

# Использование искусственного интеллекта в патентных исследованиях

*Антон Черненко (ИКИ РАН)*

## Аннотация

Искусственный интеллект в патентных исследованиях используется для автоматизации процессов анализа патентных данных, выявления тенденций развития технологий, прогнозирования патентоспособности и определения классов МПК. В данной статье рассмотрены различные аспекты применения искусственного интеллекта (ИИ) в области патентного исследования: использование алгоритмов обработки естественного языка (NLP) для анализа текстовых описаний патентов и выделения ключевых слов, фраз и терминов, методы машинного обучения для анализа признаков патентов, связанных с патентоспособностью, и прогнозирования вероятности патентоспособности. Другие аспекты включают идентификацию семантически схожих патентов, создание структурированных отчетов на основе результатов и анализ больших объемов патентных данных с помощью ИИ.

**Ключевые слова:** искусственный интеллект, машинное обучение, патенты, патентные исследования, патентный поиск, патентоспособность, уровень техники, патентование, патентная чистота, artificial intelligence, machine learning, deep learning, prior art, search queries, search strategies

## 1 Значение патентных исследований в современном мире

Патентные исследования играют ключевую роль в современном мире, особенно в области инноваций и технологического развития. Патенты представляют собой юридические документы, которые защищают интеллектуальную собственность и права инноваторов на их изобретения, технические решения и технологии.

Важность патентных исследований обусловлена несколькими факторами:

- 1. Стимулирование инноваций:** Патенты предоставляют инноваторам экономические стимулы для создания новых технологий и изобретений. Они обеспечивают монопольные права на использование, продажу и распространение изобретений, что способствует привлечению инвестиций в исследования и разработки.

- 2. Защита интеллектуальной собственности:** Патенты предоставляют юридическую защиту авторским правам и интеллектуальной собственности.
- 3. Повышение конкурентоспособности компаний:** Благодаря патентам компании могут предлагать уникальные и инновационные продукты или услуги, что помогает улучшить их позицию на рынке и привлечь новых клиентов.
- 4. Исследование и анализ технической информации:** Патенты содержат детальные технические описания изобретений. Исследователи и инженеры могут использовать патентные базы данных для доступа к информации о существующих технологиях и тенденциях развития в различных отраслях.
- 5. Предотвращение нарушений:** Патентные исследования позволяют предотвратить нарушение чужих патентных прав.

## **2 Роль искусственного интеллекта в улучшении патентных исследований**

Одной из основных областей применения искусственного интеллекта в патентных исследованиях является обработка естественного языка (NLP) [1]. Алгоритмы обработки естественного языка позволяют автоматически извлекать ключевые слова и фразы из текстовых описаний патентов, а также классифицировать документы в соответствии с определенными темами или областями техники. Это значительно упрощает процесс поиска и анализа патентной информации, особенно при работе с большими объемами данных.

Искусственный интеллект также применяется для прогнозирования патентоспособности изобретений. С помощью методов машинного обучения и анализа признаков патентов, искусственный интеллект может предсказывать вероятность успешной регистрации патента [2]. Это помогает сократить время и ресурсы на подачу патентных заявок на изобретения, которые могут быть не патентоспособны.

Кроме того, искусственный интеллект способен автоматизировать процесс создания отчетов по патентным исследованиям. Путем автоматического поиска и анализа патентных данных, системы на основе искусственного интеллекта могут генерировать структурированные отчеты, содержащие информацию о технических тенденциях, ключевых игроках на рынке и прочих патентных метриках.

Еще одной важной функцией искусственного интеллекта в патентных исследованиях является выявление патентных коллизий. Искусственный интеллект может автоматически анализировать патентные данные и идентифицировать ситуации, когда предполагаемое изобретение может нарушить уже существующий патент. Это позволяет предотвратить возможные юридические проблемы и минимизировать риски для компании или инноватора.

### **3. Основные направления использования искусственного интеллекта в патентных исследованиях**

#### **3.1 Использование искусственного интеллекта для анализа текстовых описаний патентов**

Современные патентные исследования сталкиваются с огромными объемами текстовых данных, содержащихся в патентных документах. Искусственный интеллект предоставляет мощные инструменты для обработки и анализа этих текстовых описаний, что существенно улучшает процесс извлечения значимой информации из патентных документов.

