

Маркировка ИИ-контента в китайских СМИ: социальная дифференциация доверия аудитории и вызовы медиарегулированию

Научный руководитель – Вартанова Елена Леонидовна

У Цзиньфань

Аспирант

Московский государственный университет имени М.В.Ломоносова, Факультет журналистики, Кафедра теории и экономики СМИ, Москва, Россия
E-mail: xiaofan66272@gmail.com

Основной вопрос.

ИИ-технологии не столько ставят под сомнение художественную правдивость, сколько «историческую» — фактологическую достоверность информации. Быстрый рост генеративных алгоритмов вызывает критические вопросы доверия к новостям и необходимости эффективной маркировки ИИ-контента. Наше исследование целенаправленно разъясняет когнитивные пути восприятия ИИ-новостей и их последствия для теории убеждения. Конкретно, мы исследуем, как метки «ИИ-сгенерирован» и качество подачи влияют на восприятие объективности и точности сообщения. Особое внимание уделяется социальной неоднородности реакции аудитории, что имеет значение для разработки регуляторной политики. В частности, анализ ориентирован на последствия для практики регулирования СМИ и выработку сбалансированных норм маркировки в медиа-среде Китая.

Теоретическая рамка

Используемая теоретическая база – модель вероятности осознанной обработки информации (ELM), по которой различают два пути усвоения сообщения. Центральный путь предполагает вдумчивую оценку аргументов и опирается на логику и факты, тогда как периферийный путь использует поверхностные подсказки (цитаты экспертов, массовость и т.д.) и не требует глубокой переработки. Согласно ELM, вдумчивое внимание приводит к устойчивому изменению установок, а вторичный анализ — к кратковременному эффекту. Однако в условиях цифрового неравенства эти когнитивные механизмы могут по-разному проявляться в зависимости от социальных характеристик аудитории, что делает необходимым анализ не только индивидуальных, но и социокультурных факторов восприятия ИИ-контента. Применительно к новостям, мы рассматриваем восприятие объективности (периферийный путь) и точности фактов (центральный путь) как отдельные когнитивные механизмы обработки информации.

Однако в условиях цифрового неравенства эти когнитивные механизмы могут по-разному проявляться в зависимости от социальных характеристик аудитории, что делает необходимым анализ не только индивидуальных, но и социокультурных факторов восприятия ИИ-контента.

Методика исследования

Проведен интернет-эксперимент с N=295 участниками (платформа «Credamo [U+89C1] [U+6557]»). Участникам демонстрировались нейтральные новостные фрагменты в двух версиях: с явным уведомлением «создано ИИ» и без такого уведомления. Дополнительно варьировался голос диктора (человеческий или синтезированный). Собирались оценки доверия, объективности и точности контента. Учтено влияние ИИ-грамотности аудитории как социально-дифференцирующего модератора в медиасфере. Отметим, что в Китае ужесточается регулирование: к сентябрю 2025 года национальные правила обязывают пометать ИИ-синтетический контент видимыми метками.

Учитывая, что к сентябрю 2025 года в Китае вступают в силу обязательные требования по маркировке ИИ-контента, наш эксперимент моделирует именно этот пострегуляторный сценарий, оценивая не только эффект самой маркировки, но и её взаимодействие с уровнем ИИ-грамотности различных социальных групп.

Результаты

Эксперимент выявил три ключевых эффекта:

1. Эффект маркировки. Фрагменты с пометкой «создано ИИ» получают значительно более низкие оценки доверия ($M = 5,04$), чем традиционные новости ($M = 5,65$; $p < 0,01$), подтверждая данные Wang et al. (2025) о снижении воспринимаемой точности ИИ-контента.

2. Доминирование центрального пути. Оценка точности фактов объясняет 30% дисперсии доверия, тогда как оценка объективности — лишь 10%, что указывает на приоритет содержательного анализа над периферийными сигналами.

3. Парадокс ИИ-грамотности. Предупреждение об ИИ значительно снижает доверие только у групп с низкой ($\beta = 0,200$) и средней ($\beta = 0,090$) грамотностью; у высокограмотных эффект отсутствует ($\beta = -0,003$, $p = 0,959$).

