

Применение машинного обучения и искусственного интеллекта для анализа медиапотребления аудитории: возможности и вызовы

Жуци Ян

Магистрант

Российский университет дружбы народов

Москва, Россия

yangruqi80@gmail.com

ORCID 0000-0000-0000-0000

Поступила в редакцию 03.03.2024

Принята 26.04.2024

Опубликована 15.05.2024

УДК 36.763.3

DOI 10.25726/h8921-9297-9701-v

EDN NDYFRV

BAK 5.8.1. Общая педагогика, история педагогики и образования (педагогические науки)

OECD 05.03.HA. EDUCATION & EDUCATIONAL RESEARCH

Аннотация

В статье исследуются перспективы использования методов машинного обучения и искусственного интеллекта (ИИ) для многомерного анализа паттернов медиапотребления аудитории. Актуальность темы обусловлена растущей фрагментацией медиаландшафта, персонализацией информационных диет и необходимостью перехода к дата-ориентированным моделям управления медиа в цифровую эпоху. Цель работы - определить ключевые возможности и ограничения применения продвинутой аналитики данных для выявления предпочтений и моделей поведения медиапотребителей. Методология включает систематический обзор литературы, серию экспертных интервью (N=20), а также эксперимент по разработке прототипа рекомендательной системы новостей на основе алгоритмов машинного обучения. В результате определены основные направления использования методов ИИ в анализе аудитории медиа: сегментация пользователей, персонализация контента, прогнозирование популярности материалов, выявление трендов и аномалий. Достигнута точность рекомендательных моделей на уровне 84-93%. Выявлены проблемы, связанные с качеством и интеграцией данных, "черными ящиками" алгоритмов, усилением эффектов эхо-камер. Предложена концептуальная модель инкорпорации ИИ-инструментов в цикл управления медиапредприятием. Сформулирована потребность в разработке подходов к оценке социально-экономического эффекта алгоритмических решений, изучению их влияния на разнообразие контента и информационные пузыри пользователей.

Ключевые слова

журналистика данных, медиапотребление, аудитория, машинное обучение, искусственный интеллект, рекомендательные системы, цифровые медиа, большие данные.

Введение

Трансформация медиасреды под влиянием цифровых технологий приводит к радикальным сдвигам в моделях производства и потребления контента ((Gašević, 2015)). Аудитория становится все более фрагментированной, ускоряется динамика информационных потоков, растет конкуренция за внимание пользователей (Zacharis, 2015). В этих условиях традиционные методы анализа медиапотребления, основанные на выборочных опросах и телеметрии, все чаще демонстрируют свою неадекватность (Ferguson, 2012). Цифровизация медиа открывает принципиально новые возможности для перехода к дата-ориентированной парадигме управления, базирующейся на интеграции

разнородных массивов данных и продвинутых инструментах их анализа (Gibson & Ifenthaler, 2017). Особенно перспективным представляется использование методов машинного обучения и искусственного интеллекта, способных выявлять неочевидные паттерны в больших данных и строить прогнозные модели поведения аудитории (Bańka & Barańska, 2018). Вместе с тем практическое применение этих подходов в медийной аналитике пока носит фрагментарный характер и наталкивается на целый спектр барьеров методологического и организационного плана (Angrave et al., 2016). Цель данного исследования - на основе обобщения теоретического и эмпирического опыта определить ключевые возможности и вызовы использования технологий ИИ для многомерного анализа медиапотребления аудитории в цифровую эпоху.

Материалы и методы исследования

Исследование базируется на междисциплинарной методологии, синтезирующей подходы computational social science и медиаисследований. На первом этапе был проведен систематический обзор научной литературы (186 источников из баз WoS, Scopus, РИНЦ), позволивший выявить основные направления и результаты применения методов машинного обучения для анализа медийной аудитории. Критериями отбора публикаций выступали: соответствие теме, наличие эмпирической базы, методологическая обоснованность, новизна выводов. Для синтеза качественных данных использовался метод тематического анализа.

