МРНТИ: 82.13.11; 28.23.01 DOI: https://doi.org/10.62687/ESM.3.2.2026.85

ЦИФРОВАЯ РЕПУТАЦИЯ ГОСУДАРСТВА В УСЛОВИЯХ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: МОДЕЛЬ МЕХАНИЗМА ОБРАТНОЙ СВЯЗИ И ЭПИСТЕМИЧЕСКИЕ РИСКИ

[1]Б.Б. Кожа-Ахмет*, [2]С.А. Рахимова, [3]Э.Г. Матюгина
[1,2]Международный университет Астана, г. Астана, Казахстан
[3]Томский государственный университет, г. Томск, Российская Федерация
*e-mail: bagdad.pr@gmail.com

Б.Б. Кожа-Ахмет – докторант Высшей школы экономики, Международный университет Астана, г. Астана, Казахстан, e-mail: bagdad.pr@gmail.com, https://orcid.org/0009-0008-3701-5958

С.А. Рахимова – профессор, кандидат экономических наук, Международный университет Астана, г. Астана, Казахстан, e-mail: Rakhimova.saule.econ@mail.ru, https://orcid.org/0000-0003-0553-9606

Э.Г. Матюгина – профессор, доктор экономических наук, Томский государственный университет, г. Томск, Российская Федерация, e-mail: emk512542@mail.ru, https://orcid.org/0000-0001-9788-8672

Аннотация. В статье рассматривается, как использование искусственного интеллекта влияет на формирование цифровой репутации государства. Предлагается модель Digital Reputation Loop: созданная алгоритмом интерпретация цифровых данных попадает в публичную коммуникацию, способна повлиять на мнения и действия людей, а возникшие вследствие этого цифровые следы вновь используются для анализа. Чтобы описать этот процесс, авторы разграничивают человеческое восприятие репутации, ее цифровое выражение и оценку, сформированную ИИ. Эмпирическая часть — это вторичный анализ опроса 217 человек из государственных и смежных сфер Казахстана, проведённого 17–21 августа 2026 года. Опрос не измерял репутацию напрямую, но показал отношение к ИИ. 71,4 % считают ИИ-ассистента полезным или очень полезным, но только 49,8 % доверяют его рекомендациям. Доверие к рекомендациям показывает готовность полагаться на ИИ, а не измеряет эпистемическую легитимность. 77,4 % считают объяснимость обязательной или желательной, 87,1 % поддерживают проверку человеком, 88,0 % хотят, чтобы окончательное решение принимал человек. Мы обнаружили, что восприятие полезности ИИ связано с доверием ($\rho = 0{,}490$; $p < 0{,}001$) и поддержкой внедрения ($\rho = 0{,}606$; $p < 0{,}001$). Поддержка внедрения тоже связана с поддержкой человеческого контроля ($\rho = 0{,}411$; $p < 0{,}001$). Авторы делают вывод о риске самоподтверждающейся репутационной оценки, если алгоритмические выводы не сопоставлять с независимыми данными о результатах работы государства и мнениями людей.

Ключевые слова: цифровая репутация, искусственный интеллект, государственные коммуникации, алгоритмическая медиация, общественное доверие, эпистемическая легитимность, human oversight, AI governance, public value, механизм обратной связи.

DIGITAL REPUTATION OF THE STATE IN THE CONTEXT OF ARTIFICIAL INTELLIGENCE: THE FEEDBACK MECHANISM AND EPISTEMIC RISKS

[1]B.B. Kozha-Akhmet*, **[2]S.A. Rakhimova**, **[3]E.G. Matyugina**
[1,2]International University Astana, Astana, Republic of Kazakhstan
[3]Tomsk State University, Tomsk, Russian Federation
*e-mail: bagdad.pr@gmail.com

B.B. Kozha-Akhmet – PhD Student, Graduate School of Economics, International University Astana, Astana, Kazakhstan, e-mail: bagdad.pr@gmail.com, https://orcid.org/0009-0008-3701-5958
S.A. Rakhimova – Professor, Candidate of Economic Sciences, International University Astana, Astana, Kazakhstan, e-mail: Rakhimova.saule.econ@mail.ru, https://orcid.org/0000-0003-0553-9606
E.G. Matyugina – Professor, Doctor of Economics, Tomsk State University, Tomsk, Russian Federation, e-mail: emk512542@mail.ru, https://orcid.org/0000-0001-9788-8672

Abstract. This article considers how artificial intelligence may affect the formation of a state's digital reputation. It proposes the Digital Reputation Loop: an algorithmic interpretation of digital evidence enters public communication, may influence people's views and behaviour, and thereby contributes to the data used in the next assessment. The paper distinguishes human perceptions of reputation, their digital expression, and AI-generated assessments. It also sets out the Human–AI Reputation Gap, instrumental and epistemic legitimacy, and Human Verification of Reputational Reality. The empirical section reanalyses a survey of 217 respondents in Kazakhstan's public and related sectors, conducted on 17–21 August 2026. The survey did not directly measure state reputation; it offers evidence only about some organisational conditions for the use of AI. An AI assistant was considered useful or very useful by 71.4% of respondents, whereas 49.8% expressed direct trust in AI recommendations. The latter is an indicator of willingness to rely on recommendations, not a validated measure of epistemic legitimacy. Explainability was considered mandatory or desirable by 77.4%; human review was supported by 87.1%, and 88.0% wanted final decisions to remain with humans. Perceived benefits were positively associated with direct trust ($\rho=0.490$; $p<0.001$) and support for adoption ($\rho=0.606$; $p<0.001$). Human–AI Governance was also associated with adoption support ($\rho=0.411$; $p<0.001$). The paper concludes that an AI-based reputation assessment can become self-reinforcing unless it is checked against independent evidence on institutional performance and human perceptions.

