

АРХИТЕКТУРНАЯ МОДЕЛЬ ДОМЕННО-АДАПТИРОВАННОГО LLM-RAG С ПОЛИТИКАМИ БЕЗОПАСНОСТИ ДЛЯ ОБЪЯСНИМОЙ ДИАГНОСТИКИ СЕТЕВЫХ И ИНТЕРНЕТ-ИНЦИДЕНТОВ В КРИТИЧЕСКОЙ ИНФРАСТРУКТУРЕ ЦЕНТРАЛЬНОЙ АЗИИ

Gaipnazarov Rustam¹[0000-0002-0049-8967]

Мухтарова Гулнора Хатамовна²[0009-0006-4514-2653]

Rakhimov Giyos³[0009-0005-3584-0207]

¹*Tashkent University of information technologies named after Muhammad al Kwarazmi, Faculty of Computer Engineering, Teacher, Tashkent 100084, Amir Temur Avenue 108, Uzbekistan*
r.gaipnazarov@tuit.uz

²*Tashkent University of information technologies named after Muhammad al Kwarazmi, Faculty of Computer Engineering, Teacher, Tashkent 100084, Amir Temur Avenue 108, Uzbekistan*
gulnora6530@inbox.ru

³*Tashkent University of information technologies named after Muhammad al Kwarazmi, Faculty of Computer Engineering, Teacher, Tashkent 100084, Amir Temur Avenue 108, Uzbekistan*
g.raximov@tuit.uz

Аннотация

Сетевые и интернет-инциденты в критической цифровой инфраструктуре все чаще описываются не языком протоколов, журналов событий и KPI, а через разрозненные жалобы операторов, пользователей и прикладных специалистов. На этом стыке появляется соблазн передать первичную диагностику универсальным большим языковым моделям. Однако в сетевом домене такой подход быстро упирается в три ограничения: слабую чувствительность к отраслевой терминологии, риск галлюцинаций и недостаточный контроль над политиками безопасности и приватности. В статье предложена авторская архитектурная модель **DAPT-Guard LLM-RAG**, ориентированная на безопасную и объяснимую обработку сетевых и интернет-инцидентов в условиях Центральной Азии. Модель объединяет доменную адаптацию запроса, защищенное извлечение знаний, контур анонимизации, policy-checking, независимую JudgeLLM-валидацию и слой объяснимости. Работа выполнена как обзорно-аналитическое исследование с синтезом авторской модели, без эмпирического датасета и без имитации экспериментальных результатов. В качестве результата сформирована целостная схема обработки инцидента, набор критериев ее оценки и аналитическая карта рисков для внедрения LLM в эксплуатационные сетевые процессы. Практическая ценность статьи состоит в том, что предложенная архитектура может служить основой для прототипов NOC/SOC-ассистентов, систем интеллектуального troubleshooting и безопасных LLM-сервисов для операторов связи, дата-центров, транспортных платформ и цифровых государственных сервисов региона.

Ключевые слова: LLM-RAG, диагностика сетевых инцидентов, кибербезопасность, domain adaptation, policy-aware AI, Central Asia.

1. Введение

Цифровая инфраструктура Центральной Азии строится как регионально связанная среда, где стоимость доступа, надежность сетей, развитие цифровых госуслуг и интеграция инфраструктурных платформ прямо зависят от качества сетевого управления [12]. На этом

фоне даже типовой сетевой сбой перестает быть локальной технической неприятностью. Он может затронуть облачные сервисы, транспортные узлы, корпоративные сети, медицинские учреждения и государственные платформы. Одновременно в 3GPP усиливается линия на стандартизированное внедрение AI/ML в сетевое управление: в Rel-19 уже расширяются use cases для network slicing, coverage and capacity optimization, а в рамках управления 5GS обсуждаются explainability, federated learning, мониторинг доверия к моделям и кросс-доменная оркестрация AI/ML [8], [9].

Параллельно растет интерес к использованию LLM как интерфейса для интерпретации инцидентов, операторских запросов и полужформализованных жалоб. У этой идеи есть сильная сторона: модель может быстро свести в одну точку естественный язык, технический контекст и накопленную базу знаний. Но в реальных сетях почти сразу обнаруживается слабое место. Универсальная LLM не различает эксплуатационно значимые нюансы с той точностью, которая требуется в телекоме и кибербезопасности. Там, где инженер ищет признак перегрузки интерфейса, деградации канала, ошибок маршрутизации или нарушения политики доступа, общая модель нередко производит внешне правдоподобный, но небезопасный ответ [1], [2], [10], [11].