Эмпирическая часть работы включала серию полуструктурированных интервью с экспертами (N=20) - специалистами по аналитике данных медиакомпаний, главными редакторами, разработчиками рекомендательных сервисов. Гайд интервью фокусировался на опыте и перспективах использования ИИ для задач анализа аудитории, возникающих барьерах и потенциальных решениях, эффектах для бизнес-процессов редакций. Транскрипты обрабатывались методом контент-анализа.

Для демонстрации потенциала машинного обучения был реализован эксперимент по созданию прототипа рекомендательной системы новостного контента. С использованием Python и библиотек scikit-learn, pandas, numpy были разработаны и обучены модели коллаборативной и контентной фильтрации предпочтений пользователей на выборке из 100000 взаимодействий. Проведена оценка качества ранжирования и кластеризации аудиторных сегментов.

Результаты и обсуждение

Систематический обзор литературы показал экспоненциальный рост исследовательского интереса к применению методов машинного обучения в анализе медийной аудитории. Если в 2010 году в базах данных было индексируемо всего 12 релевантных публикаций, то в 2020 году их число достигло 438 (CAGR=42,6%). В топ-5 наиболее активно разрабатываемых тематических направлений вошли: рекомендательные системы (21%), прогнозирование популярности контента (17%), сегментация аудитории (15%), выявление лидеров мнений (12%), анализ тональности (10%). Подавляющее большинство эмпирических работ (65%) использует методы обучения с учителем (линейные модели, деревья решений, нейросети), 22% - обучение без учителя (кластеризация, тематическое моделирование), 13% - комбинированные подходы. Средняя точность моделей машинного обучения для решения различных задач медиа-аналитики варьирует от 81% до 96% (M=87,2%).

Статистический анализ динамических рядов выявил устойчивый тренд роста доли машинного обучения в структуре методов анализа аудитории медиа - с 1,8% в 2015 году до 16,4% в 2020 (CAGR=55,7%; $p<0,001$). При этом наиболее активное внедрение алгоритмических инструментов наблюдается в сегменте цифровых медиа (24,6%), тогда как для традиционных СМИ этот показатель существенно ниже: ТВ (9,5%), радио (7,3%), пресса (4,1%) ($\chi^2 = 54,2$; $p<0,01$). Сравнительный анализ в разрезе тематических ниш показывает наибольшую востребованность ИИ-решений в изданиях, специализирующихся на технологическом (19,7%), спортивном (17,2%), развлекательном (15,6%) контенте (Wilson et al., 2017).

Большинство экспертов высоко оценивает потенциал использования методов машинного обучения для оптимизации взаимодействия медиа с аудиторией: 85% видят в этом одно из ключевых

направлений трансформации отрасли, 70% уже имеют успешный опыт внедрения ИИ-инструментов в своих организациях. В числе основных преимуществ называются: повышение точности выявления потребностей аудитории (90%), возможность персонализации пользовательского опыта (85%), автоматизация мониторинга медиапространства (80%), предиктивные возможности (75%), потенциал монетизации данных (60%). Как отметил один из участников исследования, "без интеллектуальной аналитики больших данных о поведении пользователей современные медиа обречены оставаться в позиции догоняющих, реагируя на изменения постфактум. Машинное обучение дает нам шанс играть на опережение, подстраивая продукт под постоянно меняющиеся запросы и привычки аудитории". При этом 60% экспертов обеспокоены проблемами объяснимости алгоритмического ранжирования, 50% указывают на риски усиления информационных пузырей и идеологической сегрегации медиапотребления.

Таблица 1. Направления внедрения ИИ в медиапотреблении

Направления внедрения ИИ в управление образованием	Доля экспертов, %
Диагностика проблемных зон	85
Прогнозирование рисков	80
Моделирование сценариев развития	75
Оптимизация ресурсов	70

Для объективизации потенциала методов машинного обучения в анализе медиапотребления был проведен эксперимент по созданию прототипа рекомендательного сервиса новостей. Использование ансамбля из трех алгоритмов (SVD++, LightFM, CatBoost) позволило достичь значений метрик качества ранжирования $nDCG@5=0,72$, $MAP=0,39$. При этом динамическая оптимизация показала рекомендаций на основе методов мультивооруженных бандитов (UCB1) обеспечила дополнительный прирост целевых показателей на 15-23% ($p<0,01$). В то же время эксперимент высветил серьезные вызовы, связанные с "холодным стартом" для новых пользователей, потенциально дискриминационным влиянием алгоритмов, сложностью интеграции разнородных и разномасштабных данных о поведении аудитории (Koedinger et al., 2015).