Keywords: digital reputation, artificial intelligence, government communication, algorithmic mediation, public trust, epistemic legitimacy, human oversight, AI governance, public value, feedback mechanism.

ЖАСАНДЫ ИНТЕЛЛЕКТ ЖАҒДАЙЫНДАҒЫ МЕМЛЕКЕТТІҢ ЦИФРЛЫҚ БЕДЕЛІ: КЕРІ БАЙЛАНЫС ТЕТІГІ ЖӘНЕ ТАНЫМДЫҚ ТӘУЕКЕЛДЕР

[1]Б.Б. Қожа-Ахмет*, **[2]С.А. Рахимова**, **[3]Э.Г. Матюгина**
[1,2]Астана халықаралық университеті, Астана қ., Қазақстан
[3]Томск мемлекеттік университеті, Томск қаласы, Ресей Федерациясы
*e-mail: bagdad.pr@gmail.com

Б.Б. Қожа-Ахмет – докторант, Экономика жоғары мектебі, Астана халықаралық университеті, Астана қ., Қазақстан, e-mail: bagdad.pr@gmail.com, https://orcid.org/0009-0008-3701-5958
С.А. Рахимова – профессор, экономика ғылымдарының кандидаты, Астана халықаралық университеті, Астана қ., Қазақстан, e-mail: Rakhimova.saule.econ@mail.ru, https://orcid.org/0000-0003-0553-9606
Э.Г. Матюгина – профессор, экономика ғылымдарының докторы, Томск мемлекеттік университеті, Томск қаласы, Ресей Федерациясы, e-mail: emk512542@mail.ru, https://orcid.org/0000-0001-9788-8672

Аңдатпа. Мақалада жасанды интеллект мемлекеттің цифрлық беделінің қалыптасуына қалай ықпал етуі мүмкін екені қарастырылады. Авторлар Digital Reputation Loop моделін ұсынады: алгоритм цифрлық деректерді түсіндіреді, бұл түсіндірме қоғамдық коммуникацияға енеді, адамдардың пікірі мен әрекетіне ықпал етуі мүмкін, ал жаңа цифрлық іздер келесі талдауға негіз болады. Адамдардың бедел туралы пікірі, оның цифрлық көрінісі және ЖИ қалыптастырған бағасы өзара ажыратылады. Human–AI Reputation Gap, ЖИ-дің құралдық және эпистемикалық легитимділігі, сондай-ақ Human Verification of Reputational Reality ұғымдары ұсынылады. Эмпирикалық бөлім 2026 жылғы 17–21 тамызда Қазақстанның мемлекеттік және оған жақын секторларындағы 217 адам қатысқан сауалнаманы қайта талдауға негізделген. Сауалнама мемлекет беделін тікелей өлшемеген; оның нәтижелері ЖИ қолданудың кейбір ұйымдық алғышарттарын ғана сипаттайды. Қатысушылардың 71,4%-ы ЖИ-ассистентті пайдалы не өте пайдалы деп бағалады, ал 49,8%-ы оның ұсынымдарына тікелей сенетінін көрсетті. Соңғы көрсеткіш эпистемикалық легитимділіктің тексерілген өлшемі емес, ұсынымдарға сүйенуге дайындықтың индикаторы ретінде қарастырылады. Түсіндірмелілікті 77,4%, адам тексеруін 87,1%, соңғы шешімнің адамда қалуын 88,0% қолдады. Қорытындыда алгоритм жасаған бедел бағасын мемлекеттік қызмет нәтижелері мен адам пікірін тәуелсіз өлшеу арқылы тексеру қажет екені көрсетіледі.

Түйін сөздер: цифрлық бедел, жасанды интеллект, мемлекеттік коммуникациялар, алгоритмдік медиация, қоғамдық сенім, эпистемикалық легитимділік, адам бақылауы, AI governance, public value, кері байланыс тетігі.

Введение. В цифровой среде восприятие деятельности государства всё в большей степени формируется посредством алгоритмически отобранной и интерпретированной информации. Поисковые и рекомендательные алгоритмы решают, какие новости увидит человек, а генеративный ИИ собирает их и выдаёт готовый смысл(Gillespie, 2014, p. 167). Чтобы понять репутацию, нужны не только реальные действия государства, но и то, как граждане их воспринимают, что пишут официальные источники, а в сети добавляются медиа-материалы и следы поведения пользователей.

Если упростить, государство делает сообщения, люди их читают и формируют мнение. Теперь часть этих реакций фиксируется в цифрах и может быть проанализирована. ИИ делает больше, чем просто искать: он собирает всё, что есть, и формулирует ответ о государстве. Исследования показывают, что ИИ-агент может править или даже писать сообщения от имени государства (Hancock et al., 2020, p. 89).