В свежей работе по кросс-доменной трансляции сетевых запросов показано, что разрыв между языком пользователя и языком телеком-диагностики остается самостоятельной проблемой. Пользователь описывает симптомы через поведенческие признаки, тогда как сетевой специалист мыслит через KPI, топологию, привязку устройств и измеримые отклонения [1]. В другой работе policy-aware LLM-RAG для критической киберфизической среды показано, что даже retrieval grounding недостаточно, если в системе нет отдельного механизма проверки политики и независимой валидации ответа [2]. Именно на этом пересечении и возникает задача нашего исследования.

Цель работы – предложить архитектурную модель доменно-адаптированного LLM-RAG для безопасной, объяснимой и политически контролируемой диагностики сетевых и интернет-инцидентов в критической инфраструктуре Центральной Азии. Исследовательский вопрос формулируется так: какой набор функциональных блоков нужен, чтобы LLM в узком сетевом домене не только понимала инцидент, но и выдавала ответ, пригодный для эксплуатации с точки зрения безопасности, приватности, проверяемости и инженерной интерпретируемости [1], [2], [7], [9]?

Гипотеза работы состоит в том, что надежный LLM-ассистент для troubleshooting не может строиться вокруг одной генеративной модели. Ему нужен связанный контур из доменной адаптации, извлечения знаний, анонимизации, политики безопасности, независимой проверки и механизма объяснения решения [1]–[3], [7].

Научная новизна статьи заключается в том, что она не ограничивается общим тезисом о полезности LLM в сетях. Вместо этого предлагается единая архитектурная модель **DAPT-Guard LLM-RAG**, ориентированная именно на эксплуатационную диагностику сетевых и интернет-инцидентов, с учетом регионального контекста, ограничений приватности и требований к policy-aware reasoning. Практическая значимость работы связана с возможностью использовать предложенную схему как основу для NOC/SOC-ассистентов, интеллектуальных систем обработки инцидентов и безопасных сервисов поддержки операторов связи и цифровых платформ региона [8]–[12].

2. Методы

Работа выполнена как обзорно-аналитическое исследование с синтезом авторской модели. Методологически статья опирается на четыре слоя анализа.

Первый слой – тематический – обзор работ по domain-adaptive LLM, multi-agent reasoning, retrieval-augmented generation, LLM-as-a-Judge, privacy-preserving query processing и policy-aware orchestration [1]–[7]. Второй слой – сравнительный анализ архитектур, уже предложенных для критических или квазикритических сред, где недопустимы непроверенные ответы [1], [2], [7]. Третий слой – учет сетевого и регионального контекста: стандартизационной линии 3GPP и инфраструктурной логики цифровой связности в Центральной Азии [8], [9], [12]. Четвертый слой – авторский синтез архитектурной модели, в которой каждый блок отвечает за отдельный тип риска.