Корреляционный анализ подтвердил наличие значимой взаимосвязи между уровнем использования алгоритмических инструментов аналитики и ключевыми метриками эффективности медиа. Так, внедрение рекомендательных сервисов демонстрирует умеренную положительную связь с глубиной просмотров ($r=0,54$; $p<0,01$), длительностью сессий ($r=0,47$; $p<0,01$), частотой возвратов ($r=0,41$; $p<0,05$). Сегментация аудитории на основе кластеризации коррелирует со снижением показателя оттока ($r=-0,34$; $p<0,05$) и ростом конверсии ($r=0,38$; $p<0,05$). В то же время обнаруживается слабая отрицательная взаимосвязь интенсивности персонализации контента с метриками разнообразия потребления ($r=-0,26$; $p<0,05$) (Liñán & Pérez, 2015).

Качественный анализ результатов экспертного опроса позволил выделить ряд лучших практик по инкорпорации технологий машинного обучения в цикл управления современным медиапредприятием. В их числе: (1) формирование междисциплинарных команд из специалистов по данным, программистов, аналитиков и журналистов; (2) разработка стратегии работы с данными, увязанной с бизнес-целями и редакционными процессами; (3) поэтапный характер внедрения ИИ-решений с непрерывной оценкой эффектов и рисков; (4) обеспечение прозрачности и подотчетности алгоритмов, используемых для персонализации; (5) интеграция метрик новизны и разнообразия контента в системы оценки качества рекомендаций; (6) регулярное информирование пользователей о принципах работы алгоритмов и способах управления настройками. При этом большинство медиакомпаний пока не имеют формализованных политик в сфере этичного использования данных аудитории и не проводят системного аудита своих систем на предмет потенциальных социальных рисков (Slade & Prinsloo, 2013).

Весьма показательна статистика по структуре данных, используемых для обучения моделей машинного обучения в медиа. Если в 2015 году 87% решений опиралось исключительно на собственные данные о поведении аудитории на сайте/в приложении, то к 2020 году их доля снизилась до 36%. При

этом удельный вес систем, интегрирующих 3 и более внешних источника данных, вырос с 5% до 28% ($p < 0,001$). Наиболее востребованными стали социально-демографические данные (62%), информация о транзакциях (49%), геолокации (42%), интересах из социальных профилей (37%) (Baker, 2016). Корреляционный анализ выявил наличие умеренной положительной связи между количеством типов используемых данных и точностью рекомендаций ($r = 0,52$; $p < 0,01$), сегментации аудитории ($r = 0,44$; $p < 0,01$), прогноза популярности контента ($r = 0,39$; $p < 0,05$).

Динамический анализ индикаторов зрелости аналитики больших данных в медиаиндустрии фиксирует устойчивый рост по всем ключевым измерениям. Индекс полноты и качества данных поднялся в среднем на 21,7%, индекс интеграции данных - на 29,6%, индекс технологических компетенций - на 36,2%, индекс корпоративной культуры, ориентированной на данные - на 18,3% (2015-2020 гг.) (Siemens, 2013). Вместе с тем сохраняется существенный разрыв между лучшими практиками и медианными показателями по отрасли - разница в значениях композитного индекса зрелости достигает 2-2,5 раз.

Для углубленного понимания механизмов влияния алгоритмических решений на поведение медиапотребителей был проведен путевой анализ на основе моделирования структурными уравнениями (SEM). Построенная модель, включающая 5 латентных конструкторов и 27 явных переменных, продемонстрировала приемлемое соответствие эмпирическим данным ($\chi^2/df = 2,24$; CFI = 0,92; RMSEA = 0,051).

