DOI 10.37882/2223-2982.2026.02.07

ВЛИЯНИЕ ФОРМАТА ЭТИЧЕСКОЙ МАРКИРОВКИ НА ДОВЕРИЕ К НОВОСТНОМУ КОНТЕНТУ, СОЗДАННОМУ С ИСПОЛЬЗОВАНИЕМ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ

THE IMPACT OF ETHICAL LABELING FORMAT ON TRUST IN NEWS CONTENT CREATED USING LARGE LANGUAGE MODELS

A. Vasilev

Summary: The introduction of large language models (LLMs) into editorial practices raises ethical transparency issues for the media industry. This article explores how different labeling formats for AI-generated content affect audience perceptions. The focus is on trust, perceived accuracy, and sharing intentions among digital media audiences. The primary method is an online experiment (N=468) conducted using a 3 (labeling format) × 2 (text topic) design between subjects. The results show that detailed labeling, which explains the distribution of roles between AI and humans («AI was used for the draft, and the editor verified the facts»), significantly increases trust and willingness to share the material compared to brief labeling («Created with the help of AI») or no labeling at all. However, the effect of no labeling is similar to the effect of brief labeling, indicating that brief labeling is counterproductive.

Keywords: large language models, media ethics, media trust, content labeling, artificial intelligence in journalism, media psychology, and experimental methods in communications.

Васильев Андрей Михайлович
Аспирант, Российский Университет Дружбы Народов Им. Патриса Лумумбы»
114223013@pfur.ru

Аннотация: Проникновение больших языковых моделей (БЯМ) в редакционную практику ставит перед медиаиндустрией вопрос этической прозрачности. В статье исследуется, каким образом различные форматы маркировки контента, созданного с участием искусственного интеллекта, влияют на восприятие аудитории. Предметом исследования выступает доверие, воспринимаемая точность и готовность к шерингу у аудитории цифровых медиа. Основным методом выступает онлайн-эксперимент (N=468), построенный по схеме 3 (формат маркировки) × 2 (тематика текста) между субъектами. Результаты показывают, что развернутая маркировка, объясняющая распределение ролей между ИИ и человеком («ИИ использован для черновика, редактор проверил факты»), статистически значимо повышает доверие и готовность поделиться материалом по сравнению с краткой маркировкой («Создано с участием ИИ») или ее отсутствием. При этом эффект от отсутствия маркировки не отличается от эффекта краткой маркировки, что указывает на контрпродуктивность последней. Обнаружено, что влияние формата маркировки сильнее для тем, связанных с высокой субъективностью (культурная аналитика), чем для фактологических текстов (отчетность). На основе полученных данных формулируются практические рекомендации для редакций по внедрению этичной и эффективной коммуникации об использовании технологий искусственного интеллекта.

Ключевые слова: большие языковые модели, медиаэтика, доверие к медиа, маркировка контента, искусственный интеллект в журналистике, медиапсихология, экспериментальные методы в коммуникациях.

Введение

Цифровая трансформация медиаиндустрии, ускоренная развитием больших языковых моделей (БЯМ), привела к появлению нового класса инструментов для создания контента. Способность БЯМ генерировать связные, стилистически выверенные тексты ставит медиакомпании перед дилеммой: как использовать технологический потенциал для повышения эффективности, не подрывая фундаментальную основу отношений с аудиторией — доверие. Доверие к медиа является сложным конструктом, включающим воспринимаемую точность, беспристрастность и прозрачность процессов [1]. Внедрение БЯМ напрямую затрагивает последний компонент, создавая информационную асимметрию между производителем и потребителем контента. В этих условиях маркировка материалов, созданных с участием ИИ, обсуждается как ключевой элемент этического ответа медиаиндустрии на технологический вызов [2, 3].

Однако дискуссия часто сводится к бинарному выбору «маркировать или нет». Вопрос о том, как именно следует маркировать подобный контент, чтобы коммуникация была эффективной, остается недостаточно изученным, особенно в контексте российского медиаполя. Краткая, неопределенная маркировка (например, «сгенерировано ИИ») может вызвать у аудитории скепсис и отторжение, ассоциируясь с полной автоматизацией и девальвацией труда журналиста [4]. В то же время подробное объяснение процесса (распределение задач между ИИ и редактором) может, согласно теории прозрачности Т. Планта [5], усилить доверие, демонстрируя ответственный подход и контроль со стороны человека.