Эти изменения влияют на управленческие решения. Госслужбы могут брать данные о цифровой репутации, искать проблемы, менять сообщения и проверять работу отделов по KPI. Но если способ получения этих данных меняется, нужно пересмотреть, как их использовать: от диагностики общественной реакции до принятия решений и оценки результата.

В предложенной модели Digital Reputation Loop госорган сначала получает алгоритмическую оценку информационной среды, потом меняет коммуникацию, учитывая эту оценку. Люди реагируют, их реакция создаёт новые цифровые данные, система снова оценивает репутацию. При таком подходе показатель может частично отражать эффект от предыдущих настроек коммуникации, а не только независимое изменение общественного мнения.

Вопрос, который стоит перед нами, прост: когда оценка алгоритма перестаёт лишь описывать репутацию и начинает её менять? Если вывод ИИ определяет, что будет в сообщении, а потом это сообщение меняет поведение аудитории, новые данные уже нельзя считать полностью независимыми от предыдущего вывода. Мы назвали такой повторяющийся процесс Digital Reputation Loop.

Наша цель — рассказать, как работает эта обратная связь, какие риски она несёт для достоверности знаний и как её проверять. Мы провели опрос в Казахстане, чтобы понять, как ИИ применяется в организациях. Опрос не даёт прямой оценки репутации государства, но помогает увидеть условия использования ИИ.

Для этой цели мы сформулировали четыре исследовательских вопроса:

- Как алгоритмическая интерпретация цифровых данных влияет на связь между мнением людей и оценкой репутации государства?
- Когда оценка ИИ из средства наблюдения превращается в фактор последующих изменений репутации?
- Как различить практическую пользу ИИ и обоснованность доверия к его выводам?
- Какую роль должен играть человек при проверке и использовании алгоритмической оценки?

В статье мы выдвигаем пять теоретических положений:

- Модель Digital Reputation Loop описывает, как оценка ИИ воздействует на последующие данные;
- Человеческая репутация, цифровая репутация и репутация, опосредованная ИИ рассматриваются как отдельные объекты анализа;
- Разрыв в репутации человека и ИИ обозначает расхождение между независимой оценкой людей и оценкой ИИ;
- Практическая полезность ИИ отделяется от обоснованности его выводов;
- Human Verification of Reputational Reality (HVRR) дополняет проверку решения человеком, проверяя, насколько верно ИИ описывает общественное восприятие.

Мы сосредоточились на теоретическом объяснении механизма обратной связи и на условиях, при которых оценку ИИ можно сопоставить с реальными результатами деятельности государства.

Обзор литературы и теоретические основания.

Мы изучили несколько ключевых позиций, чтобы построить основу нашего исследования.

Первая точка — репутация государства, цифровое правительство и public value. Public value говорит, что надо смотреть не только на скорость и объём работы, но и на качество услуг, открытость, справедливость, доверие и влияние на людей. Цифровые сервисы могут сделать услуги лучше и упростить общение с гражданами, но просто наличие сервиса ещё не гарантирует рост доверия. Исследования электронного правительства показывают, что граждане оценивают работу по качеству, прозрачности, защите данных и особенностям институтов. Поэтому то, что видим в интернете, нужно сравнивать с тем, как государство реально работает и как люди это воспринимают.

Gillespie пишет, что алгоритмы в медиа отбирают, что увидит публика (Gillespie, 2014, p. 167). Для государства важны выбор источников, порядок выдачи результатов и способ их объединения. Генеративный ИИ может дать читателю готовый, связный ответ, а не просто список всех материалов, на которых он основан (Hancock et al., 2020, p. 89).

Обзоры о восприятии ИИ в государственном секторе обсуждают пользу технологии, её риски, свойства сервисов, уровень доверия и кто их использует (Azzahro et al., 2025, p. 1). Важно ещё одно: ИИ меняет, как новости о работе государства попадают к людям.

Следующие работы говорят о переходе от простой автоматизации к получению знаний. ИИ в госслужбе используют для автоматизации, прогнозов, выявления рисков и поддержки решений (Madan & Ashok, 2023, p. 1; Wirtz et al., 2019, p. 596). Straub с коллегами выделяют три понятия: operational fitness, epistemic alignment и normative divergence (Straub et al., 2023, p. 1). Оценивая репутацию, важно, чтобы вывод ИИ точно соответствовал тому, о чём он говорит. Цифровой массив охватывает не всех граждан одинаково — на него влияют платформа и язык. Поэтому нельзя просто принимать алгоритмический рейтинг за мнение общества без проверки.

Много исследований задают вопрос о доверии, объяснении и человеческом контроле. Даже если система работает без ошибок, пользователи могут ей не доверять (Kleizen et al., 2023, p. 1; Agbabiaka et al., 2025, p. 1). На оценку ИИ влияют качество данных, понятность объяснений, возможность проверить вывод и оспорить вывод. UNESCO советует связывать прозрачность, прослеживаемость и ответственность с тем, что человек контролирует ИИ (UNESCO, 2021, p. 20). NIST предлагает управлять рисками на всех этапах жизни системы (Tabassi, 2023, p. 1). OECD подчёркивает, что решения должны быть понятными и их можно

оспорить (OECD, 2025, p. 1).

Для репутации нельзя просто дать человеку окончательное слово. Нужно проверять, правильно ли система показывает мнение людей. Механизм HVRR сравнивает вывод алгоритма с данными о том, как люди воспринимают ситуацию.

