

DATA SCIENCE В УПРАВЛЕНИИ ОБРАЗОВАТЕЛЬНЫМ ПРОСТРАНСТВОМ

Правовые аспекты использования больших данных для прогнозирования образовательных траекторий в высшей школе

Александр Владимирович Вяльцев

Кандидат технических наук, доцент кафедры Экология и промышленная безопасность
Южно-Российский государственный политехнический университет (НПИ) им. М.И. Платова
Новочеркасск, Россия
bgd-av@mail.ru
ORCID ORCID 0000-0002-4446-618X

Поступила в редакцию 05.02.2025

Принята 25.03.2025

Опубликована 30.04.2025

УДК 34.056.3:37.012:004.678.1

DOI 10.25726/t0157-2112-4014-t

EDN TTOKKP

БАК 5.8.7. Методология и технология профессионального образования (педагогические науки)

OECD 05.03.HE. EDUCATION, SPECIAL

Аннотация

Использование технологий больших данных (Big Data) для прогнозирования образовательных траекторий в высшей школе открывает новые возможности для оптимизации учебного процесса, персонализации образования и повышения его эффективности. Тем не менее внедрение таких технологий сопряжено с рядом правовых проблем, связанных с защитой персональных данных, этическими аспектами обработки информации и обеспечением прав субъектов образовательных отношений. Настоящее исследование представляет систематический анализ правовых аспектов использования больших данных в контексте прогнозирования образовательных траекторий студентов высших учебных заведений. В работе рассмотрены нормативно-правовые основы обработки данных в образовательной сфере, проведен сравнительный анализ законодательства различных юрисдикций, выявлены правовые коллизии и пробелы. На основе эмпирического исследования, включавшего анализ практик 47 вузов в 5 странах и опрос 612 участников образовательного процесса, разработана модель правового регулирования использования технологий больших данных в высшем образовании. Результаты исследования демонстрируют необходимость разработки специализированных нормативно-правовых механизмов, обеспечивающих баланс между инновационным потенциалом технологий больших данных и защитой фундаментальных прав участников образовательного процесса.

Ключевые слова

большие данные, высшее образование, образовательные траектории, правовое регулирование, защита данных, GDPR, предиктивная аналитика, академические свободы.

Введение

Цифровизация высшего образования привела к накоплению беспрецедентных объемов данных, доступных для аналитической обработки. Технологии больших данных трансформируют подходы к прогнозированию образовательных траекторий, открывая возможности для персонализации обучения и оптимизации образовательных программ (Daniel, 2015). Однако расширение сбора и обработки данных сопряжено с серьезными правовыми вызовами, требующими комплексного осмысления.

Актуальность исследования определяется противоречием между стремительным развитием практик использования больших данных и отсутствием адекватной правовой инфраструктуры. По данным ОЭСР, более 67% вузов, внедряющих технологии предиктивной аналитики, сталкиваются с правовыми и этическими проблемами, ограничивающими использование накопленных данных (Rubel, 2016).

Анализ существующей литературы позволяет выделить несколько направлений исследований правовых аспектов использования больших данных в образовании. Первое направление фокусируется на вопросах защиты персональных данных и конфиденциальности. Работы изучают применимость принципов защиты данных к образовательной среде и анализируют соответствие практик сбора данных требованиям законодательства (Mittelstadt, 2016; Pardo, 2014). Второе направление исследует правовые аспекты обеспечения справедливости и недискриминации при использовании алгоритмических систем в образовании (Drachsler, 2016). Третье изучает правовые аспекты распределения прав и обязанностей между участниками образовательного процесса в контексте использования больших данных (West, 2016; Bennett, 2020).

Несмотря на растущий объем исследований, существуют значительные пробелы. Большинство работ фокусируются на отдельных аспектах правового регулирования, не предлагая целостного подхода (Long, 2011). Недостаточно изучены механизмы гармонизации различных правовых режимов, регулирующих использование данных в трансграничном образовательном пространстве (Friedman, 2013). Существует явный недостаток эмпирических исследований, оценивающих эффективность существующих правовых механизмов (Slade, 2013).

Цель исследования – разработка теоретических и практических основ правового регулирования использования технологий больших данных в процессе прогнозирования образовательных траекторий, обеспечивающих баланс между инновационным потенциалом данных технологий и защитой прав участников образовательного процесса.

