], Artem A. Vyatkin[1]*

[1] Russian Academy of Sciences, Saint Petersburg Federal Research Center, Saint Petersburg, Russian Federation
[2] Russian Presidential Academy of National Economy and Public Administration, Moscow, Russian Federation; *tvt@dscs.pro

* Статья выполнена в рамках научно-исследовательской работы по государственному заданию СПб ФИЦ РАН Mol_Lab (молодежная_лаб) № FFZF-2024-0003.

The article was carried out as part of the research the state order of SPC RAS no. FFZF-2024-0003.

ABSTRACT

In recent years, large language models and agent systems based on them have been considered a promising tool for the digital transformation of public administration. However, the practical effectiveness of such systems is determined not only by the quality of text generation, including grammatical correctness, coherence, and general awareness, but also by their ability to reliably retain context, retrieve previously acquired information, and reproduce procedural rules within long-term managerial processes. For public organizations, this aspect is particularly critical, as administrative activity relies on stable memory of regulatory requirements, organizational rules, facts of specific cases, and their chronological order.

Purpose. The purpose of the article is to determine the limits of applicability of agent systems based on large language models in public administration through a comparative analysis of human memory models and the memory architectures of such systems.

Methods. The study employs a comparative analytical approach. Basic cognitive models of human memory are examined, and their characteristics relevant to managerial activity are identified. Subsequently, architectural mechanisms of information storage and retrieval in agent systems based on large language models are analyzed as a functional analogue of human memory.

Results. The analysis demonstrates that agent systems reproduce certain external functions of human memory through a combination of short-term contextual representations and external knowledge repositories. At the same time, fundamental differences are identified, including the absence of autobiographical memory, experiential chronology, embedded responsibility mechanisms, and causal verification. These limitations increase the risk of contextual distortion and complicate the validation of generated outputs.

Conclusions. It is concluded that, at present, agent systems based on large language models cannot be used for autonomous decision-making in responsible administrative procedures. Nevertheless, they show significant potential as cognitive assistants for public servants, provided that mandatory human oversight is maintained and personal responsibility for decisions is preserved.

Keywords: artificial intelligence, agent systems, context, knowledge, responsibility.

For citation: Veiber E. N., Grudinin M. A., Tulupyeva T. V., Vyatkin A. A. Limits of Applicability of Artificial Intelligence in Public Administration: Memory of LLM-Based Agents vs. Human Memory // Administrative Consulting. 2026. N 2. P. 54–67. EDN XKGYNH

1. Введение

Интерес к большим языковым моделям (англ. Large Language Models, LLM) в государственном управлении обусловлен их способностью быстро обрабатывать массивы текстовых данных: выделять основную суть документов, формировать справки, подготавливать проекты писем, выявлять противоречия и т. д. [11; 12; 17]. Эти навыки LLM во многом соотносятся с повседневной деятельностью государственных и муниципальных служащих [5; 6; 7], значительная часть которой связана с нормативными актами, служебными документами, перепиской и т. д. В литературе отмечаются как потенциальные преимущества внедрения LLM — ускорение административных процессов и повышение доступности информации для граждан за счет автоматизации подготовки ответов на типовые запросы и навигации по документальным материалам, так и сопутствующие риски, включающие ошибки генерации, непрозрачность принятия решений и правовую неопределенность несения ответственности [3; 4].

Во многих работах LLM описываются преимущественно как средства генерации текста, реагирующие на отдельные запросы пользователей [10; 11; 12]. Такой подход во многом ограничен: большая языковая модель, функционирующая в режиме «вопрос-ответ», не удерживает устойчивый контекст, не отслеживает историю взаимодействий и не способна самостоятельно соотносить текущие действия с целями и процедурами административного процесса [16; 17; 20].

Переход к LLM-агентам означает надстройку над моделью, например, в виде модулей декомпозиции и последовательного выполнения задач, внешних инструментов (доступа к базам данных, интернету, государственным информационным системам и т. д.), а также модулей хранения запросов, ответов и иной внутренней и внешней по отношению к пользователю информации (памяти). В данном контексте важным становится устойчивость контекста, знаний и правил, что смещает фокус анализа с отдельных LLM на LLM-агентов [17; 19; 26; 27].

Проблема исследования связана с риском несоответствия требований управленческих процедур возможностям долговременного контекстного удержания у LLM-агентов. В условиях государственного управления это несоответствие становится критичным, поскольку принятие решений опирается на устойчивую память, накопленный опыт и способность учитывать длительные причинно-следственные связи. Отсутствие сопоставления между когнитивными моделями человеческой памяти и архитектурами памяти LLM-агентов затрудняет оценку реальных границ их применения в управленческой практике.

Цель статьи — определить границы применимости LLM-агентов в государственном управлении на основе сравнительного анализа моделей памяти человека и LLM-агентов.

Методы исследования. В работе применяется сравнительно-аналитический подход. Сначала рассматриваются базовые когнитивные модели человеческой памяти и выделяются ее свойства, критически значимые для государственного управления. Затем анализируются архитектурные механизмы памяти LLM-агентов как функциональной аналогии человеческой памяти. На основе сопоставления этих двух уровней выявляются ключевые сходства и принципиальные расхождения, имеющие практическое значение для применения LLM-агентов в органах государственной власти. Завершающая часть работы посвящена анализу управленческих рисков и формулированию условий и сценариев безопасной интеграции LLM-агентов в управленческие процессы при сохранении ведущей роли человека как субъекта ответственности.