В работе Strathern рассматривается эффект, при котором показатель теряет часть своей аналитической силы, когда его начинают использовать как цель (Strathern, 1997, p. 314). Например, если ведомство хочет поднять репутационный балл, рассчитанный ИИ, оно может менять сообщения, чтобы метрика понравилась. Такие изменения попадают в цифровые данные, а потом их оценивают, как подтверждение результата. В статье этот сценарий называют Reputation Goodhart Effect.

Исследования о public value, цифровом государстве, внедрении ИИ и коммуникации рассматривают отдельные части процесса (Moore, 1995, p. 28; Twizeyimana & Andersson, 2019, p. 167; Criado & Villodre, 2021, p. 253; Fifaliana, 2026, p. 1; Azzahro et al., 2025, p. 1; Madan & Ashok, 2023, p. 1; Straub et al., 2023, p. 1; OECD, 2025, p. 1; UNESCO, 2021, p. 20; Tabassi, 2023, p. 1; Wirtz et al., 2019, p. 596; Kleizen et al., 2023, p. 1; Agbabiaka et al., 2025, p. 1; Harrison & Luna-Reyes, 2022, p. 494). Чтобы проверить репутационные решения, нужно собрать их в одну управленческую модель: от анализа данных, выбора сообщения, реакции аудитории до новых цифровых следов. Модель Digital Reputation Loop соединяет всё это и помогает понять, когда вывод ИИ можно сравнивать с независимыми данными о работе государства.

Это простая модель Digital Reputation Loop. Как работает обратная связь. В упрощённой линейной схеме после действий государства идут: коммуникация, восприятие людей и оценка репутации:

Обратная связь возникает, если подготовленная ИИ интерпретация влияет на коммуникацию, а ее последствия становятся входными данными следующей оценки. Схема и место независимой проверки показаны на рисунке 1.

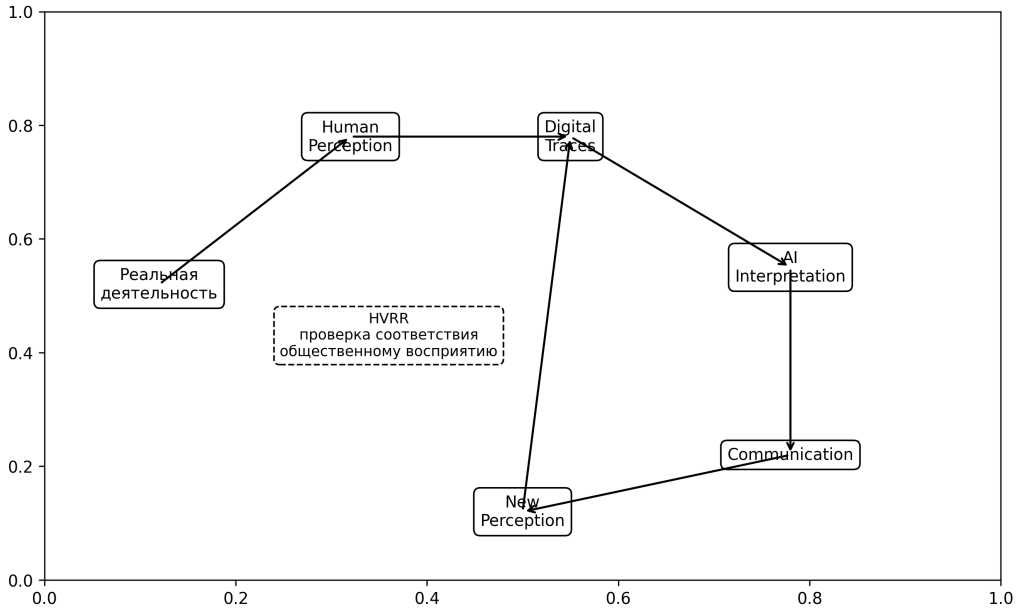


Рисунок 1. Рекурсивная модель цифровой репутации и Human Verification of Reputational Reality

* HVRR - это независимая проверка того, что алгоритм сказал, сравнивая с реальными данными и тем, как люди воспринимают.

Мнение людей, их цифровые следы и то, как ИИ их интерпретирует, часто не совпадают. Различия между тремя уровнями репутации показаны в таблице 1.

Таблица 1. Уровни репутации государства в цифровой и AI-mediated среде

Уровень	Содержание	Источники	Основной риск
Human Reputation	Человеческое восприятие государства	Опросы, интервью, опыт услуг	Подмена мнения цифровыми прокси
Digital Reputation	Цифровое выражение восприятия и поведения	СМИ, соцсети, поиск, публикации	Selection bias, языковые и платформенные эффекты
AI-mediated Reputation	Алгоритмически сформированная интерпретация цифрового массива	LLM, classifiers, recommendation systems	Hallucination, ranking/synthesis bias, feedback amplification

Даже детальный алгоритмический анализ цифровых данных остается интерпретацией доступного информационного массива, который может не полностью отражать общественное восприятие. Поэтому соответствие алгоритмической оценки восприятию людей требует независимой проверки.

Human-AI Reputation Gap – Чтобы сравнить две оценки, берём разницу G = A − H.

H – это то, как люди реально воспринимают тему за определённый период (измерено независимо).

A – это оценка ИИ по тем же параметрам.