Теоретическая основа и гипотезы

Теоретический фундамент исследования строится на интеграции двух подходов. Во-первых, это концепция процессуальной прозрачности (process transparency),

которая постулирует, что раскрытие методов работы может повысить доверие к результату, особенно в ситуациях неопределенности [5]. Во-вторых, модель elaboration likelihood (ELM) Петти и Каччоппо [6] позволяет прогнозировать, как аудитория будет обрабатывать информацию о маркировке. Детализированная маркировка может запустить центральный путь обработки, активируя рациональную оценку, в то время как краткая — остаться на периферийном, вызывая поверхностные ассоциации (например, «машинное» равное «неавторитетное»).

На основе этого были выдвинуты следующие гипотезы:

H1: Развернутая маркировка, поясняющая роль редактора, приведет к более высоким показателям воспринимаемой достоверности и точности по сравнению с краткой маркировкой или ее отсутствием.

H2: Краткая маркировка («создано с участием ИИ») вызовет самый низкий уровень поведенческих намерений (готовности поделиться) среди всех условий эксперимента.

H3: Влияние формата маркировки будет сильнее выражено для текстов, требующих интерпретации и анализа (культурная аналитика), чем для строго фактологических текстов (отчетность), так как в первом случае аудитория более чувствительна к вопросу авторства и субъективной оценки.

Методология

Дизайн исследования. Для проверки гипотез был проведен количественный онлайн-эксперимент, построенный по межсубъектному факторному плану 3×2. Независимыми переменными выступили: 1) Формат маркировки (3 уровня: отсутствие маркировки (контроль), краткая маркировка, развернутая маркировка); 2) Тематика текста (2 уровня: фактологический текст на тему статистики рынка труда, аналитический текст о трендах в современном театре).

Материалы и процедура

Были созданы два оригинальных текста длиной 450-500 слов, близких по структуре и стилю к материалам качественных онлайн-изданий (например, «РБК» или «Ведомости»). Текст о рынке труда содержал данные Росстата, цитаты экспертов и график. Текст о театре строился на анализе нескольких премьер сезона, содержал интерпретацию и культурологический контекст. Для обеспечения экологической валидности и исключения влияния узнаваемости стиля конкретной БЯМ, тексты были написаны профессиональным журналистом, а затем три независимых эксперта-лингвиста подтвердили, что текст не содержит характерных «артефактов» известных БЯМ (например, избыточной грамматической правильности, шаблонных конструкций). Маркировка предъявлялась в виде стандартного блока под заголовком.

Выборка

В эксперименте приняли участие 468 респондентов (52% женщин, 48% мужчин) в возрасте от 20 до 55 лет (M = 34.7, SD = 9.2), отобранные через сервис онлайн-рекрутинга «Анкетолог» с применением квотной выборки по возрасту, полу и частоте потребления новостей. Участие было добровольным и анонимным.

Зависимые переменные и инструментарий. После прочтения текста респонденты заполняли опросник, включавший адаптированные надежные шкалы:

Воспринимаемая достоверность (Perceived Credibility): 6-пунктовая шкала на основе McCroskey (α = .88) (напр., «Насколько вы считаете этот источник надежным?»).

Воспринимаемая точность (Perceived Accuracy): 4-пунктовая шкала (α = .82) (напр., «Насколько вы уверены, что факты в этом материале верны?»).

Намерение поделиться (Intention to Share): 3-пунктовая шкала (α = .85) (напр., «Вероятно, я поделюсь этим материалом в социальных сетях»).

Все пункты оценивались по 7-балльной шкале Ликерта (от 1 – «совершенно не согласен» до 7 – «полностью согласен»). Также собирались демографические данные.

Результаты

Для проверки гипотез был проведен двухфакторный дисперсионный анализ (ANOVA). Статистическая обработка данных осуществлялась в программе IBM SPSS Statistics 27.

Влияние на воспринимаемую достоверность и точность. Основной эффект формата маркировки на воспринимаемую достоверность оказался статистически значимым ($F(2, 462) = 9.87$, $p < .001$, $\eta^2 = .04$). Post hoc сравнения по методу Тьюки подтвердили H1: группа с развернутой маркировкой показала значимо более высокий средний балл достоверности (M = 5.21, SD = 0.98) по сравнению с группой с краткой маркировкой (M = 4.52, SD = 1.12, $p < .01$) и контрольной группой (M = 4.81, SD = 1.05, $p < .05$). Различия между контрольной группой и группой с краткой маркировкой оказались незначимыми ($p = .12$). Аналогичная, но менее выраженная картина наблюдалась для воспринимаемой точности ($F(2, 462) = 4.95$, $p < .01$, $\eta^2 = .02$).