Теоретическая значимость работы заключается в развитии междисциплинарного подхода к анализу памяти LLM-агентов путем систематического сопоставления когнитивных моделей человеческой памяти и инженерных механизмов хранения и извлечения информации в языковых агентных системах в контексте государственного управления. Впервые память LLM-агента рассматривается не только как техническая характеристика архитектуры модели, но и как аналитическая категория, имеющая институциональное значение для оценки надежности, воспроизводимости и обоснованности управленческих решений. Работа уточняет границы применимости функциональной аналогии между человеческой и машинной памятью и показывает, какие свойства человеческой памяти не воспроизводятся в современных LLM-архитектурах и почему это принципиально важно для управленческой теории.

Практическая значимость исследования заключается в возможности использования полученных выводов при разработке и корректировке регламентов внедрения LLM-агентов в деятельность органов государственной власти. Результаты работы могут быть использованы при формировании требований к архитектурам агентных систем. Выводы исследования также могут служить методологической основой для оценки рисков применения LLM-агентов в длительных управленческих процессах и при подготовке юридически значимых решений, а также для определения допустимых ролей таких систем как когнитивных ассистентов государственных служащих.

Научная новизна исследования заключается в том, что память LLM-агента впервые концептуализируется как институционально значимый фактор, влияющий на распределение ответственности, подотчетность и воспроизводимость управленческих

решений в государственном управлении. В отличие от существующих работ, в которых память языковых моделей анализируется преимущественно с инженерной или когнитивно-метафорической точки зрения, в данной работе она рассматривается через призму управленческой и правовой логики. Показано, что ограничения памяти LLM-агентов носят не только технический, но и институциональный характер, что требует особых режимов их использования и нормативного закрепления.

2. Теоретические основы: когнитивные модели человеческой памяти

В данном разделе рассматриваются теоретические основы для анализа памяти, опирающиеся на классические когнитивные модели человеческой памяти. В подразделе 2.1 анализируются многоуровневые и функциональные модели памяти, предложенные Р. Аткинсоном, Р. Шиффрином и Э. Талвингом, с акцентом на различие между кратковременной и долговременной памятью, а также между семантической и эпизодической компонентами. Подраздел 2.2 посвящен характеристикам человеческой памяти, критически важным для государственного управления.

2.1. Многоуровневые модели памяти (Аткинсона — Шиффрина, Талвинга)

Одной из распространенных рамок описания человеческой памяти является многохранилищная модель, предложенная Р. Аткинсоном и Р. Шиффрином [8]. Она используется в когнитивной психологии и смежных дисциплинах как базовая схема, позволяющая структурировать процессы запоминания, удержания и извлечения информации без обращения к нейрофизиологическим деталям.

В модели Аткинсона — Шиффрина для анализа управленческой деятельности принципиальное значение имеют кратковременная и долговременная память, поскольку именно они обеспечивают, с одной стороны, оперативную обработку информации в ходе текущей деятельности, а с другой — накопление и воспроизведение устойчивых знаний и опыта, лежащих в основе институциональных и процедурных решений [8; 22; 25]. Говоря более детально, кратковременная память обеспечивает удержание и обработку информации, непосредственно вовлеченной в текущую деятельность. Долговременная память, в свою очередь, отвечает за устойчивое хранение знаний о правовых нормах, ведомственных процедурах, организационной структуре, типовых рисках и «истории вопроса», что существенно важно для институционального воспроизводства управленческой практики [2; 8; 22].

Дополняя структурное разделение по времени хранения, Э. Талвинг предложил функциональную типологию долговременной памяти, выделив семантическую и эпизодическую память [24]. Семантическая память включает обобщенные знания о мире, правилах и значениях, тогда как эпизодическая память связана с воспоминаниями о конкретных событиях, пережитых в определенном контексте времени, места и участников. Для государственного управления это различие имеет практическое значение: нередко ошибки возникают не из-за незнания нормативных требований как таковых (дефицита семантической памяти), а из-за искаженной реконструкции последовательности событий конкретного случая: дат, оснований, согласований, предыдущих решений и вовлеченных акторов, то есть сбоев на уровне эпизодической памяти [2; 3].

2.2. Критические для государственного управления свойства человеческой памяти

Специфика государственной службы, как правило, характеризуется высокой степенью нормативной регламентации, необходимостью принятия решений в условиях неполной информации, а также институциональными требованиями последующей проверки и обоснования принятых действий [2; 7]. В этих условиях значимым

оказывается не столько объем усвоенной информации, сколько функциональные свойства человеческой памяти, обеспечивающие контекстуальную интерпретацию, временную связность и воспроизводимость управленческих решений [23; 24].

Контекстуальность. Человеческая память тесно связана с контекстом: использование информации зависит от целей деятельности, социальных отношений и конкретных условий взаимодействия. Одни и те же сведения могут иметь различный смысл в зависимости от административного контекста — например, в зависимости от того, идет ли речь о сроке, установленном нормативным актом, сроке исполнения служебного поручения или сроке хранения документа. Контекстуальная организация памяти позволяет соотносить факты не изолированно, а в рамках конкретной управленческой ситуации, что имеет существенное значение для правоприменительной практики и межведомственного взаимодействия [24].

Реконструктивный характер. Память человека не функционирует как точное хранилище данных; воспоминания формируются в процессе реконструкции на основе фрагментов информации, когнитивных схем и обобщенных знаний [9; 21]. В практике публичного управления это может проявляться в необходимости воссоздания целостной картины дела при распределенности информации по различным источникам (письма, протоколы, отчеты). Реконструктивность памяти может сопровождаться риском искажения, однако одновременно выступает механизмом когнитивной гибкости, позволяющим уточнять интерпретации и корректировать управленческие действия по мере поступления новых данных.