Если G сильно отличается, нужно посмотреть, откуда берутся данные, какой у них язык, какие сроки и как их обрабатывают. В этой статье G не считается, потому что в исходной анкете не было нужных цифр для расчёта H и A.

Инструментальная и эпистемическая легитимность. Практическую ценность ИИ можно оценить по тому, помогает ли он сотрудникам быстрее разбирать информацию и готовить сообщения.

Вопрос о достоверности вывода ИИ как отражения общественного мнения сложнее: в исследовании нет проверенной шкалы, которая бы измеряла эту «эпистемическую легитимность».

Ответы о доверии лишь показывают, готовы ли участники полагаться на рекомендации ИИ. Поэтому разница между оценкой полезности (71,4 %) и прямым доверием (49,8 %) описывает отношение респондентов к двум сторонам применения ИИ, но не измеряет эпистемическую легитимность.

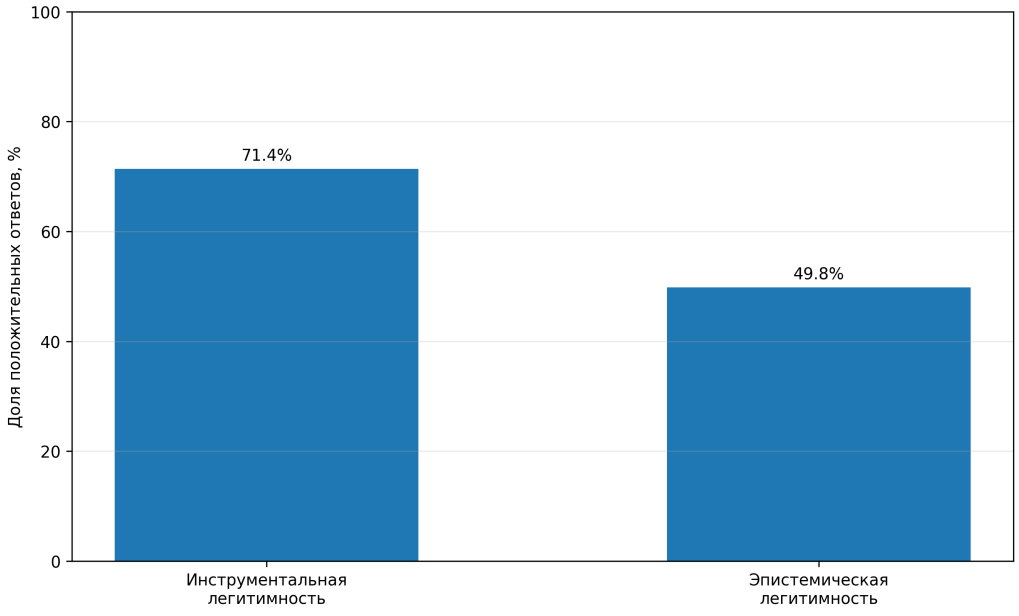


Рисунок 2. Разрыв между воспринимаемой практической полезностью и прямым доверием к рекомендациям ИИ

* Оценка полезности характеризует воспринимаемую практическую ценность ИИ; ответ о прямом доверии отражает готовность использовать его рекомендации. Валидированной шкалы эпистемической легитимности здесь нет; n=217.

При обычной проверке человек может принять или отклонить решение. HARG (Human-AI Reputation Gap) добавляет проверку основания: совпадает ли картина, предложенная ИИ, с независимым изучением восприятия граждан. Для этих двух задач нужны разные данные и процедуры.

Таблица 2. Исследовательская модель, гипотезы и теоретические пропозиции

Гипотеза	Формулировка	Статус
H1	Чем выше воспринимаемая польза ИИ, тем выше прямое доверие к его рекомендациям.	Тестируется
H2	Воспринимаемая польза ИИ положительно связана с поддержкой его внедрения.	Тестируется
H3	Поддержка механизмов человеческого контроля положительно связана с поддержкой внедрения ИИ.	Тестируется
P1	Если алгоритмическая оценка влияет на последующую коммуникацию и реакцию аудитории, новые цифровые данные могут усиливать исходный репутационный сигнал.	Теоретическая
P2	Значительное расхождение между оценками ИИ и людей в сочетании с высокой автономностью системы может повышать риск неверного представления общественного мнения.	Теоретическая

Когда оценки ИИ сильно отличаются от оценок людей и система работает почти самостоятельно, растёт риск неверного представления общественного мнения. Гипотезы H1–H3 можно проверить по данным одного опроса. Положения P1 и P2 относятся к теоретической модели, их здесь не проверяют. Чтобы их оценить, нужны наблюдения во времени или эксперимент.

Материалы и методы. Мы проанализировали ответы 217 человек из государственных и смежных секторов Казахстана. Данные собирали в 2026 году. Выборка не репрезентативна, поэтому нельзя обобщать результаты на всех государственных служащих. Здесь они нужны лишь для описания отношения к ИИ и условий его применения в организациях.

Таблица 3. Структура выборки по типу организации

Тип организации	n	%
Государственная организация/учреждение	85	39,2
Местный исполнительный орган	54	24,9
Квазигосударственная организация	12	5,5
Центральный государственный орган	8	3,7
Другое	58	26,7
Всего	217	100,0

Свою должность в области управления, анализа, коммуникаций или цифровизации указали 143 участника (65,9%). Остальные 74 (34,1%) выбрали категорию «Другое». Ее содержание в исходных данных не раскрыто, поэтому нельзя установить, какие именно должности занимали эти респонденты.