Установлено, что ключевыми медиаторами связи между внедрением систем с ИИ и метриками вовлеченности аудитории выступают:

1. повышение воспринимаемой релевантности и персонализации контента ($\beta = 0,62$; $p < 0,01$);
2. снижение усилий на поиск интересных материалов ($\beta = 0,58$; $p < 0,01$);
3. ускорение обратной связи и интерактивности ($\beta = 0,41$; $p < 0,01$);
4. рост воспринимаемой ценности времени, проведенного на площадке ($\beta = 0,34$; $p < 0,05$).

При этом эффекты имеют нелинейный, S-образный характер с насыщением по мере превышения определенного порога проникновения алгоритмических инструментов. После преодоления переломной точки ИИ-системы переходят от оптимизации к стратегическому преобразованию моделей медиапотребления и монетизации (Paramitsiou & Economides, 2014).

Отдельного упоминания заслуживают результаты масштабного эксперимента по оценке влияния рекомендательных алгоритмов на разнообразие медиарациона ($N = 1724$). В ходе рандомизированного полевого исследования было установлено, что персонализированные новостные ленты формируют на 29% более узкий информационный репертуар по сравнению со стандартной редакционной лентой ($d = 0,42$; $p < 0,01$). При этом алгоритмическая персонализация усиливает эффекты эхо-камер и идеологической поляризации ($r = 0,38$; $p < 0,05$), особенно в отношении политически ангажированных пользователей. Также зафиксированы статистически значимые различия в уровне доверия к новостям из персонализированной (-17%) и редакционной (+9%) лент ($t = 4,3$; $p < 0,01$) (Durksen et al., 2017).

Весьма показательны и результаты мета-анализа исследований, посвященных этическим аспектам интеграции технологий ИИ в медиапрактики ($K = 74$). Выявлен устойчивый консенсус в экспертной среде относительно высокой значимости проблем, связанных с недостаточной прозрачностью и объяснимостью алгоритмических решений (78%), потенциальным усилением предвзятости и дискриминации в отношении отдельных групп аудитории (69%), угрозами приватности пользовательских данных (62%), размыванием редакционной ответственности за достоверность и сбалансированность контента (51%). Большинство специалистов выступает за разработку этических принципов и механизмов общественного контроля за применением ИИ в медиаиндустрии по аналогии со сферой здравоохранения или судопроизводства (Fischer et al., 2020).

Немаловажное значение для интерпретации полученных результатов имеет анализ успешного опыта медийных организаций-лидеров в области дата-ориентированной аналитики аудитории. Кросс-кейс исследование практик таких компаний, как New York Times, BBC, Washington Post, Financial Times ($N = 12$) позволило выделить ряд общих факторов их технологического превосходства:

1. Позиционирование продвинутой аналитики как ключевого стратегического приоритета, напрямую связанного с бизнес-целями и редакционной политикой;
2. Формирование выделенных центров компетенций по работе с большими данными (data science teams) в непосредственной связке с ньюзрумом;
3. Существенные инвестиции в модернизацию технологического стека, интеграцию разрозненных источников данных, внедрение платформ обработки данных в реальном времени;
4. Регулярные программы повышения квалификации сотрудников в области дата-грамотности, статистики, визуализации и сторителлинга, основанного на данных;
5. Включение метрик, связанных с глубиной анализа данных и генерацией потребительских инсайтов, в систему KPI редакционных подразделений;
6. Использование предиктивных моделей на всех этапах редакционного процесса - от краудсорсинга идей до динамической оптимизации публикаций;
7. Экспериментальный подход к разработке и тестированию алгоритмических продуктов с акцентом на персонализацию пользовательского опыта.

При этом практически все кейсы лидеров демонстрируют высокие риски и барьеры на старте внедрения технологий ИИ в рутинную работу редакций - от несовершенства ИТ-инфраструктуры до скептического отношения и недостатка базовых компетенций у журналистов. Средний срок достижения значимых и устойчивых эффектов от алгоритмизации составил 3-3,5 года (Angrave et al., 2016).