Хронологическая связность. Для государственного управления существенную роль играет способность удерживать и воспроизводить временну́ю последовательность управленческих действий и событий [15; 24; 25]. Профессиональная память субъекта государственного управления обеспечивает связывание фактов с конкретными обстоятельствами времени, места и участников, формируя связное представление о развитии управленческого процесса. Такая временная организованность памяти оказывается значимой в ситуациях служебного контроля, проверок и правовой оценки ранее принятых решений [13].

Ответственность. В государственном управлении память неотделима от ответственности: в процедурах служебного контроля и правовой оценки обычно требуется возможность реконструкции мотивов и оснований. Требование воспроизводимости мотивации управленческих решений является институционально закрепленной характеристикой государственной службы и опирается на особенности человеческой памяти как средства осмысленного действия в нормативной среде [2; 3].

3. Архитектура памяти LLM-агентов: типы и реализация (как функциональная аналогия человеческой памяти)

Важно различать разные аспекты «памяти» LLM-агентных систем. В процессе обучения большая языковая модель усваивает обобщенные знания о языке и мире, которые закрепляются в ее параметрах и используются при генерации ответов [10; 20; 23]. В этом смысле LLM обладают семантическим компонентом долговременной памяти. Сама LLM не «помнит» прошлые взаимодействия; эпизодическая память реализуется внешними хранилищами и извлечением информации [17; 19]. В результате такие системы приобретают многоуровневую организацию памяти, которая в функциональном отношении может быть сопоставлена с человеческой, однако отличается по своей природе, а также по способам формирования и хранения информации. В рамках данной многоуровневой архитектуры условно выделяются несколько функциональных слоев памяти. В подразделе 3.1 рассматривается кратковременная память LLM-агента, реализуемая через контекстное окно модели, его ограничения и влияние на связность рассуждений. Подраздел 3.2 посвящен

долговременной памяти, в которой разграничиваются семантический, эпизодический и процедурный компоненты, а также анализируются их технические реализации и ограничения в управленческих сценариях.

3.1. Кратковременная память: контекстное окно, его ограничения и роль в рассуждении

Функциональным аналогом кратковременной памяти выступает контекстное окно — ограниченный объем текста, который модель «видит» одновременно [11; 12; 26]. В рамках этого окна агент может поддерживать связное рассуждение, следовать инструкции, удерживать промежуточные выводы и ссылки на части документа. Однако при выходе за пределы окна возникает эффект «потери» ранних фрагментов: либо они исключаются, либо требуют компрессии (сжатых резюме), что влечет риск искажения деталей. Такая память обеспечивает оперативную обработку информации и не предназначена для накопления знаний [11; 16; 26].

В контексте государственного управления это может означать, что при работе с длинными делами (многостраничные материалы, переписка за месяцы) агент неизбежно сталкивается с ограничением кратковременного удержания: он может обработать документ «порциями», но связность результата зависит от качества разбиения, резюмирования и повторной проверки.

3.2. Долговременная память: семантический, эпизодический и процедурный слои

При анализе архитектуры LLM-агентов можно выделить несколько механизмов, обеспечивающих устойчивое использование информации за пределами одного запроса. Такая дифференциация опирается на когнитивные модели человеческой памяти, в которых долговременная память подразделяется на семантическую, эпизодическую и процедурную [22; 23; 24; 25], и используется в современных исследованиях искусственных агентов в адаптированном методологическом виде.

Семантическая память в человеческой когнитивной системе представляет собой совокупность обобщенных знаний о мире: фактов, понятий, категорий и связей между ними, не привязанных к конкретному эпизоду опыта. В LLM-агентах функциональным аналогом такого слоя выступают параметры модели, сформированные в процессе предварительного обучения на больших корпусах текстов. Этот слой обеспечивает способность к обобщению, логическому выводу и применению общих правил, однако не хранит явных ссылок на источники конкретных утверждений, как правило, не позволяет надежно восстановить их происхождение и не гарантирует их актуальность. В контексте государственного управления это обстоятельство обусловливает необходимость внешней верификации знаний и использования обновляемых источников данных.

Эпизодическая память в человеческой системе связана с хранением информации о конкретных событиях и взаимодействиях, привязанных к временному и ситуационному контексту. В LLM-агентах приближение к такому типу памяти реализуется преимущественно за счет внешних механизмов, в частности подходов Retrieval-Augmented Generation (RAG), при которых перед генерацией ответа осуществляется извлечение релевантных документов из поисковых индексов или векторных баз данных [17]. В управленческих сценариях такие эпизоды могут включать обращения граждан, служебную переписку, протоколы и ранее принятые решения. В LLM-агентах эпизодический слой нередко используется для персонализации и хранения истории взаимодействий. Исследования, подобные MemoryBank, предлагают механизмы отбора и обновления сохраняемых эпизодов, моделируя ограниченность и избирательность долговременного хранения информации [27].

При этом эмпирические данные показывают, что устойчивость такого рода «долгосрочной памяти» остается ограниченной: способность LLM корректно использовать информацию из прошлых взаимодействий снижается с увеличением временной дистанции, а простое накопление эпизодов не всегда приводит к улучшению качества воспроизведения [16].

Процедурная память в когнитивной психологии связана со способностью выполнять действия и следовать устойчивым способам решения задач. В LLM-агентах данный тип памяти не представлен в виде самостоятельного хранилища, а проявляется через устойчивые паттерны обработки инструкций и рассуждений. К таким механизмам относятся дообучение на примерах корректных действий, использование системных подсказок и шаблонов рассуждений, а также интеграция инструментов планирования и исполнения. Важно отметить, что в отличие от человеческой процедурной памяти, формирующейся через практическое взаимодействие с миром, процедурные компоненты LLM-агентов имеют преимущественно текстовую природу и выражаются в повторяемости стратегий обработки информации, а не в освоении действий в физической или социальной среде.