Влияние на намерение поделиться. Основной эффект для намерения поделиться также был значимым ($F(2, 462) = 7.34$, $p < .001$, $\eta^2 = .03$). Результаты частично подтвердили H2: группа с краткой маркировкой показала наименьшее намерение поделиться (M = 3.45, SD = 1.31),

что было значимо ниже, чем в группе с развернутой маркировкой ($M = 4.18$, $SD = 1.24$, $p < .01$). Однако разница с контрольной группой ($M = 3.78$, $SD = 1.29$) была на грани значимости ($p = .051$).

Эффект взаимодействия с тематикой. Обнаружено значимое взаимодействие факторов «формат маркировки» × «тематика» для зависимой переменной «достоверность» ($F(2, 462) = 3.98$, $p < .05$). Простой анализ эффектов показал, что для аналитического текста о театре позитивный эффект развернутой маркировки был выражен сильнее (разница с контролем $\Delta = 0.72$), чем для фактологического текста ($\Delta = 0.31$), что подтверждает H3. Для краткой маркировки негативный эффект также был сильнее для аналитической темы.

Обсуждение

Результаты исследования демонстрируют, что не всякая маркировка одинаково полезна для поддержания доверия аудитории. Ключевой вывод заключается в том, что развернутая, процессуально-ориентированная маркировка, подчеркивающая конечный контроль со стороны человека, является наиболее эффективной стратегией. Она не просто информирует, но и рефреймирует использование ИИ — не как замену журналиста, а как инструмент в его руках. Это снижает угрозу, связанную с технологией, и соответствует принципам процессуальной прозрачности.

Напротив, краткая маркировка («с участием ИИ») оказалась контрпродуктивной. Её эффект на доверие и готовность к шерингу не отличался от эффекта полного сокрытия информации или был негативным. Это можно интерпретировать в рамках ELM: такая маркировка служит негативным периферийным сигналом, активируя у аудитории смутные опасения по поводу «ненастоящести» контента, не предлагая при этом рациональных оснований для доверия.

Подтверждение третьей гипотезы указывает на необходимость дифференцированного подхода в зависимости от жанра и темы. Аудитория более критично и внимательно относится к маркировке в тех сферах, где ценится авторский взгляд, интерпретация и экспертиза. В фактологических жанрах доверие, видимо, в большей степени зависит от качества самих данных и источника, чем от раскрытия процесса создания.

Ограничения и дальнейшие исследования. К ограничениям можно отнести лабораторный характер эксперимента, где респонденты были сфокусированы на задаче. В реальных условиях потребления медиа эффект маркировки может быть слабее. Кроме того, не изучалось долгосрочное влияние такой маркировки на имидж издания. Перспективным направлением является исследование восприятия маркировки в контексте разных типов медиабрендов (традиционные СМИ vs. новые медиа) и изучение оптимальных визуальных форматов маркировки (иконки, цветовые коды).

Заключение

Внедрение больших языковых моделей в медиакоммуникации требует от редакций взвешенной этической и коммуникационной политики. Настоящее исследование эмпирически обосновывает, что путь к сохранению доверия лежит не через формальное соблюдение этических норм (простая маркировка), а через смысловую и развернутую коммуникацию с аудиторией. Разъяснение роли ИИ как ассистента, а не автора, и подчеркивание неизменной ответственности редакции за конечный продукт позволяют превратить потенциальный фактор недоверия в демонстрацию профессионализма и открытости. Полученные данные предоставляют научно обоснованные рекомендации для разработки редакционных стандартов, которые будут способствовать устойчивому развитию медиа в условиях технологической трансформации.

ЛИТЕРАТУРА

1. Meyer P. Defining and Measuring Credibility of Newspapers: Developing an Index // Journalism Quarterly. 1988. Vol. 65. P. 567–574.
2. Diakopoulos N. et al. Toward Understanding the Role of AI-Generated Content in News Media // Proc. ACM Hum.-Comput. Interact. 2023. Vol. 7. P. 1–24.
3. Калгин А.С. Искусственный интеллект в журналистике: этические дилеммы и профессиональные риски // Медиаальманах. 2022. № 5. С. 34–45.
4. Thierer A., O'Sullivan A., Russell R., Mahaffey J. How to Label AI-Generated Content: A Policy Analysis // Mercatus Center Policy Brief. 2023.
5. Plantin J.-C. The Politics of Transparency Platforms: Seeing Like a Dataset // Social Media + Society. 2019. Vol. 5(2).
6. Petty R.E., Cacioppo J.T. The Elaboration Likelihood Model of Persuasion // Advances in Experimental Social Psychology. 1986. Vol. 19. P. 123–205.
7. McCroskey J.C., Teven J.J. Goodwill: A reexamination of the construct and its measurement // Communication Monographs. 1999. Vol. 66. P. 90–103.

Журнал «Современная наука: актуальные проблемы теории и практики»