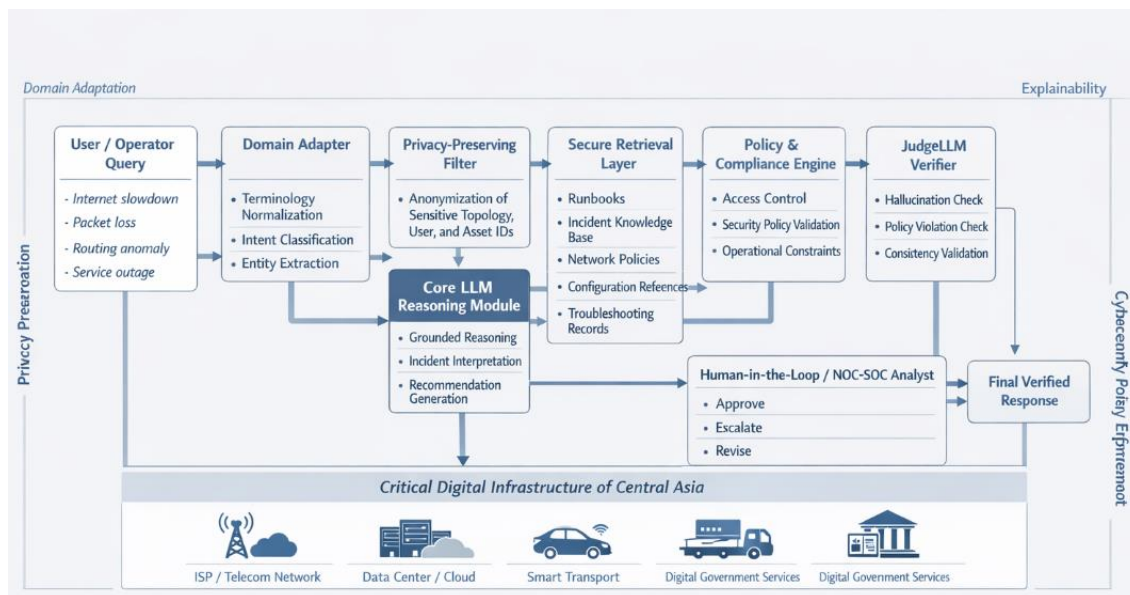


Рис. 1. DAPT-Guard LLM-RAG Architecture for Secure and Explainable Network Incident Diagnosis

В работе не использовались персональные данные, производственные журналы событий или закрытые операторские дампы. По этой причине статья не заявляет количественную валидацию на эмпирическом датасете. Результатом является не simulated benchmark, а проектная исследовательская модель с формализованными компонентами, сценариями применения и критериями будущей оценки.

2.1. Постановка задач

В рамках цели исследования были поставлены три задачи. Первая – выявить, какие ограничения мешают использовать универсальные LLM для сетевой диагностики. Вторая – определить архитектурные компоненты, которые снимают эти ограничения. Третья – собрать их в единую модель, пригодную для развертывания в сетях и интернет-сервисах с повышенными требованиями к надежности, приватности и управляемости [1], [2], [7], [10].

2.2. Авторская модель DAPT-Guard LLM-RAG

Предлагаемая модель включает шесть взаимосвязанных модулей.

Модуль 1. Domain Adapter. Он преобразует исходный запрос из произвольного пользовательского языка в представление, удобное для сетевой диагностики. Здесь выполняются нормализация терминов, распознавание типа инцидента, выделение сущностей и привязка запроса к доменной онтологии. Такой блок нужен, потому что

сетевые жалобы редко приходят в форме, пригодной для прямого анализа телеком-моделью [1].

Модуль 2. Privacy-Preserving Filter. Перед передачей запроса в ядро reasoning выполняется семантически щадящая анонимизация. Ее логика близка к идее semantic-preserving anonymization: топологически значимые сущности не уничтожаются безусловно, а преобразуются так, чтобы система не потеряла диагностический смысл, но снизила риск утечки чувствительных сведений [1], [6].

Модуль 3. Secure Retrieval Layer. После доменной адаптации система обращается к проверяемой базе знаний: gunbooks, регламентам, конфигурационным шаблонам, словарям инцидентов, журналам типовых сбоя, политикам реагирования и локальным ограничениям эксплуатации. Здесь используется RAG как механизм заземления ответа во внешнем знании [2], [3].

Модуль 4. Policy and Compliance Engine. Этот блок не дает модели работать в режиме «что сочинила, то и выдала». Он сопоставляет проект ответа с политиками безопасности, ролями доступа, допустимыми действиями, локальными нормативами и ограничениями на эксплуатационные команды [2], [7].

Модуль 5. JudgeLLM Verifier. Отдельная модель или отдельный режим той же модели повторно оценивает кандидатный ответ. Ее задача – не сгенерировать решение, а проверить, не содержит ли оно галлюцинации, нарушения политики, логической непоследовательности или опасного эксплуатационного совета [2].

Модуль 6. Explainability Layer. Система должна показать, почему она пришла к выводу. Объяснение не обязано превращаться в длинный chain-of-thought, но должно включать источник опорных знаний, тип инцидента, ключевые признаки и правило, которое стало основанием для рекомендации [7], [9].

2.3. Умеренная формализация

Полезность ответа модели предлагается описывать как свертку четырех параметров:

$$U = \alpha R + \beta S + \gamma E + \sigma C$$

где (R) – релевантность инциденту, (S) – безопасность ответа, (E) – объяснимость, (C) – соответствие политике и сетевому контексту.