4. Сравнительный анализ и ключевые ограничения памяти LLM-агентов

Сопоставление человеческой памяти и «памяти» LLM-агентов уместно рассматривать как условную аналогию. Она описывает не тождество механизмов, а сходство функций, которые инженерно воспроизводятся в прикладных системах.

В этом смысле оперативное удержание информации в контексте запроса можно соотнести с кратковременной памятью; знания, закрепленные в параметрах модели, — с семантической; внешние хранилища и извлечение документов (например, в RAG подходах) — с эпизодической; устойчивые стратегии выполнения задач и следования инструкциям — с процедурной памятью. Эта аналогия полезна как язык описания архитектуры и как способ определить требования к надежности.

Вместе с тем данная аналогия имеет принципиальные ограничения, критически важные для государственного управления. Прежде всего, у LLM-агентов **отсутствует истинная автобиографичность**. Человеческая эпизодическая память структурно организована вокруг биографии субъекта и фиксирует события как лично пережитые и временно локализованные. У LLM-агента нет внутреннего «Я» и непрерывного переживаемого опыта; его «эпизоды» — это записи во внешнем хранилище, которые могут быть неполными, ошибочно извлеченными или неправильно интерпретированными.

Существенным расхождением является и **слабая встроенная хронология**. Человек удерживает временную ось событий и способен отличать «до» и «после» не только по текстовым маркерам, но и через причинно-следственные связи и социальные контексты. У LLM-агента временная структура должна быть явно задана данными и правилами извлечения; иначе «память» становится набором фрагментов без гарантированного порядка. Именно поэтому исследования долгосрочных диалогов выделяют деградацию воспроизведения прошлых фактов во времени и необходимость специальных механизмов «хронологической памяти» [16].

Также стоит отметить **проблему причинности и обоснования**. В управлении важно не только «вспомнить факт», но и объяснить основания выводов и действий. Модель может сгенерировать правдоподобное объяснение без гарантии, что оно следует из достоверного источника. Это связано с природой генеративных моделей, ориентированных на статистически правдоподобное продолжение текста, а не на внутреннюю проверку истинности каждого утверждения [11].

Наконец, в прикладном контексте особое значение приобретает феномен «**галлюцинаций памяти**». В прикладном смысле это ситуация, когда агент уверенно

сообщает «воспоминание» (например, о якобы полученном письме или ранее принятом решении), которого нет в материалах или оно извлечено неверно. Для государственного управления это не техническая мелочь, а риск управленческой ошибки: ложная «память» может стать основанием неправильного действия [9; 18; 21].

Итоговое сравнение памяти человека и LLM-агентов по различным критериям представлено в табл. 1.

Таблица 1 не предполагает прямого сопоставления «эквивалентных» механизмов памяти человека и LLM-агентов. Критерии сравнения фиксируют ключевые функции памяти, значимые для управленческих процедур. Сопоставление показывает ограничения памяти LLM-агентов и необходимость их компенсации внешними механизмами контроля и проверки.

Таблица 1

Сравнительная характеристика памяти человека и памяти LLM-агентов

Table 1. Comparative characteristics of human memory and LLM-agent memory

Критерий сравнения	Память человека	Память LLM-агента
Природа памяти	Когнитивная, связанная с сознанием и личным опытом	Инженерная, реализованная через параметры модели и внешние хранилища
Автобиографичность	Имеет устойчивую связь с биографией субъекта и пережитым опытом	Отсутствует; «эпизоды» представляют собой записи без личного переживания
Хронологическая организация	Формируется естественным образом, включает причинно-следственные связи	Не гарантирована; требует явного задания временных меток и правил извлечения
Кратковременное удержание информации	Гибкое, адаптивное, зависит от контекста деятельности	Ограничено размером контекстного окна
Семантическая память	Обновляется через опыт и обучение	Закреплена в параметрах модели и может быть неактуальной
Эпизодическая память	Связана с конкретными событиями, местом, временем и участниками	Реализуется через внешние базы данных и механизмы извлечения
Процедурная память	Формируется через практику и повторение действий	Выражена в устойчивых шаблонах обработки инструкций
Проверяемость источников	Возможность указать основания и обстоятельства принятия решений	Ограниченная; возможны ответы без явного источника
Ошибки памяти	Искажения возможны, но поддаются осознанию и корректировке	Возможны «галлюцинации памяти», не отличимые системой от достоверных сведений
Связь с ответственностью	Непосредственно связана с личной и институциональной ответственностью	Ответственность отсутствует, лежит на пользователе системы

Источник: составлено авторами.

4.1. Практические последствия для государственного управления

Выявленные ограничения памяти LLM-агентов имеют прикладное значение не сами по себе, а через то, как они проявляются в типичных управленческих процедурах: при подготовке решений, их согласовании, последующей проверке и работе с персональными данными. Ниже представлены ключевые последствия, которые прямо следуют из особенностей контекстного, параметрического и эпизодического слоев памяти.

Хрупкость контекста. В реальных управленческих процессах задачи редко укладываются в короткую цепочку: есть согласования, уточнения, переговоры, промежуточные поручения. Ограничение контекстного окна и необходимость резюмирования повышают вероятность потери деталей.

Слабая верификация источников. Если ответ опирается на параметрические знания, модель, как правило, не может указать конкретный источник утверждения. Для административной логики это принципиально: обоснование решения должно опираться на норму и документ, а не на правдоподобную формулировку. Частично проблему решают RAG-подходы, поскольку они могут предоставлять цитируемые фрагменты, но тогда надежность зависит от качества корпуса, индекса, политики обновления, правил доступа и протоколов аудита.