Одновременно мы обнаружили «двойственный» эффект уровня ИИ-грамотности: у пользователей с высоким уровнем образования в области ИИ негативное влияние маркировки на доверие ослаблено, тогда как у менее грамотных снижение доверия усиливается. Этот результат подтверждает наблюдение Ali (2025), что ИИ-грамотность снижает штраф доверия за пометку «ИИ». С одной стороны, понимание технологий помогает смягчать скепсис; с другой — у слишком доверчивых пользователей без грамотности возникают сильные эмоции по поводу «недоверности».

Интересно, что не все исследования однозначны. Li и Yang (2024) сообщают об минимальном эффекте ИИГС-меток на восприятие достоверности и доверия. Это указывает на возможное смягчение реакции при знакомстве пользователей с новыми технологиями. Тем не менее, наши результаты демонстрируют явные различия по ключевым показателям, что позволяет сделать практические выводы.

Основные выводы и практическое значение

Полученные результаты свидетельствуют о том, что регуляторные меры по маркировке ИИ-контента могут иметь неоднородные социальные последствия.

Социальная дифференциация доверия: Мы выделили два пути влияния новостей — оценка объективности и оценка точности, — которые по-разному реагируют на информацию об источнике. Это расширяет классическую ELM-рамку, акцентируя внимание на специфике ИИ-контента.

Парадокс ИИ-грамотности как вызов социальной политике: Найден «двойной» эффект образования в области ИИ: грамотные пользователи менее пугаются ИИ-маркировок, но при этом могут становиться придирчивее к фактам. Такая модерация подчеркивает необходимость учитывать характеристики аудитории. Это указывает на необходимость не только технической маркировки, но и социально-ориентированных стратегий медиаобразования. Регуляторная маркировка ИИ-контента может усиливать социальное неравенство доверия при отсутствии системных мер по повышению медиаграмотности аудитории.

Маркировка и доверие: Четко подтверждено, что минимально помеченный как ИИ контент воспринимается менее достоверным. Это согласуется с практикой регулирования: китайские нормы уже требуют явного обозначения ИИ-контента.

Практически наши выводы важны для медиа и регуляторов. С одной стороны, открытая маркировка ИИ-контента повышает прозрачность и соответствует нормативным требованиям. С другой — она требует балансировки, чтобы не подрывать доверие аудитории без необходимости. Фокус на повышении медийной грамотности пользователей (включая понимание ИИ) может смягчить негативный эффект «algorithm aversion». В сумме, пред-

ставленные данные дают новое видение классической теории убеждения и указывают на пути минимизации рисков при интеграции генеративного ИИ в журналистику.

Источники и литература

- 1) Wang C., Sturgis P., de Kadt D. AI labeling reduces the perceived accuracy of online content but has limited broader effects [Электронный ресурс] // arXiv:2506.16202, 2025. URL: <https://arxiv.org/abs/2506.16202>.
- 2) Li F., Yang Y. Impact of Artificial Intelligence–Generated Content Labels on Perceived Accuracy, Message Credibility, and Sharing Intentions for Misinformation // JMIR Formative Research, 2024. – С. 1–13. DOI: 10.2196/XYZ.
- 3) Ali A. Transparency and Disclosure of ИИ-Use in Newsrooms: Impact on Trust and Credibility // Research Consortium Archive. – Vol. 3, No. 4 (2025). – P. 715–733.
- 4) Nickerson C. Elaboration Likelihood Model of Persuasion [Электронный ресурс] // SimplyPsychology. – 2023. – URL: <https://www.simplypsychology.org/elaboration-likelihood-model.html>.
- 5) Киберпространства Китайской Народной Республики. Меры по маркировке синтетического контента, сгенерированного ИИ [Электронный ресурс]. – Проект, сентябрь 2025. – URL: <https://www.xinhuanet.com/digital/20250902/d08b592223954340af8c02ac3d7161e1/c.html> (дата обращения).