Интегральный риск можно представить так:

$$Risk = w_1 H + w_2 P + w_3 D - w_4 V$$

где (H) – вероятность галлюцинации, (P) – вероятность policy violation, (D) – доменное расхождение между пользовательским запросом и сетевой интерпретацией, (V) – сила верификационного контура.

Такая запись не претендует на завершенную математическую теорию, но удобна как рамка для дальнейшей прототипной оценки.

2.4. Алгоритм обработки инцидента

Алгоритм 1. DAPT-Guard Pipeline

1. Получить пользовательский или операторский запрос (q).
2. Выполнить классификацию типа инцидента и доменную нормализацию.
3. Обнаружить чувствительные сущности и провести семантически допустимую анонимизацию.
4. Сформировать доменно-адаптированный запрос (q^*).
5. Выполнить retrieval из доверенной базы знаний (K).

6. Сконструировать grounded prompt с учетом (q^*), (K) и политик (P).
7. Сгенерировать кандидатный ответ (r).
8. Передать (r) в Policy Engine и JudgeLLM.
9. Если ответ нарушает политику или не проходит верификацию, вернуть его на этап уточнения или эскалации.
10. Сформировать объяснимый проверенный ответ (r_v).
11. Передать (r_v) оператору или в систему поддержки решений.

3. Результаты

Поскольку исследование носит архитектурно-аналитический характер, результаты статьи представлены не в форме числового benchmark, а в форме проектного исследовательского результата.

Первый результат – это сама архитектура DAPT-Guard LLM-RAG. В отличие от более простых схем, она строится не вокруг одного LLM-узла, а вокруг контура с разделением ролей. Генерация, проверка, политика и объяснимость разведены по разным функциональным слоям. Такой подход ближе к инженерной практике сетевого управления, где критично не только получить ответ, но и понимать, на чем он основан.

Второй результат – карта рисков, которую закрывает модель. Для сетевой и интернет-диагностики особенно существенны пять рисков: ложная интерпретация пользовательского запроса, утрата диагностически важных сведений при грубой анонимизации, галлюцинация при отсутствии надежного retrieval, нарушение политики безопасности и непрозрачность рекомендации для оператора. В предложенной модели каждый риск связан с отдельным защитным блоком, а не размазан по общей логике ответа.

Третий результат – аналитическая матрица применимости для инфраструктурных сценариев региона. В городской сети и дата-центре акцент смещается к конфигурационным и policy-рискам. В транспортной инфраструктуре возрастает чувствительность к задержке и объяснимости. В цифровых государственных сервисах на первый план выходят приватность, аудит действий и совместимость с регламентами доступа. Это делает предложенную архитектуру достаточно гибкой для разных вертикалей, сохраняя единое ядро reasoning.

Четвертый результат – набор критериев будущей оценки, который можно использовать в следующей работе уже с прототипом и тестовым датасетом.

Таблица 1

Сравнительный анализ подходов к диагностике сетевых и интернет-инцидентов: универсальная LLM, RAG и предложенная модель DAPT-Guard LLM-RAG

Критерий	Смысл оценки
Domain relevance	Насколько ответ действительно опирается на сетевой контекст, а не на общие шаблоны
Policy compliance	Нарушает ли рекомендация локальные регламенты и правила безопасности
Privacy preservation	Сохраняется ли диагностическая ценность после анонимизации
Explainability	Может ли инженер быстро понять ход решения
Actionability	Дает ли система действия, пригодные для эксплуатации
Escalation quality	Корректно ли система передает сложные случаи человеку