Этико-правовые риски: ответственность и приватность. Государственное управление предполагает юридически значимые последствия. Если агент предлагает решение, возникает вопрос: кто несет ответственность за ошибку, и можно ли воспроизвести цепочку оснований? Правовые работы по регулированию искусственного интеллекта (ИИ) подчеркивают необходимость ясного распределения ответственности и соблюдения прав граждан [4; 14]. Дополнительно возникает риск приватности: «эпизодическая память» в виде логов, векторных представлений и архивов диалогов может включать персональные данные и служебную информацию, что требует режимов хранения, доступа и удаления.

В совокупности этих факторов можно утверждать: память LLM-агента недостаточна для автономных решений в ответственных процедурах; допустима роль ассистента под контролем человека.

5. Перспективы и модели интеграции LLM-агентов в государственное управление

Перспективы интеграции LLM-агентов в практику государственного управления во многом определяются архитектурными особенностями их памяти и выявленными ограничениями ее надежности. Современная инженерная траектория связана с улучшением извлечения, ранжирования, компрессии контекста, введением механизмов обновления и «забывания», а также усилением хронологической организации эпизодов (в том числе через специализированные датасеты и метрики долгосрочной памяти) [16; 27]. Для публичного сектора важно, чтобы эти улучшения сопровождались протоколами доказуемости: логированием источников, неизменяемыми журналами обращений к данным, формальными политиками доступа.

Анализ практических примеров показывает, что на текущем этапе интеграция больших языковых моделей в управленческие процессы реализуется преимущественно в форме вспомогательных сервисов и инструментов поддержки, а не в виде полноценной замены субъекта принятия решений. В России обсуждается и тестируется использование решений на базе LLM для консультаций граждан при взаимодействии с государственными цифровыми сервисами. Так, федеральные органы власти и Минцифры РФ заявляли о планах интеграции отечественных языковых моделей, основанных на архитектурах класса GPT, в функционал портала

«Госуслуги» для консультирования пользователей и помощи при заполнении заявлений. Соответствующие решения проходят стадию экспериментального тестирования в формате цифровых ассистентов[1].

Еще одним заметным кейсом является применение генеративной языковой модели GigaChat[2] для обработки и анализа входящих обращений граждан в рамках публичных коммуникаций, таких как «прямая линия» с участием высшего государственного руководства. В данном сценарии ИИ-система используется для первичной систематизации, категоризации и маршрутизации большого массива обращений[3].

Помимо государственных сервисов, аналогичные технологии внедряются в корпоративной среде и системах электронного документооборота, где LLM применяются для подготовки текстов, классификации документов и поиска по крупным корпусам документов [11; 12; 17]. Эти практики демонстрируют расширение применения LLM-инструментов в операционных и вспомогательных управленческих задачах как в государственном, так и в частном секторе.

Анализ перечисленных кейсов позволяет выделить общий принцип их использования: во всех рассмотренных сценариях LLM-агенты выступают в роли ассистента, тогда как утверждение решений и ответственность за их последствия сохраняются за человеком как субъектом управления. В прикладном плане это выражается в устойчивом наборе ассистентских сценариев, в которых LLM-агенты дополняют деятельность государственного служащего, не подменяя его функций.

К числу наиболее реалистичных сценариев использования LLM-агентов в государственном управлении относятся:

1) анализ документов (выделение сущностей, сроков, противоречий, подготовка кратких справок);

2) подготовка проектов ответов по типовым обращениям с обязательной юридической проверкой экспертом-человеком;

3) контроль процедур (чек-листы по регламентам, напоминание о необходимых согласованиях, сопоставление реквизитов);

4) поиск по корпусам (RAG-доступ к нормативным актам, внутренним письмам и методическим материалам при наличии правовых оснований доступа).

В указанных сценариях агент работает как «когнитивный усилитель», а окончательное решение и подпись остаются за человеком. Такая постановка согласуется с акцентом исследований на осторожном внедрении и оценке рисков использования LLM-агентов в государственном управлении [1; 6].

5.1. Условия безопасного внедрения LLM-агентов

Безопасная интеграция LLM-агентов в практику государственного управления предполагает выполнение ряда институциональных и технических условий, непосредственно вытекающих из выявленных ограничений их памяти.

Базовый набор условий, вытекающих из анализа.

1. **Четкое функционально-правовое разграничение ролей** между человеком и LLM-агентом, при котором агент используется для подготовки вариантов, поиска

---

[1] Интеграция LLM-модели на «Госуслуги» до конца года [Электронный ресурс] // Open Vision. 2024. URL: https://www.open-vision.ru/about/news/integracziya-llm-modeli-na-gosuslugi-do-koncza-goda; В Госдуме ответили на вопрос о ChatGPT на «Госуслугах» [Электронный ресурс] // РИА Новости. 2024. URL: https://ria.ru/20240201/chatgpt-1924878904.html (дата обращения: 05.01.2026).

[2] Описание моделей GigaChat [Электронный ресурс] // Sber Developers. 2025. URL: https://developers.sber.ru/docs/ru/gigachat/models (дата обращения: 09.01.2026).

[3] Путин подтвердил применение GigaChat в обработке обращений граждан на прямой линии [Электронный ресурс] // ТАСС. 2024. URL: https://tass.ru/politika/25672459 (дата обращения: 05.01.2026).

и структурирования информации, тогда как утверждение решений, их правовая фиксация и ответственность за последствия их принятия остаются за человеком как субъектом государственного управления.

2. **Прозрачность работы агента.** Это предполагает хранение ссылок на использованные источники, протоколирование операций извлечения документов и возможность аудита изменений в эпизодических хранилищах. Прозрачность памяти системы выступает предпосылкой проверяемости и подотчетности управленческих решений.

3. Соблюдение **режимов защиты данных и приватности**. Эпизодическая память агента должна проектироваться с учетом минимизации персональных данных, ограничений сроков хранения, разграничения доступа и формализованных процедур удаления информации.