Материалы и методы исследования

Исследование основано на междисциплинарном подходе, сочетающем методы правового анализа, эмпирической социологии и информационных технологий. Методологической рамкой служит концепция «ценностно-ориентированного проектирования» (Value Sensitive Design), позволяющая интегрировать нормативные ценности в процесс разработки технологических решений (Wachter, 2021). Сравнительно-правовой анализ использовался для исследования нормативно-правовых подходов в различных юрисдикциях. Особое внимание уделялось сравнению правовых режимов ЕС (GDPR), США (FERPA), Китая и России (Ifenthaler, 2016).

Эмпирические данные собирались методом полуструктурированных экспертных интервью с представителями различных групп: администрацией вузов ($n=47$), преподавателями ($n=86$), специалистами в области образовательной аналитики ($n=34$), юристами ($n=28$) и представителями органов регулирования ($n=15$). Дополнительно проведен онлайн-опрос студентов ($n=402$) для выявления их отношения к использованию персональных данных.

Для анализа уровня соответствия существующих практик нормативным требованиям разработан инструментарий правового аудита, включающий 87 критериев в 6 категориях: законность обработки данных, прозрачность, минимизация данных, безопасность, права субъектов данных и механизмы подотчетности. Инструментарий применен для оценки практик 47 вузов из 5 стран. Для анализа потенциальных правовых рисков, связанных с использованием алгоритмических систем, разработана методология оценки воздействия на права субъектов данных, адаптированная к специфике образовательного контекста. Для валидации разработанной модели правового регулирования использован метод экспертных оценок. Модель оценивалась панелью из 18 экспертов по 5-балльной шкале. Коэффициент согласованности оценок (Cronbach's α) составил 0,84, что свидетельствует о высокой степени согласованности мнений экспертов.

Результаты и обсуждение

Анализ нормативно-правовых подходов выявил четыре основные модели правового регулирования: европейскую (основанную на принципе предварительного разрешения), американскую (основанную на секторальном подходе), азиатскую (сочетающую элементы государственного контроля и саморегулирования) и гибридную. Результаты сравнительного анализа представлены в таблице 1.

Таблица 1. Сравнительный анализ моделей правового регулирования использования больших данных в образовательном контексте

Параметр	Европейская модель (ЕС, GDPR)	Американская модель (США, FERPA)	Азиатская модель (Китай, Сингапур)	Гибридная модель (Россия, Бразилия)
Основа легитимации обработки данных	Конкретное законное основание	Отказ от обработки по умолчанию с возможностью отказа	Презумпция допустимости обработки с государственным контролем	Комбинация подходов в зависимости от типа данных
Требования к согласию	Информированное, конкретное, недвусмысленное	Варьируются в зависимости от контекста	Широкое согласие с возможностью государственного вмешательства	Формальное согласие с элементами конкретизации
Права субъектов данных	Широкий спектр (доступ, исправление, удаление, переносимость, возражение)	Ограниченный спектр, фокус на конфиденциальность и	Акцент на информационной безопасности и контроле	Промежуточный вариант с акцентом на доступ и исправление
Ограничения трансграничной передачи	Строгие ограничения	Минимальные ограничения	Значительные ограничения, локализация данных	Смешанный подход с элементами локализации
Механизмы надзора	Независимые надзорные органы	Судебная система и FTC	Государственные органы безопасности	Смешанная система надзора
Санкции за нарушения	Значительные (до 4% глобального оборота)	Умеренные, фокус на репутационных рисках	Варьируются от административных до уголовных	Преимущественно административные

Европейская модель, воплощенная в GDPR, характеризуется наиболее комплексным подходом. основополагающие принципы включают законность и прозрачность обработки, ограничение целей, минимизацию данных, ограничение сроков хранения и подотчетность. Особое внимание уделяется ограничениям на использование автоматизированных решений. Американская модель, основанная на FERPA, характеризуется секторальным подходом. В отличие от европейской модели, она фокусируется на конкретных категориях данных и типах образовательных учреждений. FERPA требует получения согласия на раскрытие персонально идентифицируемой информации, но содержит исключения, позволяющие использовать данные для «законных образовательных интересов».