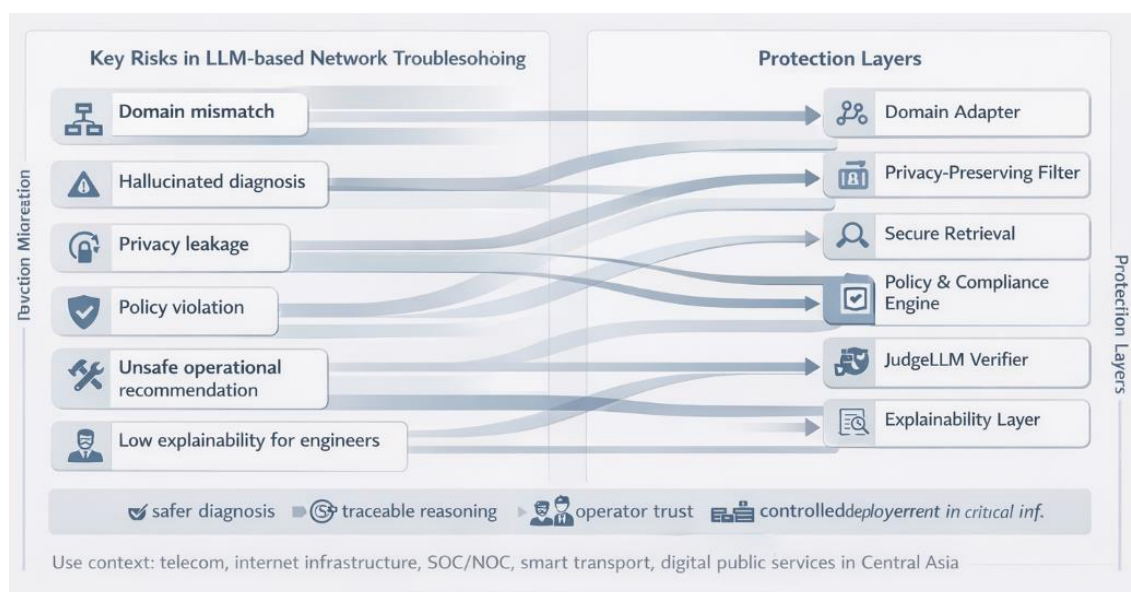
Для короткой conference paper этого набора достаточно, чтобы показать завершенность авторской модели и одновременно не приписывать ей несуществующие экспериментальные достижения.

4. Обсуждение

Сопоставление с литературой показывает, что существующие подходы обычно сильны в одном-двух направлениях, но не удерживают весь контур сразу. Работа по cross-domain query translation убедительно показывает, что один из самых ранних сбоя происходит уже на входе: система просто не понимает запрос как сетевой инцидент и не умеет бережно перевести его в язык телеком-диагностики. Policy-aware архитектура для edge-LLM-RAG, в свою очередь, хорошо показывает необходимость второго слоя – независимой JudgeLLM-проверки и привязки ответа к политике и telemetry-aware контексту. Однако эти идеи пока редко соединяются в единую модель для сетевой и интернет-диагностики, тем более с учетом эксплуатационного и регионального контекста.

На этом фоне главное достоинство DAPT-Guard LLM-RAG состоит не в том, что в нее добавлено больше блоков. Важнее другое: каждый блок встроен в причинно понятную цепочку. Domain Adapter снижает риск терминологической ошибки. Privacy Filter не дает потерять диагностический смысл при защите чувствительных данных. Retrieval привязывает ответ к реальному знанию, а не к памяти модели. Policy Engine проверяет допустимость действия. JudgeLLM разрывает опасную привычку верить первому сгенерированному ответу. Explainability Layer возвращает человеку контроль над интерпретацией решения.

Есть и более широкий контекст. 3GPP уже рассматривает AI/ML не как декоративный инструмент, а как элемент системной сетевой автоматизации, причем с управлением жизненным циклом, inference orchestration, explainability и privacy-aware training practices. NIST, со своей стороны, отдельно подчеркивает, что доверие к GenAI строится не вокруг красивого интерфейса, а вокруг управляемого риск-менеджмента и процедур оценки. Для телеком-отрасли это особенно чувствительно, потому что GSMA прямо указывает на разрыв между общими возможностями frontier-моделей и реальной готовностью к сетевому troubleshooting. Именно поэтому статья сознательно не предлагает «еще одного чат-бота для инженера», а строит проверяемую архитектурную рамку.



При этом у работы есть ограничения. Во-первых, статья не содержит эмпирической валидации на реальном операторском датасете. Во-вторых, региональный контекст Центральной Азии введен на уровне инфраструктурной логики и сценариев применения, а не через корпус локальных журналов событий и инцидентных тикетов. В-третьих, модель JudgeLLM сама нуждается в калибровке и не должна рассматриваться как абсолютный арбитр. В-четвертых, качество всей схемы будет зависеть от зрелости базы знаний: если runbooks устарели, retrieval просто быстрее принесет устаревший ответ.

Несмотря на эти ограничения, предложенная модель полезна уже сейчас. Она превращает обсуждение LLM в сетевой диагностике из общего разговора о потенциале в конкретную архитектурную задачу: как спроектировать систему, в которой каждый ответ проходит через доменное, политическое и верификационное сито. Для короткой conference paper такой уровень конкретности уже является самостоятельным вкладом.

