



ИИ-МЕТОДЫ В ИССЛЕДОВАНИИ ПРОЦЕССОВ МОДЕРНИЗАЦИИ В УЗБЕКСКОЙ ПРЕССЕ НАЧАЛА XX ВЕКА

Нигора Алимкуловна САТИБАЛДИЕВА

Преподаватель

Узбекский государственный университет мировых языков

Аннотация. Статья посвящена применению искусственного интеллекта (ИИ) для изучения процессов модернизации, отражённых в узбекской прессе первых десятилетий XX века. Рассматриваются методы автоматического анализа идеологических, культурных и социальных сдвигов, зафиксированных в газетах и журналах периода джадидизма и ранних советских реформ. Особое внимание уделяется алгоритмам тематического моделирования, анализу тональности, семантическому поиску и методам выявления дискурсивных паттернов. Показано, что ИИ-технологии позволяют реконструировать динамику модернизационных сюжетов – от школьных реформ до трансформации идентичности и урбанизации. Обозначены ограничения цифровых подходов, связанные с качеством оцифровки, графической неоднородностью источников и многослойностью дискурса эпохи.

Ключевые слова: ИИ, узбекская пресса, модернизация, джадидизм, тематическое моделирование, анализ тональности, исторические медиатексты, цифровые методы.

Газетная пресса Узбекистана начала XX века представляет собой уникальное зеркало модернизационных процессов, происходивших в Средней Азии на фоне культурных реформ, политической трансформации и появления новых форм общественной коммуникации. Джадидские газеты, издания общественных объединений, экономические бюллетени и раннесоветские печатные органы фиксировали конкурентные идеологии, социальные конфликты, образовательные проекты и дискурсы обновления. Однако исследование этих источников традиционными методами сталкивается с рядом трудностей: разрозненность архивов, устаревшие шрифтовые системы, неполнота фондов и объём массивов. В связи с этим искусственный интеллект становится ключевым инструментом для анализа модернизационных сюжетов, выявления дискурсивных закономерностей и реконструкции текстового пространства эпохи.

Оцифровка и подготовка материалов

Первые этапы работы с историческими газетами включают оцифровку и очистку текстов. Нейросетевые модели оптического распознавания символов позволяют обрабатывать дореволюционные гарнитуры, арабографические версии узбекского языка и сложные рукопечатные макеты [1]. На основе специализированных наборов данных современные OCR-системы автоматически исправляют искажения, распознают лигатуры и корректируют архаичные графемы, что повышает точность последующего анализа.

Оцифрованный массив становится основой для построения исследовательского корпуса, где каждое издание снабжается метаданными: год, место издания, политическая ориентированность, жанровая структура, сведения об авторах. Корпусная организация позволяет проводить статистические и сравнительные исследования модернизационного дискурса в разных регионах и периодах.

Динамика модернизационных тем: алгоритмы тематического моделирования

Модернизационные процессы начала XX века проявлялись в узбекской прессе через множество сюжетов: школьная реформа, распространение научных знаний, санитарная культура, экономическое развитие, положение женщин, строительство инфраструктуры, критика традиционных социальных структур.

Тематическое моделирование (LDA, NMF и др.) позволяет выделить устойчивые «пучки» обсуждаемых вопросов и проследить их интенсивность в разные годы [2]. Например:

- Тема школьной реформы усиливается в газетах Самарканда в 1908–1913 гг.
- Вопросы женского образования набирают выраженность ближе к 1917 году.
- Экономический дискурс преобладает в изданиях Туркестанского края после 1920 года.

Автоматическое моделирование выявляет также региональные различия в модернизационной повестке. Газеты Ферганы уделяли больше внимания аграрным вопросам, тогда как ташкентские издания активно обсуждали промышленность и городское благоустройство.

Анализ тональности и идеологических изменений

С помощью инструментов анализа тональности исследователи могут проследить эмоционально-оценочные характеристики материалов. Газеты начала XX века часто использовали публицистический стиль, насыщенный оценочными конструкциями, призывами и риторическими приёмами.