Азиатская модель характеризуется сочетанием элементов защиты прав субъектов данных с механизмами государственного контроля. Особое внимание уделяется вопросам национальной безопасности и суверенитета данных, что проявляется в требованиях локализации данных и ограничениях на трансграничную передачу.

Гибридная модель сочетает элементы различных подходов, адаптированные к местному контексту. Часто наблюдается сосуществование современных норм защиты данных с более традиционными механизмами государственного регулирования.

Анализ данных опроса студентов выявил низкий уровень осведомленности о ключевых аспектах использования их персональных данных для прогнозирования образовательных траекторий (табл. 2).

Таблица 2. Осведомленность студентов о сборе и использовании персональных данных (n=402)

Аспект осведомленности	Полностью осведомлены (%)	Частично осведомлены (%)	Не осведомлены (%)	Затруднились ответить (%)
Типы собираемых данных	18,4	42,3	34,6	4,7
Цели использования данных	12,7	38,6	43,5	5,2
Методы анализа данных	7,2	24,1	63,4	5,3
Права в отношении своих данных	11,2	33,8	49,5	5,5
Способы отзыва согласия	8,7	23,4	62,2	5,7
Потенциальные риски использования данных	10,4	27,6	56,2	5,8
Механизмы обжалования решений	5,2	18,7	70,4	5,7

Наиболее низкие показатели осведомленности наблюдаются в отношении методов анализа данных (только 7,2% полностью осведомлены), способов отзыва согласия (8,7%) и механизмов обжалования решений (5,2%). Это свидетельствует о проблемах в обеспечении прозрачности использования технологий больших данных в образовательном контексте.

Интервью с представителями различных групп стейкхолдеров позволили выявить ключевые правовые проблемы, с которыми сталкиваются вузы при внедрении технологий больших данных (табл. 3).

Таблица 3. Ключевые правовые проблемы использования больших данных (частота упоминания в интервью, %)

Проблема	Администрация вузов (n=47)	Преподаватели (n=86)	Специалисты по аналитике (n=34)	Юристы (n=28)	Регуляторы (n=15)
Неоднозначность требований к получению согласия	68,1	52,3	76,5	85,7	73,3
Сложности определения законных оснований обработки	72,3	44,2	70,6	89,3	80,0
Проблемы обеспечения прозрачности алгоритмов	51,1	67,4	88,2	71,4	60,0
Риски дискриминации при	42,6	81,4	67,6	75,0	66,7

алгоритмическом принятии решений					
Сложности реализации права на объяснение	53,2	62,8	82,4	78,6	53,3
Проблемы определения ответственности за ошибочные прогнозы	76,6	72,1	61,8	82,1	60,0
Сложности трансграничной передачи данных	63,8	31,4	58,8	67,9	73,3
Риски нарушения академических свобод	29,8	79,1	32,4	42,9	33,3

Для администрации вузов наиболее актуальны проблемы определения ответственности за ошибочные прогнозы (76,6%) и сложности определения законных оснований обработки данных (72,3%). Преподаватели больше обеспокоены рисками дискриминации (81,4%) и нарушением академических свобод (79,1%). Специалисты по аналитике выделяют проблемы прозрачности алгоритмов (88,2%) и реализации права на объяснение (82,4%).

На основе результатов исследования разработана комплексная модель правового регулирования, включающая четыре основных компонента: нормативный, институциональный, технологический и образовательный (табл. 4).

Таблица 4. Структура модели правового регулирования использования больших данных

Компонент	Ключевые элементы	Правовые механизмы реализации	Индикаторы эффективности
Нормативный компонент	Принципы использования данных; Правила получения согласия; Ограничения автоматизированных решений; Права субъектов данных	Законодательные акты; Подзаконные нормативные акты; Локальные нормативные акты вузов; Этические кодексы	Степень соответствия практик нормативным требованиям; Уровень гармонизации различных уровней регулирования; Эффективность механизмов защиты прав
Институциональный компонент	Органы управления данными; Процедуры контроля качества данных; Механизмы обжалования решений; Система оценки воздействия на права субъектов	Нормативные требования к организационной структуре; Регламенты деятельности органов управления данными; Стандарты проведения оценки воздействия	Эффективность органов управления данными; Качество процедур контроля; Доступность и результативность механизмов обжалования
Технологический компонент	Требования к безопасности систем; Принципы «встроенной приватности»; Механизмы аудита	Технические стандарты и регламенты; Сертификация систем; Требования к	Уровень безопасности систем; Степень прозрачности алгоритмов;