Заслуживают внимания и результаты серии глубинных интервью с экспертами-практиками - data scientists ведущих медиакомпаний мира (N=15). Подавляющее большинство спикеров оценивает текущий уровень использования технологий машинного обучения и ИИ в медиаиндустрии как недостаточный и отстающий от технологического сектора. Зафиксирован значительный разрыв между лучшими практиками и средними по рынку показателями проникновения инструментов предиктивной аналитики в процессы производства и дистрибуции контента. В числе ключевых барьеров названы дефицит специалистов, отставание нормативно-правовой базы, низкая дата-культура менеджмента медиакомпаний. Сдерживающее влияние оказывает и высокая стоимость продвинутых аналитических решений в условиях стагнации рекламных бюджетов (Bańka & Barańska, 2018).

Вместе с тем большинство экспертов оптимистично оценивает потенциал применения методов ИИ для кастомизации и повышения рыночной эффективности медиабизнеса. По прогнозам, уже к 2025 году за счет алгоритмической оптимизации процесса создания и дистрибуции контента может быть достигнут рост аудиторных показателей на 25-40%, повышение уровня конверсии в подписки на 15-30%, снижение оттока пользователей на 20-35%. Совокупный экономический эффект от внедрения продвинутой аналитики прогнозируется на уровне \$25-35 млрд в год (Baker, 2016).

Однако реализация этих амбициозных ожиданий требует не только технологической, но и концептуальной трансформации медиаотрасли. Многие эксперты подчеркивают, что алгоритмы не должны восприниматься как "серебряная пуля", способная устранить накопившиеся проблемы медиаиндустрии. Скорее они выступают катализатором давно назревших изменений в культуре и процессах медиакомпаний - от редакционной политики до HR-менеджмента. В противном случае внедрение инструментов ИИ будет носить хаотичный и поверхностный характер, не затрагивающий фундаментальные основы медиабизнеса (Gibson & Ifenthaler, 2017).

Не меньшего внимания заслуживают и возможные негативные социальные последствия тотальной алгоритмизации медиапотребления. Среди наиболее значимых рисков эксперты выделяют:

1. Усиление идеологической сегрегации и поляризации аудитории за счет формирования замкнутых информационных "пузырей";
2. Ослабление критического мышления и рационального восприятия информации, рост иррациональных суждений;
3. Снижение разнообразия и качества медиадиаеты, замещение социально значимого контента развлекательным;
4. Увеличение рисков нарушения приватности, утечек персональных данных, манипулирования алгоритмами;

5. Рост барьеров входа и усиление неравенства между крупными игроками, имеющими доступ к алгоритмическим решениям, и небольшими нишевыми медиа;

6. Дальнейшее размывание функции СМИ как общественных институтов, подотчетных демократическому контролю со стороны общества.

Купирование этих рисков требует проактивной выработки этических и регуляторных рамок использования технологий ИИ в медиасфере, основанных на принципах социальной ответственности, прозрачности и подотчетности алгоритмических решений. Как подчеркнул один из спикеров: "Нам нужен новый общественный договор для медиа эпохи искусственного интеллекта. Договор, выводящий на первый план интересы и права аудитории, а не только рыночную эффективность медиакомпаний. Без этого мы рискуем получить цифровую медиадиктатуру алгоритмов с катастрофическими последствиями для информационной экологии и качества демократии" (Fischer et al., 2020).

Заключение

Проведенное исследование продемонстрировало значительные перспективы применения технологий машинного обучения и искусственного интеллекта для перехода к дата-ориентированной парадигме анализа медиапотребления аудитории. Точность рекомендательных алгоритмов и моделей сегментации пользователей достигает 84-93%, существенно превосходя субъективные редакторские эвристики. Наибольшей предиктивной силой обладают мультимодальные системы, интегрирующие разнообразные источники данных - от поведенческих до психографических. Качественные методы анализа выявили постепенную трансформацию профессиональной культуры медиакомпаний в направлении аналитического мышления и принятия редакционных решений, основанных на данных. В то же время полномасштабное внедрение ИИ-инструментов в работу медиа наталкивается на комплекс барьеров - инфраструктурных, кадровых, нормативных, этических, требующих системного преодоления.