4. **Практическая валидация систем** на типовых управленческих сценариях. Такая валидация должна включать оценку характерных ошибок памяти: потери контекста, ложных «воспоминаний», нарушений хронологии, а также использоваться как инструмент настройки и ограничения областей применения агента.

6. Заключение

В данной работе проведен сравнительный анализ моделей памяти человека и LLM-агентов для анализа потенциала применения последних в государственном управлении. Представлены модели памяти человека, а также указаны критические для государственного управления свойства этой памяти: контекстуальность, реконструктивный характер, хронологическая связность, наличие ответственности.

В противовес ей — с инженерной стороны — рассмотрена память LLM-агентов как аналогия человеческой памяти. Сравнительный анализ показал, что принципиальными расхождениями между двумя видами памяти являются присущие памяти LLM-агентов отсутствие истинной автобиографичности, слабая встроенная хронология, проблемы причинности и обоснования, а также галлюцинации в «воспоминаниях».

Завершающая часть статьи посвящена раскрытию основного вывода, заключающегося в том, что LLM-агенты могут быть мощными инструментами-ассистентами в государственном управлении, однако их интеграция должна строиться на понимании фундаментальных ограничений их памяти, а не на поверхностных аналогиях с человеческим интеллектом.

Дальнейшие исследования могут быть направлены на создание хронологической долговременной памяти, усиление механизмов верификации используемых источников при ответах LLM-агентов, а также на развитие систем аудита хранилищ, модели подотчетности ИИ человеку.

Литература

1. *Кабанова Е. Е.* Искусственный интеллект в государственном управлении: ключевые проблемы и перспективы применения // Russian Journal of Management. 2025. Т. 13, № 2. С. 1–14. DOI: 10.29039/2500-1469-2025-13-2-1-14
2. *Кирилловых А. А.* Цифровое государство (digital state): организационно-правовые аспекты. Москва : Юрлитинформ, 2022. 184 с. ISBN 978-5-4396-2425-6.
3. *Михальченкова Н. А.* Искусственный интеллект в контексте публичного управления // Власть. 2021. Т. 29, № 5. С. 122–127. DOI: 10.31171/vlast.v29i5.8545
4. *Троян Н. А.* Искусственный интеллект и его регулирование в сфере государственного управления: правовые аспекты // Мониторинг правоприменения. 2025. № 3. С. 59–65. EDN SNUECT
5. *Федорченко С. Н.* Искусственный интеллект в сфере политики, медиапространства и государственного управления // Журнал политических исследований. 2020. Т. 4, № 2. С. 3–9. URL: https://naukaru.ru/ru/nauka/article/38587/view

6. *Федорченко С. Н.* Цифровое государство и искусственный интеллект в публичном управлении : монография. Москва : Юрайт, 2021. 256 с. DOI: 10.33910/2707-0144-2021-3-1-7
7. *Черешнева И. А.* Искусственный интеллект в государственном управлении и транспарентность: европейский опыт // Государственная служба (Public Administration). 2022. № 2 (136). С. 80–87. DOI: 10.22394/2070-8378-2022-24-2-80-87
8. *Atkinson R. C., Shiffrin R. M.* Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), The Psychology of Learning and Motivation (Vol. 2, pp. 89–195). New York: Academic Press, 1968. DOI: 10.1016/S0079-7421(08)60422-3
9. *Bartlett F. C.* Remembering: A Study in Experimental and Social Psychology. Cambridge: Cambridge University Press, 1932. 317 p. URL: https://archive.org/details/rememberingstudy0000bart_u8n7
10. *Bender E. M., Gebru T., McMillan-Major A., Shmitchell S.* On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? // Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT). New York, NY, USA: ACM, 2021. P. 610–623. DOI: 10.1145/3442188.3445922
11. *Bommasani R., Hudson D. A., Adeli E.*, et al. On the Opportunities and Risks of Foundation Models. Stanford, CA, USA: Stanford Center for Research on Foundation Models (CRFM), Stanford University, 2021. 212 p. URL: https://arxiv.org/abs/2108.07258
12. *Brown T. B., Mann B., Ryder N.*, et al. Language Models are Few-Shot Learners // Advances in Neural Information Processing Systems (NeurIPS). Red Hook, NY, USA: Curran Associates, Inc., 2020. Vol. 33. P. 1877–1901. URL: https://arxiv.org/abs/2005.14165
13. *Burgess N.* The hippocampus, space and viewpoints in episodic memory // Quarterly Journal of Experimental Psychology A. 2002. 55 (4), P. 1057–1080. DOI: 10.1080/02724980244000224
14. *Cajueiro D. O., Celestino V. R. R.* A comprehensive review of Artificial Intelligence regulation: Weighing ethical principles and innovation // Journal of Economy and Technology. 2026. Vol. 4. P. 77–91. DOI: 10.1016/j.ject.2025.07.001
15. *Conway M. A., Pleydell-Pearce C. W.* The Construction of Autobiographical Memories // Psychological Review. 2000. Vol. 107, No. 2. P. 261–288. DOI: 10.1037/0033-295X.107.2.261
16. *Jia Z., Liu Q., Li H.*, et al. Evaluating the Long-Term Memory of Large Language Models // Findings of the Association for Computational Linguistics (ACL). 2025. 18 p. URL: https://arxiv.org/abs/2402.03300
17. *Lewis P., Perez E., Piktus A.*, et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks // Advances in Neural Information Processing Systems (NeurIPS). Red Hook, NY, USA: Curran Associates, Inc., 2020. Vol. 33. P. 9459–9474. URL: https://arxiv.org/abs/2005.11401
18. *Loftus E. F.* False Memories // American Psychologist. 1995. Vol. 50, No. 9. P. 720–725. DOI: 10.1037/0003-066X.50.9.720
19. *Park J. S., O'Brien J., Cai C. J., Morris M. R., Liang P., Bernstein M. S.* Generative Agents: Interactive Simulacra of Human Behavior // arXiv.org. 2023. 32 p. URL: https://arxiv.org/abs/2304.03442
20. *Petroni F., Rocktäschel T., Riedel S., Lewis P., Bakhtin A., Wu Y., Miller A.* Language Models as Knowledge Bases? // Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong: Association for Computational Linguistics, 2019. P. 2463–2473. DOI: 10.18653/v1/D19-1250
21. *Schacter D. L.* The Seven Sins of Memory: Insights from Psychology and Cognitive Neuroscience // American Psychologist. 1999. Vol. 54, No. 3. P. 182–203. DOI: 10.1037/0003-066X.54.3.182
22. *Squire L. R.* Mechanisms of Memory // Science. 1986. Vol. 232, No. 4758. P. 1612–1619. DOI: 10.1126/science.3086978
23. *Squire L. R., Zola S. M.* Structure and Function of Declarative and Nondeclarative Memory Systems // Proceedings of the National Academy of Sciences of the United States of America. 1996. Vol. 93, No. 24. P. 13515–13522. DOI: 10.1073/pnas.93.24.13515
24. *Tulving E.* Elements of Episodic Memory. Oxford: Oxford University Press, 1983. 351 p. URL: https://archive.org/details/elementsofepisod0000tulv
25. *Tulving E.* Episodic and Semantic Memory // Organization of Memory / Eds. E. Tulving, W. Donaldson. New York: Academic Press, 1972. P. 381–403. DOI: 10.1016/B978-0-12-702050-2.50008-4
26. *Vaswani A., Shazeer N., Parmar N.*, et al. Attention Is All You Need // Advances in Neural Information Processing Systems (NeurIPS). Red Hook, NY, USA: Curran Associates, Inc., 2017. Vol. 30. P. 5998–6008. URL: https://arxiv.org/abs/1706.03762
27. *Zhong W., Yu S., Wu C.*, et al. MemoryBank: Enhancing Large Language Models with Long-Term Memory // Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, CA, USA: AAAI Press, 2024. URL: https://arxiv.org/abs/2305.10260.