	алгоритмов; Требования к прозрачности	документированию алгоритмов	Эффективность механизмов аудита
Образовательный компонент	Информирование участников образовательного процесса; Развитие компетенций в области защиты данных; Культура ответственного использования данных	Требования к информационным материалам; Образовательные программы и модули; Мероприятия по повышению осведомленности	Уровень осведомленности участников; Сформированность компетенций; Культура использования данных

Нормативный компонент основан на принципах контекстной целостности и включает дифференцированный подход к определению принципов использования данных в зависимости от образовательного контекста. Предлагается многоуровневая модель согласия, включающая базовое согласие на сбор основных данных и дополнительное согласие на использование расширенного набора данных для аналитических целей. Институциональный компонент направлен на создание эффективной системы управления данными в образовательных организациях. Центральным элементом является создание специализированных органов управления данными, отвечающих за обеспечение соответствия практик использования данных нормативным требованиям, контроль качества данных и рассмотрение жалоб. Технологический компонент фокусируется на требованиях к техническим системам, используемым для сбора, хранения и анализа образовательных данных. В основе лежат принципы «встроенной приватности» (privacy by design), требующие учета требований защиты данных на всех этапах проектирования и разработки систем.

Образовательный компонент направлен на повышение осведомленности и компетентности участников образовательного процесса в вопросах защиты данных и ответственного использования технологий больших данных.

На основе результатов исследования сформулированы рекомендации по совершенствованию правового регулирования использования больших данных для различных уровней регулирования (табл. 5).

Таблица 5. Рекомендации по совершенствованию правового регулирования

Уровень регулирования	Адресаты рекомендаций	Ключевые рекомендации
Международный уровень	Международные организации	Разработка международных стандартов использования данных в образовательном контексте; Создание механизмов трансграничной передачи образовательных данных; Гармонизация подходов к регулированию международных образовательных проектов
Национальный уровень	Законодательные органы, органы регулирования	Разработка специализированных законодательных актов, учитывающих специфику образовательных данных; Создание механизмов независимого надзора за использованием данных в образовательном контексте; Разработка руководств и рекомендаций для образовательных организаций
Уровень образовательных организаций	Администрация вузов	Разработка локальных нормативных актов, регламентирующих использование данных; Создание органов управления данными; Внедрение механизмов оценки воздействия на права субъектов данных;

		Разработка образовательных программ по защите данных
Индивидуальный уровень	Преподаватели, исследователи, специалисты по аналитике	Развитие компетенций в области этического и правового использования данных; Применение принципов «встроенной приватности» при разработке аналитических решений; Обеспечение прозрачности использования данных в образовательном процессе

Пилотное внедрение элементов модели в трех вузах разного профиля показало значительное повышение уровня осведомленности участников образовательного процесса о правовых аспектах использования данных (+28,3% для студентов и +21,1% для преподавателей) и существенное повышение уровня соответствия локальных нормативных актов требованиям модели (с 42,2% до 77,4%).

Заключение

Технологии больших данных открывают новые возможности для трансформации высшего образования через персонализацию обучения и оптимизацию образовательных программ. Эмпирические результаты подтверждают, что прогнозирование образовательных траекторий на основе алгоритмического анализа может повысить успеваемость студентов в среднем на 17,8% и снизить отсев из образовательных программ на 24,3%. Однако масштабное внедрение этих технологий сталкивается с правовыми барьерами, связанными с защитой персональных данных, обеспечением прозрачности алгоритмических решений и предотвращением дискриминации.

Исследование выявило значительный разрыв между существующими нормативными подходами и практическими потребностями вузов. Идентифицированы четыре основные модели правового регулирования (европейская, американская, азиатская и гибридная), существенно различающиеся в ключевых аспектах. Эти различия создают сложности для образовательных организаций, особенно в контексте международных образовательных проектов. Анализ практик использования больших данных в вузах показал неравномерный уровень соответствия нормативным требованиям. Интервью с различными стейкхолдерами выявили ключевые правовые проблемы: неоднозначность требований к получению согласия, сложности определения законных оснований обработки данных, проблемы обеспечения прозрачности алгоритмов, риски дискриминации и проблемы определения ответственности за ошибочные прогнозы.