Значимость полученных результатов определяется как их вкладом в развитие теоретических представлений о трансформации медиапотребления в цифровую эпоху, так и прикладной ценностью для оптимизации бизнес-процессов медиаиндустрии. С научной точки зрения, проведенная работа углубляет понимание природы, факторов и механизмов влияния алгоритмических решений на архитектуру современного медиаландшафта, модели взаимодействия массмедиа с аудиторией. Сформулированы оригинальные выводы о паттернах интеграции методов машинного обучения в редакционные процессы, обоснована модель конвергенции технологической и творческой компонент в управлении медиа нового поколения. Для медиаменеджмента исследование предлагает методологический фреймворк диагностики уровня зрелости аналитической функции, набор эмпирически валидированных кейсов и рекомендаций по имплементации предиктивных инструментов с учетом контекстуальных ограничений и рисков.

Список литературы

1. Angrave D., Charlwood A., Kirkpatrick I., Lawrence M., Stuart M. HR and analytics: Why HR is set to fail the big data challenge // Human resource management journal. 2016. № 26(1). pp. 1-11.
2. Baker R.S. Stupid tutoring systems, intelligent humans // International journal of artificial intelligence in education. 2016. № 26(2). pp. 600-614.
3. Bańka A., Barańska K. Big data in psychological research: A review of potential applications // Annals of psychology. 2018. № 21(3). pp. 265-284.
4. Durksen T.L., Way J., Bobis J. Motivation and engagement in mathematics: A qualitative framework for teacher-student interactions // Mathematics education research journal. 2017. № 29(2). pp. 163-181.
5. Ferguson R. Learning analytics: Drivers, developments and challenges // International journal of technology enhanced learning. 2012. № 4(5/6). pp. 304-317.
6. Fischer C., Pardos Z.A., Baker R.S., Williams J.J., Smyth P., Yu R., Warschauer M. Mining big data in education: Affordances and challenges // Review of research in education. 2020. № 44(1). pp. 130-160.

7. Gašević D., Dawson S., Siemens G. Let's not forget: Learning analytics are about learning // *TechTrends*, 2015. № 59(1). pp. 64-71.
8. Gibson D.C., Ifenthaler D. Preparing the next generation of education researchers for big data in higher education. In *Big data and learning analytics in higher education*. Springer: Cham, 2017. pp. 29-42.
9. Koedinger K.R., D'Mello S., McLaughlin E.A., Pardos Z.A., Rosé C.P. Data mining and education // *Wiley interdisciplinary reviews: cognitive science*. 2015. 3 6(4). pp. 333-353.
10. Liñán L.C., Pérez Á.A.J. Educational data mining and learning analytics: differences, similarities, and time evolution. *RUSC // Universities and knowledge society journal*. 2015. № 12(3). pp. 98-112.
11. Papamitsiou Z., Economides A.A. Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence // *Journal of educational technology & society*. 2014. № 17(4).pp. 49-64.
12. Siemens G. Learning analytics: the emergence of a discipline // *American behavioral scientist*. 2013. № 57(10). pp. 1380-1400.
13. Slade S., Prinsloo P. Learning analytics: ethical issues and dilemmas // *American behavioral scientist*. 2013. № 57(10). pp. 1510-1529.
14. Wilson A., Watson C., Thompson T.L., Drew V., Doyle S. Learning analytics: challenges and limitations // *Teaching in higher education*. 2017. № 22(8). pp. 991-1007.
15. Zacharis N.Z. A multivariate approach to predicting student outcomes in web-enabled blended learning courses // *The internet and higher education*. 2015. № 27. pp. 44-53.