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

Об авторах:

Вейбер Евгения Николаевна, младший научный сотрудник, Российская академия наук, Санкт-Петербургский Федеральный исследовательский центр (Санкт-Петербург, Российская Федерация); vejber2013@gmail.com

Грудинин Михаил Артёмович, стажер-исследователь, Российская академия наук, Санкт-Петербургский Федеральный исследовательский центр (Санкт-Петербург, Российская Федерация); st119027@student.spbu.ru

Тулупьева Татьяна Валентиновна, кандидат психологических наук, доцент, Почетный работник сферы образования РФ, советник проректора по науке, Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации (Москва, Российская Федерация); ведущий научный сотрудник, лаборатория Прикладного искусственного интеллекта, Российская академия наук, Санкт-Петербургский Федеральный исследовательский центр (Санкт-Петербург, Российская Федерация); tvt@dscs.pro

Вяткин Артём Андреевич, младший научный сотрудник, Российская академия наук, Санкт-Петербургский Федеральный исследовательский центр (Санкт-Петербург, Российская Федерация); aav@dscs.pro

References

1. Kabanova E. E. Artificial intelligence in public administration: key issues and prospects of application // Russian Journal of Management. 2025. Vol. 13, No. 2, P. 1–14. (In Russ.). DOI: 10.29039/2500-1469-2025-13-2-1-14
2. Kirillovykh A. A. Digital State: Organizational and Legal Aspects. Moscow: Yurlitinform, 2022. 184 p. (In Russ.). ISBN 978-5-4396-2425-6.
3. Mikhalchenkova N. A. Artificial intelligence in the context of public administration // Power [Vlast']. 2021. Vol. 29, No. 5. P. 122–127. (In Russ.). DOI: 10.31171/vlast.v29i5.8545
4. Troyan N. A. Artificial intelligence and its regulation in public administration: legal aspects // Monitoring of Law Enforcement [Monitoring pravoprimeneniya]. 2025. No. 3. P. 59–65. (In Russ.). EDN SNUECT
5. Fedorchenko S. N. Artificial intelligence in politics, the media space and public administration // Journal of Political Research [Zhurnal politicheskikh issledovanii]. 2020. Vol. 4, No. 2. P. 3–9. (In Russ.). URL: https://naukaru.ru/ru/nauka/article/38587/view
6. Fedorchenko S. N. Digital State and Artificial Intelligence in Public Administration: Monograph. Moscow: Yurait, 2021. 256 p. (In Russ.). DOI: 10.33910/2707-0144-2021-3-1-7
7. Chereshneva I. A. Artificial intelligence in public administration and transparency: European experience. // Public Administration [Gosudarstvennaya sluzhba]. 2022. No. 2 (136). P. 80–87. (In Russ.). DOI: 10.22394/2070-8378-2022-24-2-80-87
8. Atkinson R. C., Shiffrin R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), The Psychology of Learning and Motivation (Vol. 2, pp. 89–195). New York: Academic Press, 1968. DOI: 10.1016/S0079-7421(08)60422-3
9. Bartlett F. C. Remembering: A Study in Experimental and Social Psychology. Cambridge: Cambridge University Press, 1932. 317 p. URL: https://archive.org/details/rememberingstudy0000bart_u8n7
10. Bender E. M., Gebru T., McMillan-Major A., Shmitchell S. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? // Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT). New York, NY, USA: ACM, 2021. P. 610–623. DOI: 10.1145/3442188.3445922
11. Bommasani R., Hudson D. A., Adeli E., et al. On the Opportunities and Risks of Foundation Models. Stanford, CA, USA: Stanford Center for Research on Foundation Models (CRFM), Stanford University, 2021. 212 p. URL: https://arxiv.org/abs/2108.07258
12. Brown T. B., Mann B., Ryder N., et al. Language Models are Few-Shot Learners // Advances in Neural Information Processing Systems (NeurIPS). Red Hook, NY, USA: Curran Associates, Inc., 2020. Vol. 33. P. 1877–1901. URL: https://arxiv.org/abs/2005.14165
13. Burgess N. The hippocampus, space and viewpoints in episodic memory // Quarterly Journal of Experimental Psychology A. 2002. 55 (4), P. 1057–1080. DOI: 10.1080/02724980244000224
14. Cajueiro D. O., Celestino V. R. R. A comprehensive review of Artificial Intelligence regulation: Weighing ethical principles and innovation // Journal of Economy and Technology. 2026. Vol. 4. P. 77–91. DOI: 10.1016/j.ject.2025.07.001