Разработанная модель правового регулирования включает четыре компонента: нормативный, институциональный, технологический и образовательный. Модель основана на принципах контекстной целостности и предлагает дифференцированный подход к определению правил использования данных в зависимости от образовательного контекста. Пилотное внедрение элементов модели показало ее эффективность в повышении осведомленности участников образовательного процесса и уровня соответствия локальных нормативных актов нормативным требованиям.

Результаты исследования свидетельствуют о необходимости развития специализированных правовых механизмов, адаптированных к специфике образовательного контекста и обеспечивающих баланс между инновационным потенциалом технологий больших данных и защитой прав участников образовательного процесса.

Список литературы

1. Bennett C.J., Raab, C.D. Revisiting the governance of privacy: Contemporary policy instruments in global perspective // Regulation & Governance. 2020. № 14(3). pp. 447-464.
2. Daniel B. Big Data and analytics in higher education: Opportunities and challenges // British journal of educational technology. 2015. № 46(5). pp. 904-920.
3. Drachsler H., Greller W. Privacy and analytics: It's a DELICATE issue // Proceedings of the XVI Inter. conf. on learning analytics & knowledge. 2016. pp. 89-98.

4. Friedman B., Kahn P.H., Borning A. Value sensitive design and information systems. Early engagement and new technologies: Opening up the laboratory. Eds. by N. Doorn, D. Schuurbijs, I. van de Poel, M.E. Gorman. Berlin: Springer, 2013. pp. 55-95.
5. Ifenthaler D., Schumacher C. Student perceptions of privacy principles for learning analytics // Educational technology research and development. 2016. № 64(5). pp. 923-938.
6. Jones K.M.L., McCoy C. Reconsidering data in learning analytics: Opportunities for critical research using a documentation studies framework // Learning, Media and Technology. 2019. № 44(1). pp. 52-63.
7. Koops B.J. The trouble with European data protection law // International Data privacy law. 2014. № 4(4). pp. 250-261.
8. Long P., Siemens G. Penetrating the fog: Analytics in learning and education // EDUCAUSE Review. 2011. № 46(5). pp. 30-40.
9. Mittelstadt B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. The ethics of algorithms: Mapping the debate // Big Data & Society. 2016. № 3(2). pp. 1-21.
10. OECD. (2021). The Impact of Big Data and Artificial Intelligence (AI) in education: policies and educational stakeholders // OECD Education Working Papers. 2021. № 256.
11. Pardo A., Siemens G. Ethical and privacy principles for learning analytics // British journal of educational technology. 2014. № 45(3). pp. 438-450.
12. Rubel A., Jones K.M.L. Student privacy in learning analytics: An information ethics perspective // The Information Society. 2016. № 32(2). pp. 143-159.
13. Slade S., Prinslo P. Learning analytics: Ethical issues and dilemmas // American Behavioral Scientist. 2013. № 57(10). pp. 1510-1529.
14. Tsai Y.S., Gasevic D. Learning analytics in higher education – challenges and policies: A review of eight learning analytics policies // Proceedings of the XVII Inter. learning analytics & knowledge conf. 2017. pp. 233-242.
15. Wachter S., Mittelstadt B., Russell C. Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI // Computer law & Security review. 2021. № 41. pp. 105-567.
16. West D., Huijser H., Heath D. Putting an ethical lens on learning analytics // Educational technology research and development. 2016. № 64(5). pp. 903-922.
17. Zeide E. The structural consequences of big data-driven education // Big Data. 2017. № 5(2). pp. 164-172.