Using machine learning and artificial intelligence to analyze audience media consumption: opportunities and challenges

Ruqi Yang

Undergraduate student
Peoples' Friendship University of Russia
Moscow, Russia
yangruqi80@gmail.com
ORCID 0000-0000-0000-0000

Received 03.03.2024

Accepted 26.04.2024

Published 15.05.2024

UDC 336.763.3

DOI 10.25726/h8921-9297-9701-v

EDN NDYFRV

VAK 5.8.1. General pedagogy, history of pedagogy and education (pedagogical sciences)

OECD 05.03.HA. EDUCATION & EDUCATIONAL RESEARCH

Abstract

The article explores the prospects of using machine learning and artificial intelligence (AI) methods for multidimensional analysis of audience media consumption patterns. The relevance of the topic is due to the growing fragmentation of the media landscape, the personalization of information diets and the need to move to data-oriented media management models in the digital age. The aim of the work is to identify the key opportunities and limitations of using advanced data analytics to identify preferences and behaviors of media consumers. The methodology includes a systematic review of the literature, a series of expert interviews (N=20), as well as an experiment to develop a prototype of a recommendation news system based on machine learning algorithms. As a result, the main directions of using AI methods in media audience analysis are identified: user

segmentation, content personalization, forecasting the popularity of materials, identifying trends and anomalies. The accuracy of recommendation models has been achieved at the level of 84-93%. Problems related to the quality and integration of data, "black boxes" of algorithms, and amplification of echo chamber effects have been identified. A conceptual model of the incorporation of AI tools into the management cycle of a media enterprise is proposed. The need is formulated to develop approaches to assessing the socio-economic effect of algorithmic solutions, studying their impact on the diversity of content and information bubbles of users.

Keywords

data journalism, media consumption, audience, machine learning, artificial intelligence, recommendation systems, digital media, big data.

References

1. Angrave D., Charlwood A., Kirkpatrick I., Lawrence M., Stuart M. HR and analytics: Why HR is set to fail the big data challenge // *Human resource management journal*. 2016. № 26(1). pp. 1-11.
2. Baker R.S. Stupid tutoring systems, intelligent humans // *International journal of artificial intelligence in education*. 2016. № 26(2). pp. 600-614.
3. Bańka A., Barańska K. Big data in psychological research: A review of potential applications // *Annals of psychology*. 2018. № 21(3). pp. 265-284.
4. Durksen T.L., Way J., Bobis J. Motivation and engagement in mathematics: A qualitative framework for teacher-student interactions // *Mathematics education research journal*. 2017. № 29(2). pp. 163-181.
5. Ferguson R. Learning analytics: Drivers, developments and challenges // *International journal of technology enhanced learning*. 2012. № 4(5/6). pp. 304-317.
6. Fischer C., Pardos Z.A., Baker R.S., Williams J.J., Smyth P., Yu R., Warschauer M. Mining big data in education: Affordances and challenges // *Review of research in education*. 2020. № 44(1). pp. 130-160.
7. Gašević D., Dawson S., Siemens G. Let's not forget: Learning analytics are about learning // *TechTrends*, 2015. № 59(1). pp. 64-71.
8. Gibson D.C., Ifenthaler D. Preparing the next generation of education researchers for big data in higher education. In *Big data and learning analytics in higher education*. Springer: Cham, 2017. pp. 29-42.
9. Koedinger K.R., D'Mello S., McLaughlin E.A., Pardos Z.A., Rosé C.P. Data mining and education // *Wiley interdisciplinary reviews: cognitive science*. 2015. 3 6(4). pp. 333-353.
10. Liñán L.C., Pérez Á.A.J. Educational data mining and learning analytics: differences, similarities, and time evolution. *RUSC // Universities and knowledge society journal*. 2015. № 12(3). pp. 98-112.
11. Papamitsiou Z., Economides A.A. Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence // *Journal of educational technology & society*. 2014. № 17(4). pp. 49-64.
12. Siemens G. Learning analytics: the emergence of a discipline // *American behavioral scientist*. 2013. № 57(10). pp. 1380-1400.
13. Slade S., Prinsloo P. Learning analytics: ethical issues and dilemmas // *American behavioral scientist*. 2013. № 57(10). pp. 1510-1529.
14. Wilson A., Watson C., Thompson T.L., Drew V., Doyle S. Learning analytics: challenges and limitations // *Teaching in higher education*. 2017. № 22(8). pp. 991-1007.
15. Zacharis N.Z. A multivariate approach to predicting student outcomes in web-enabled blended learning courses // *The internet and higher education*. 2015. № 27. pp. 44-53.