Для составных шкал получены следующие значения α: Communication Complexity — 0,958; Perceived AI Benefits — 0,750; Human-AI Governance — 0,746; Adoption Support — 0,875. У трехпунктовой шкалы Trust значение ниже (α=0,583); поэтому прямое доверие анализируется как отдельный порядковый показатель. Связи между переменными оценивались

коэффициентом Спирмена. Поскольку все ответы получены в один период, установить по ним причинное влияние нельзя.

Таблица 4. Результаты. Опыт использования ИИ

Частота	n	%
Регулярно	72	33,2
Иногда	80	36,9
Редко	38	17,5
Не использую	13	6,0
Только начинаю	14	6,5
Всего	217	100,0

Регулярно используют ИИ 72 человека, время от времени — 80. Вместе это 152 из 217 участников, или 70,0%. Следовательно, опыт работы с ИИ есть у значительной части опрошенных.

Таблица 5. Полезность, доверие и объяснимость

Индикатор	n	%
ИИ-ассистент полезен/очень полезен	155	71,4
Сокращение времени аналитической подготовки	159	73,3
Выявление проблем и сигналов	142	65,4
Повышение качества стратегических коммуникаций	148	68,2
Доверие к рекомендациям ИИ	108	49,8
Наличие источников повышает доверие	121	55,8
Объяснимость обязательна/желательна	168	77,4

Положительную оценку полезности и прямое доверие разделяют 21,6 процентного пункта. Респонденты могут считать ИИ полезным для работы, но при этом не готовы безусловно опираться на его рекомендации. Однако ответ о доверии не является валидированной мерой эпистемической легитимности: для оценки этого теоретического понятия понадобится отдельное исследование.

Таблица 6. Человеческий контроль

Механизм	n	%
Каждая рекомендация проходит человеческую проверку	189	87,1
Окончательное решение остается за человеком	191	88,0
Руководитель может отклонить рекомендацию	156	71,9
Обязательный контроль повышает доверие	157	72,4

217 человек ответили, насколько они поддерживают проверку рекомендаций человеком, право человека принимать окончательное решение и возможность отклонить вывод ИИ.

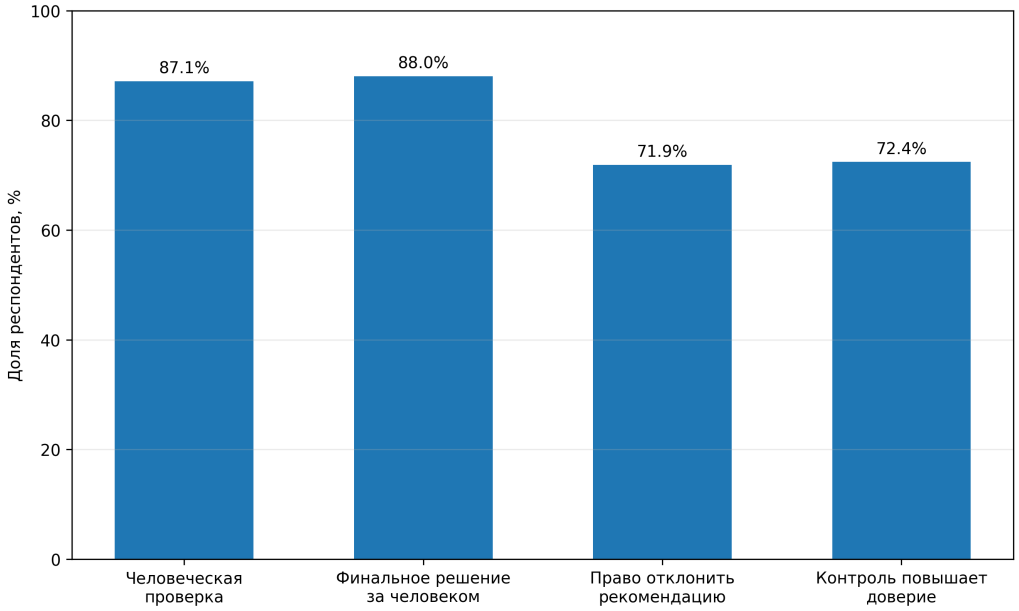


Рисунок 3. Поддержка механизмов человеческого контроля

Ответы показывают, что люди хотят, чтобы человек был в процессе принятия решений с ИИ. Предлагаемая HVRR добавляет проверку самих данных, на которых основана рекомендация.

Таблица 7. Воспринимаемые риски

Риск	Доля, %
Ошибочные рекомендации	53,5
Недостоверная информация	52,1
Утечка персональных данных	50,2
Утечка служебной/конфиденциальной информации	48,4
Чрезмерное доверие сотрудников к ИИ	38,2
Ошибочная оценка общественных настроений	19,8
Репутационный ущерб	18,9
Непрозрачность работы ИИ	17,5

Для нашей модели особенно важны риски недостоверной информации и неверной оценки общественного мнения. Если ведомство реагирует на ошибочный вывод, его сообщения могут изменить поведение людей. Затем цифровые следы усиливают ошибку при следующем анализе.