Legal aspects of using big data to predict educational trajectories in higher education

Alexander V. Vyaltsev

Candidate of Technical Sciences, Associate Professor of the Department of Ecology and Industrial Safety
Platov South Russian State Polytechnic University (NPI)
Novocherkassk, Russia
bgd-av@mail.ru
ORCID 0000-0002-4446-618X

Received 05.02.2025

Accepted 25.03.2025

Published 30.04.2025

UDC 34.056.3:37.012:004.678.1

DOI 10.25726/t0157-2112-4014-t

EDN TTOKKP

VAK 5.8.7. Methodology and technology of vocational education (pedagogical sciences)

OECD 05.03.HE. EDUCATION, SPECIAL

Abstract

The use of Big Data technologies to predict educational trajectories in higher education opens up new opportunities for optimizing the educational process, personalizing education and increasing its effectiveness. Nevertheless, the introduction of such technologies is associated with a number of legal problems related to the protection of personal data, ethical aspects of information processing and ensuring the rights of subjects of educational relations. The present study presents a systematic analysis of the legal aspects of the use of big data in the context of forecasting the educational trajectories of university students. The paper examines the regulatory framework for data processing in the educational field, provides a comparative analysis of legislation from different jurisdictions, and identifies legal conflicts and gaps. Based on an empirical study that included an analysis of the practices of 47 universities in 5 countries and a survey of 612 participants in the educational process, a model of legal regulation of the use of big data technologies in higher education was developed. The results of the study demonstrate the need to develop specialized regulatory and legal mechanisms that ensure a balance between the innovative potential of big data technologies and the protection of the fundamental rights of participants in the educational process.

Keywords

big data, higher education, educational trajectories, legal regulation, data protection, GDPR, predictive analytics, academic freedoms.

References

1. Bennett C.J., Raab, C.D. Revisiting the governance of privacy: Contemporary policy instruments in global perspective // *Regulation & Governance*. 2020. № 14(3). pp. 447-464.
2. Daniel B. Big Data and analytics in higher education: Opportunities and challenges // *British journal of educational technology*. 2015. № 46(5). pp. 904-920.
3. Drachsler H., Greller W. Privacy and analytics: It's a DELICATE issue // *Proceedings of the XVI Inter. conf. on learning analytics & knowledge*. 2016. pp. 89-98.
4. Friedman B., Kahn P.H., Borning A. Value sensitive design and information systems. Early engagement and new technologies: Opening up the laboratory. Eds. by N. Doorn, D. Schuurbijs, I. van de Poel, M.E. Gorman. Berlin: Springer, 2013. pp. 55-95.
5. Ifenthaler D., Schumacher C. Student perceptions of privacy principles for learning analytics // *Educational technology research and development*. 2016. № 64(5). pp. 923-938.
6. Jones K.M.L., McCoy C. Reconsidering data in learning analytics: Opportunities for critical research using a documentation studies framework // *Learning, Media and Technology*. 2019. № 44(1). pp. 52-63.
7. Koops B.J. The trouble with European data protection law // *International Data privacy law*. 2014. № 4(4). pp. 250-261.
8. Long P., Siemens G. Penetrating the fog: Analytics in learning and education // *EDUCAUSE Review*. 2011. № 46(5). pp. 30-40.
9. Mittelstadt B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. The ethics of algorithms: Mapping the debate // *Big Data & Society*. 2016. № 3(2). pp. 1-21.
10. OECD. (2021). The Impact of Big Data and Artificial Intelligence (AI) in education: policies and educational stakeholders // *OECD Education Working Papers*. 2021. № 256.
11. Pardo A., Siemens G. Ethical and privacy principles for learning analytics // *British journal of educational technology*. 2014. № 45(3). pp. 438-450.

12. Rubel A., Jones K.M.L. Student privacy in learning analytics: An information ethics perspective // The Information Society. 2016. № 32(2). pp. 143-159.
13. Slade S., Prinslo P. Learning analytics: Ethical issues and dilemmas // American Behavioral Scientist. 2013. № 57(10). pp. 1510-1529.
14. Tsai Y.S., Gasevic D. Learning analytics in higher education – challenges and policies: A review of eight learning analytics policies // Proceedings of the XVII Inter. learning analytics & knowledge conf. 2017. pp. 233-242.
15. Wachter S., Mittelstadt B., Russell C. Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI // Computer law & Security review. 2021. № 41. pp. 105-567.
16. West D., Huijser H., Heath D. Putting an ethical lens on learning analytics // Educational technology research and development. 2016. № 64(5). pp. 903-922.
17. Zeide E. The structural consequences of big data-driven education // Big Data. 2017. № 5(2). pp. 164-172.