5. Заключение

В статье предложена архитектурная модель DAPT-Guard LLM-RAG для безопасной и объяснимой диагностики сетевых и интернет-инцидентов в критической цифровой инфраструктуре Центральной Азии. В отличие от универсального LLM-подхода, модель строится как связанный контур из доменной адаптации, защищенного retrieval, policy-aware reasoning, JudgeLLM-валидации и слоя объяснимости. Такой подход закрывает основные риски, которые особенно заметны в сетевом troubleshooting: галлюцинации, доменное расхождение, утечку чувствительных сведений и небезопасные рекомендации.

Научная новизна работы состоит в синтезе единой архитектуры для узкого сетевого домена, где требования к точности ответа, его проверяемости и соответствию политике выше, чем в обычных диалоговых задачах. Практическая значимость состоит в том, что предложенная модель может служить основой для следующих этапов исследования: построения прототипа, создания тестового корпуса инцидентов, разработки локальной базы знаний и проведения сценарной оценки для операторов связи, дата-центров, smart transport-платформ и цифровых государственных сервисов региона.

Заявления

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Этическое заявление. В исследовании не использовались персональные данные, данные участников эксперимента и закрытые эксплуатационные логи. Работа носит концептуально-архитектурный характер.

Заявление об оригинальности. Текст представляет собой оригинальный исследовательский черновик, подготовленный для академического использования. Все заимствованные идеи и положения оформлены через ссылки на источники.

Литература

- 1) N. P. Tran, B. Jaumard, K. Premkumar, and S. Memon, "Cross-Domain Query Translation for Network Troubleshooting: A Multi-Agent LLM Framework with Privacy Preservation and Self-Reflection," arXiv:2604.13353, 2026.
- 2) O. Solanki, L. Praharaj, D. Gupta, and M. Gupta, "A Policy-Aware Edge LLM-RAG Framework for Internet of Battlefield Things Mission Orchestration," arXiv:2604.09493, 2026.

- 3) P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks,” *Advances in Neural Information Processing Systems*, vol. 33, 2020.
- 4) T. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, et al., “Language Models are Few-Shot Learners,” *Advances in Neural Information Processing Systems*, vol. 33, 2020.
- 5) S. Yao, J. Zhao, D. Yu, N. Du, I. Shafran, K. Narasimhan, and Y. Cao, “ReAct: Synergizing Reasoning and Acting in Language Models,” in *International Conference on Learning Representations (ICLR)*, 2023.
- 6) L. Sweeney, “k-Anonymity: A Model for Protecting Privacy,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 5, pp. 557–570, 2002.
- 7) C. Autio, R. Schwartz, J. Dunietz, S. Jain, M. Stanley, E. Tabassi, P. Hall, and K. Roberts, *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST AI 600-1)*. National Institute of Standards and Technology, 2024. ([NIST](#))
- 8) 3GPP, “AI/ML for NG-RAN & 5G-Advanced towards 6G,” 2026. Official 3GPP technology overview. Accessed: 17 Apr. 2026. ([3GPP](#))
- 9) 3GPP, “Engineering intelligence: Shaping AI/ML management for the 5G System,” 2025. Official 3GPP technology overview. Accessed: 17 Apr. 2026. ([3GPP](#))
- 10) GSMA, “Addressing the AI Accuracy Gap: Open-Telco LLM Benchmarks,” 2025. GSMA materials on telco-oriented LLM benchmarking and troubleshooting readiness. Accessed: 17 Apr. 2026. ([gsma.com](#))
- 11) GSMA, “The AI Telco Troubleshooting Challenge Launches to Transform Network Reliability,” 2025. GSMA industry initiative on AI for network troubleshooting. Accessed: 17 Apr. 2026. ([gsma.com](#))
- 12) World Bank, *Digital CASA - Kyrgyz Republic (P160230): Implementation Status & Results Report*. Washington, DC: World Bank, 2025. ([World Bank](#))
- 13) M. A. Jamshed, R. Singh, M. M. Saad, A. Kaushik, W. Shin, M. Dajer, and A. Mourad, “ISAC-Enabled Non-Terrestrial Networks for 6G: Design Principles, Standardization, Performance Tradeoffs, and Use Cases,” arXiv:2604.11593, 2026.