Таблица 8. Корреляционный анализ

Связь	ρ	p
Communication Complexity — Perceived AI Benefits	0,141	0,037
AI Use Frequency — Perceived AI Benefits	0,230	<0,001
AI Use Frequency — Adoption Support	0,224	<0,001
Perceived AI Benefits — Direct Trust	0,490	<0,001
Perceived AI Benefits — Adoption Support	0,606	<0,001
Direct Trust — Adoption Support	0,441	<0,001
Human-AI Governance — Adoption Support	0,411	<0,001

Люди считают ИИ полезным, когда им доверяют напрямую ($\rho = 0{,}490$) и когда поддерживают его внедрение ($\rho = 0{,}606$). Когда поддерживают механизмы Human-AI Governance, люди тоже охотнее внедряют ИИ ($\rho = 0{,}411$). Эти цифры показывают, как ответы участников связаны между собой; они не доказывают, что одно вызывает другое, и не говорят о репутации государства.

Обсуждение. Практическая польза и доверие к выводам ИИ. Полезным ИИ-ассистента считают 71,4 % участников, а его рекомендациям напрямую доверяют 49,8 %. Эта разница важна для управления: экономия времени при подготовке аналитики ещё не гарантирует, что получённая оценка общественного мнения верна. Показатель доверия отражает позицию респондентов относительно использования рекомендаций.

Human-in-the-loop оставляет окончательное решение за человеком. HVRR предлагает перед этим шагом проверить, насколько верно ИИ отразил положение дел. Предполагаемая последовательность такова: анализ ИИ → проверка источников → сопоставление с независимыми данными о восприятии → решение → коммуникация → новое независимое измерение. Желательно, чтобы проверку проводили не только сотрудники, чья работа оценивается по коммуникационным KPI.

Если повышение рассчитанного ИИ-балла становится самостоятельной целью, ведомство может улучшить цифровой показатель, не меняя реальных результатов своей деятельности. Система тогда примет новый показатель за признак улучшения репутации. Чтобы увидеть такое расхождение, нужно отдельно изучать фактические результаты, мнение людей, цифровые следы и интерпретацию ИИ.

Обсуждение одного события на казахском и русском языках может проходить на разных площадках и с разными источниками. Поэтому совпадения или расхождения оценок ИИ и людей стоит проверять отдельно по языкам и регионам. Сейчас данные не позволяют точно установить, насколько велико это различие.

Если алгоритм дал небольшую ошибку, её легко увидеть и поправить. Труднее, когда ошибочный вывод уже попал в коммуникацию, а система приняла последующие реакции за подтверждение. Тогда надо проверить и исходную оценку, и источник новых данных.

Таблица 9. Управленческая архитектура

Уровень	Функция	Контрольный вопрос	Механизм
Reality layer	Фактические результаты	Что произошло?	Независимые KPI качества услуг и политики
Data layer	Цифровые следы	Что доступно и чего нет?	Карта источников и coverage audit
AI layer	Анализ и синтез	Какова AI-интерпретация?	Версия модели, тестирование, provenance
Human governance	Проверка и ответственность	Можно ли опираться на вывод?	Human review + final decision authority
Reputation feedback	Сопоставление с восприятием	Изменилось ли человеческое восприятие?	Независимые опросы и качественные методы

- Не ставьте репутационный балл ИИ единственным KPI для коммуникаций.
- Измеряйте отдельно мнение людей о работе государства.
- Сохраняйте данные об источниках и способах получения выводов ИИ.
- Перед важным решением проверяйте рекомендации ИИ и оставляйте окончательное решение за человеком.
- Сравнивайте оценки ИИ и людей по темам, языкам и регионам.
- После больших информационных кампаний проверяйте, не усиливают ли новые данные прежнюю ошибку.
- Разделяйте результаты работы государства, мнение людей, цифровые следы и интерпретацию ИИ.
- Если алгоритмическая оценка постоянно расходится с независимой человеческой оценкой, останавливайте автоматическое распространение выводов до завершения проверки.

Ограничения и дальнейшие исследования. Исследование было сделано, чтобы понять, как люди относятся к ИИ-ассистенту, а не чтобы измерять репутацию государства. Выборка из 217 человек слишком мала, чтобы представлять всех, и данные собраны в один момент. В анкете нет независимой оценки того, как люди воспринимают ИИ, и нет балла репутации, который бы посчитал ИИ, поэтому нельзя вычислить Human–AI Reputation Gap. Найденные корреляции лишь показывают связь, а не причину. Ответ «Другое» не даёт информации о должностях участников.

Чтобы проверить предполагаемую обратную связь, можно провести эксперимент или повторные измерения. Сначала измеряем исходную оценку участников (H0). Затем показываем участникам подготовленное ИИ-резюме или ранжирование и измеряем оценку снова (H1). Разница $\Delta H = H1 – H0$ показывает, как изменилась оценка после знакомства с материалом, но для вывода о причине нужна контрольная группа. Позже можно изучить цифровые следы и посмотреть, как они влияют на следующий вывод ИИ. Эксперимент можно провести отдельно для казахского и русского языков, чтобы сравнить влияние языка, раскрытия источников и человеческой проверки.

Заключение. Оценка цифровой репутации с помощью ИИ может менять сам объект анализа. Это происходит, когда вывод алгоритма определяет дальнейшие сообщения государства, сообщения меняют реакцию людей, а реакция снова попадает в данные для нового вывода. Такая цепочка называется Digital Reputation Loop.