15. Conway M. A., Pleydell-Pearce C. W. The Construction of Autobiographical Memories // Psychological Review. 2000. Vol. 107, No. 2. P. 261–288. DOI: 10.1037/0033-295X.107.2.261
16. Jia Z., Liu Q., Li H., et al. Evaluating the Long-Term Memory of Large Language Models // Findings of the Association for Computational Linguistics (ACL). 2025. 18 p. URL: https://arxiv.org/abs/2402.03300
17. Lewis P., Perez E., Piktus A., et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks // Advances in Neural Information Processing Systems (NeurIPS). Red Hook, NY, USA: Curran Associates, Inc., 2020. Vol. 33. P. 9459–9474. URL: https://arxiv.org/abs/2005.11401
18. Loftus E. F. False Memories // American Psychologist. 1995. Vol. 50, No. 9. P. 720–725. DOI: 10.1037/0003-066X.50.9.720
19. Park J. S., O'Brien J., Cai C. J., Morris M. R., Liang P., Bernstein M. S. Generative Agents: Interactive Simulacra of Human Behavior // arXiv.org. 2023. 32 p. URL: https://arxiv.org/abs/2304.03442
20. Petroni F., Rocktäschel T., Riedel S., Lewis P., Bakhtin A., Wu Y., Miller A. Language Models as Knowledge Bases? // Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong: Association for Computational Linguistics, 2019. P. 2463–2473. DOI: 10.18653/v1/D19-1250
21. Schacter D. L. The Seven Sins of Memory: Insights from Psychology and Cognitive Neuroscience // American Psychologist. 1999. Vol. 54, No. 3. P. 182–203. DOI: 10.1037/0003-066X.54.3.182
22. Squire L. R. Mechanisms of Memory // Science. 1986. Vol. 232, No. 4758. P. 1612–1619. DOI: 10.1126/science.3086978
23. Squire L. R., Zola S. M. Structure and Function of Declarative and Nondeclarative Memory Systems // Proceedings of the National Academy of Sciences of the United States of America. 1996. Vol. 93, No. 24. P. 13515–13522. DOI: 10.1073/pnas.93.24.13515
24. Tulving E. Elements of Episodic Memory. Oxford: Oxford University Press, 1983. 351 p. URL: https://archive.org/details/elementsofepisod0000tulv
25. Tulving E. Episodic and Semantic Memory // Organization of Memory / Eds. E. Tulving, W. Donaldson. New York: Academic Press, 1972. P. 381–403. DOI: 10.1016/B978-0-12-702050-2.50008-4
26. Vaswani A., Shazeer N., Parmar N., et al. Attention Is All You Need // Advances in Neural Information Processing Systems (NeurIPS). Red Hook, NY, USA: Curran Associates, Inc., 2017. Vol. 30. P. 5998–6008. URL: https://arxiv.org/abs/1706.03762
27. Zhong W., Yu S., Wu C., et al. MemoryBank: Enhancing Large Language Models with Long-Term Memory // Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, CA, USA: AAAI Press, 2024. URL: https://arxiv.org/abs/2305.10260.

Conflict of interests
The authors declare no relevant conflict of interests.

About the authors:
Evgeniya N. Veiber, Junior Researcher, Saint Petersburg Federal Research Center of the Russian Academy of Sciences (Saint Petersburg, Russian Federation); vejber2013@gmail.com
Mikhail A. Grudinin, Trainee researcher, Saint Petersburg Federal Research Center of the Russian Academy of Sciences (Saint Petersburg, Russian Federation); st119027@student.spbu.ru
Tatiana V. Tulupyeva, PhD in Psychology, Associate Professor, Honorary Worker of the Russian Federation in the Sphere of Education, Advisor to the Vice-Rector for Research, Russian Presidential Academy of National Economy and Public Administration (Moscow, Russian Federation); Leading researcher at the Laboratory of Applied Artificial Intelligence at the St. Petersburg Federal Research Center of the Russian Academy of Sciences (Saint Petersburg, Russian Federation); tvt@dscs.pro
Artem A. Vyatkin, Junior Researcher, Saint Petersburg Federal Research Center of the Russian Academy of Sciences (Saint Petersburg, Russian Federation); aav@dscs.pro

Поступила в редакцию: 28.10.2025
Поступила после рецензирования: 12.12.2025
Принята к публикации: 20.03.2026

The article was submitted: 28.10.2025
Approved after reviewing: 12.12.2025
Accepted for publication: 20.03.2026

© Вейбер Е. Н., Грудинин М. А., Тулупьева Т. В., Вяткин А. А., 2026