Алгоритмы sentiment analysis позволяют определить:

- отношение редакции к реформам в сфере образования,
- степень критики традиционных институтов,
- эмоциональный фон обсуждения политики, войны, налогов, здравоохранения.

Важным направлением является выявление изменения тональности в зависимости от политической ситуации. Например, обсуждение модернизации в 1913 году имеет иной эмоциональный профиль, чем в 1921 году, когда происходил переход к новой административной системе [3].

Дискурсивные паттерны и семантические сети

ИИ-алгоритмы семантического поиска и кластеризации позволяют восстановить сложные дискурсивные структуры, характерные для эпохи. Модели на основе word embeddings показывают, какие понятия были устойчиво связаны друг с другом:

- *маърифат* (просвещение) – с *тарбия* (воспитание), *илм* (наука), *мактаб* (школа),



- *миллат* (народ/нация) – с *тинчлик* (благополучие), *тараққиёт* (развитие), *ҳаракат* (движение),
- *ислохот* (реформа) – с *янгича фикрлаш* (новое мышление), *замонавийлик* (современность) [4].

Эти данные помогают исследователям реконструировать интеллектуальную карту модернизационного дискурса, выявлять центры и периферии смысловых полей.

Стилометрический анализ модернизационной журналистики

Стилометрия играет важную роль в изучении газет эпохи реформ. Многие тексты выходили под псевдонимами, и вопрос авторства нередко остаётся открытым. ИИ-модели анализируют синтаксические структуры, характерные коллокации и частотные параметры, что помогает выявить идеостиль публицистов, определить принадлежность спорных текстов, выделить редакционные правки [5].

Стилометрия также позволяет проследить эволюцию журналистского языка: сокращение архаичных оборотов, увеличение доли абстрактной лексики, внедрение научных терминов, появление публицистической метафоры.

Модернизация как медийная практика: картирование сетей

ИИ-методы позволяют анализировать не только тексты, но и структуру медийного пространства. Алгоритмы построения сетевых графов выявляют связи между газетами, обмен материалами, цитирование, полемику между редакциями.

Такое графическое моделирование показывает:

- какие издания были интеллектуальными центрами,
- как распространялись модернизационные идеи между городами,
- какие авторы оказывали влияние на другие редакции [6].

Эта информация ценна для реконструкции культурной динамики и коммуникационных стратегий.

Ограничения цифровых методов

Несмотря на возможности ИИ, исследователь сталкивается с рядом трудностей:

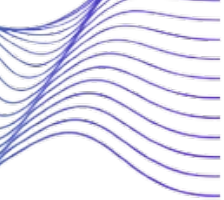
- неоднородность шрифтов и плохое качество архивов,
- ошибки OCR при работе с арабграфическими газетами,
- сложность учёта контекста эпохи при автоматической интерпретации,
- необходимость сочетать ИИ-анализ с классическими филологическими методами [7].

Тем не менее даже при этих ограничениях цифровые методы открывают новые ракурсы анализа модернизации.

Перспективы развития

Ключевыми направлениями дальнейших исследований становятся:

- создание мультязычных корпусов центральноазиатской прессы,
- разработка ИИ-моделей для анализа изображений и карикатур начала XX века,
- автоматическое выделение жанров,
- использование больших языковых моделей в реконструкции утраченных текстов,
- интеграция данных о прессе с социально-историческими источниками [8].



Эти шаги позволят сформировать комплексное цифровое представление медийных процессов эпохи.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Smith, R. An overview of the Tesseract OCR engine. *ICDAR*, 2007.
2. Springmann, M., Lüdeling, A. OCR of historical printings with neural networks. *DH Conference*, 2017.
3. Pang, B., Lee, L. Sentiment analysis in computational linguistics. *Foundations and Trends*, 2008.
4. Hamilton, W., Leskovec, J., Jurafsky, D. Semantic change in diachronic embeddings. *ACL*, 2016.
5. Holmes, D. Stylometry and authorship studies. *LLC*, 1998.
6. Moretti, F. *Graphs, Maps, Trees*. Verso, 2005.
7. Jockers, M. *Macroanalysis*. University of Illinois Press, 2013.
8. Underwood, T. *Distant Horizons*. Chicago University Press, 2019.