Опрос показал разницу между положительной оценкой полезности ИИ (71,4 %) и доверием к рекомендациям ИИ (49,8 %). Разница составляет 21,6 % пунктов. По этим цифрам нельзя сказать, насколько надёжны рекомендации и работает ли предложенная репутационная петля на практике. Данные лишь описывают отдельные условия, при которых организации могут использовать ИИ и контролировать его выводы.

87,1 % опрошенных говорят, что человек должен проверять рекомендации; 88,0 % считают, что окончательное решение должно оставаться за человеком. 71,9 % отмечают, что могут отклонить рекомендацию, а 72,4 % считают, что обязательный контроль повышает доверие. Мы видим связь: когда люди считают технологию полезной и считают важным соблюдать правила, они чаще поддерживают её внедрение.

Теоретическая модель предлагает рассматривать отдельно результаты работы государства, то, как её воспринимают люди, цифровые следы и оценку ИИ. Если оценка ИИ влияет на сообщения государства и потом меняет новые данные, появляется риск, что выводы будут подтверждать сами себя. Это надо проверить в реальных исследованиях.

Фраза «быть, а не казаться» здесь имеет смысл: репутацию нужно сравнивать с независимыми проверками работы государства. При использовании ИИ надо хранить информацию об источниках, назначать ответственных, давать возможность оспаривать решения и регулярно сравнивать алгоритмические оценки с мнением людей.

References

Agbabiaka, O., Ojo, A., & Connolly, N. (2025). Requirements for trustworthy AI-enabled automated decision-making in the public sector: A systematic review. Technological Forecasting and Social Change, 215, 124076. DOI: 10.1016/j.techfore.2025.124076. Language: English.

Azzahro, F., Hidayanto, A. N., & Shihab, M. R. (2025). Examining factors shaping citizens' perception of artificial intelligence in government: A systematic literature review. Social Sciences & Humanities Open, 11, 101518. DOI: 10.1016/j.ssaho.2025.101518. Language: English.

Criado, J. I., & Villodre, J. (2021). Delivering public services through social media in European local governments. Local Government Studies, 47(2), 253–275. DOI: 10.1080/03003930.2020.1729750. Language: English.

Fifaliana, R. M. (2026). A Systematic Literature Review of Empirical Research on the Impact of E-Government Use on Citizens' Trust in Government. DOI: 10.1177/15396754251413048. Language: English.

Gillespie, T. (2014). The Relevance of Algorithms. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), Media Technologies: Essays on Communication, Materiality, and Society, 167–194. MIT Press. DOI: 10.7551/mitpress/9042.003.0013. Language: English.

Hancock, J. T., Naaman, M., & Levy, K. (2020). AI-Mediated Communication: Definition, Research Agenda, and Ethical Considerations. Journal of Computer-Mediated Communication, 25(1), 89–100. DOI: 10.1093/jcmc/zmz022. Language: English.

Harrison, T. M., & Luna-Reyes, L. F. (2022). Cultivating trustworthy artificial intelligence in digital government. Social Science Computer Review, 40(2), 494–511. DOI: 10.1177/0894439320980122. Language: English.

Kleizen, B., Van Dooren, W., Verhoest, K., & Tan, E. (2023). Do citizens trust trustworthy artificial intelligence? Government Information Quarterly, 40(4), 101834. DOI: 10.1016/j.giq.2023.101834. Language: English.

Madan, R., & Ashok, M. (2023). AI adoption and diffusion in public administration: A systematic literature review and future research agenda. Government Information Quarterly, 40(1), 101774. DOI: 10.1016/j.giq.2022.101774. Language: English.

Moore, M. H. (1995). Creating Public Value: Strategic Management in Government. Harvard University Press. Language: English.
OECD. (2025). Governing with Artificial Intelligence: The State of Play and Way Forward in Core Government Functions. OECD Publishing. DOI: 10.1787/795de142-en. Language: English.
Strathern, M. (1997). Improving ratings: Audit in the British university system. European Review, 5(3), 305–321. DOI: 10.1002/(SICI)1234-981X(199707)5:3<305::AID-EURO184>3.0.CO;2-4. Language: English.
Straub, V. J., Morgan, D., Bright, J., & Margetts, H. (2023). Artificial intelligence in government: Concepts, standards, and a unified framework. Government Information Quarterly, 40(4), 101881. DOI: 10.1016/j.giq.2023.101881. Language: English.
Tabassi, E. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1. DOI: 10.6028/NIST.AI.100-1. Language: English.
Twizeyimana, J. D., & Andersson, A. (2019). The public value of E-Government – A literature review. Government Information Quarterly, 36(2), 167–178. DOI: 10.1016/j.giq.2019.01.001. Language: English.
UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence. UNESCO, Paris. Language: English.
Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector—Applications and challenges. International Journal of Public Administration, 42(7), 596–615. DOI: 10.1080/01900692.2018.1498103. Language: English.

Примечание об эмпирических данных. В разделах 5–6 повторно анализируются результаты ранее проведенного опроса 217 человек. Первоначальное исследование касалось воспринимаемой пользы ИИ-ассистента, доверия, человеческого контроля и отношения к внедрению. Здесь эти показатели используются только для обсуждения условий применения ИИ при анализе репутации. Исходные данные не позволяют непосредственно измерить Human–AI Reputation Gap.