Обработка естественного языка (NLP) позволяет искусственному интеллекту понимать и анализировать человеческий язык. В контексте патентных исследований, NLP используется для извлечения ключевых слов и фраз из текстовых описаний патентов [3]. Алгоритмы NLP автоматически определяют существенные термины, которые описывают технические особенности и инновационные аспекты изобретения. Это позволяет исследователям быстро и точно определить суть патентных документов и выявить ключевые понятия, связанные с темой исследования.

При анализе текстовых описаний патентов, часто возникает необходимость обработки словоформ для улучшения качества и точности анализа. Искусственный интеллект использует методы **лемматизации и стемминга** для приведения слов к их базовой форме или корню [4]. Лемматизация позволяет преобразовать слова к словарной форме, а стемминг - к общему корню. Это помогает уменьшить размерность данных и объединить различные формы слов, что улучшает обработку текстовых описаний патентов и повышает качество анализа.

Далее при помощи процесса **Part-of-Speech Tagging (POS-тегирование)** производится определение частей речи каждого слова в тексте (например,

существительное, глагол, прилагательное и т.д.) [5]. POS-тегирование помогает понять контекст и синтаксическую структуру текста, что может быть полезно при выделении ключевых слов.

Последний этап анализа - **извлечение именованных сущностей (Named Entity Recognition, NER)**, таких как названия компаний, местоположения, персоны, даты и т.д [6]. Это помогает определить конкретные технологии, продукты или источники, связанные с патентами.

Кроме текстовых описаний, патенты содержат структурированные данные, такие как классы Международной патентной классификации (МПК), даты подачи, публикации и другие метаданные. Искусственный интеллект анализирует и эти данные для выявления связей и паттернов, которые могут указывать на патентоспособность изобретения.

### **3.2 Идентификация семантически схожих патентов с помощью ИИ**

Искусственный интеллект применяет методы семантического анализа и машинного обучения для автоматического выявления семантических связей между патентными документами. В процессе идентификации семантически схожих патентов, системы искусственного интеллекта строят векторные представления для каждого текстового описания патента [7]. Эти векторы содержат числовую информацию о смысловой структуре текста и его содержании. Одним из наиболее распространенных методов является преобразование текстовых описаний патентов в числовые векторы с помощью методов **word2vec**, **GloVe** или **FastText**(см., например, [8]). Эти методы позволяют представить слова в виде векторов в многомерном пространстве, где близкие по смыслу слова имеют близкие векторы.

Затем, используя алгоритмы кластеризации, такие как **k-means**, **hierarchical clustering**, **DBSCAN** и другие искусственный интеллект группирует патентные документы на основе схожести их векторных представлений [9]. Патенты, которые имеют близкие векторные представления, считаются семантически схожими и объединяются в один кластер. Таким образом, системы искусственного интеллекта выявляют группы патентов, которые относятся к общей тематике или имеют схожие инновационные решения.

Дополнительно, искусственный интеллект может использовать методы сравнения текста, такие как **TF-IDF (Term Frequency-Inverse Document Frequency)**, для выявления наиболее схожих фрагментов между патентными документами [10]. Это позволяет выделить общие технические концепции и ключевые слова, которые характеризуют семантически схожие патенты.

### **3.3 Применение искусственного интеллекта для прогнозирования патентоспособности**

В современных патентных исследованиях искусственный интеллект активно применяется для предсказания вероятности патентоспособности конкретного изобретения. Этот процесс включает анализ различных признаков патентов, применение методов машинного обучения и предоставляет значительные преимущества в оценке патентоспособности.

#### **3.3.1 Анализ признаков патентов, связанных с патентоспособностью**

Первый шаг в прогнозировании патентоспособности - это анализ признаков, которые могут влиять на решение о выдаче патента. Такие признаки могут включать технические аспекты изобретения, степень новизны, уровень инноваций, патентоспособность по сравнению с уже существующими патентами и многое другое. Искусственный интеллект способен автоматически извлечь и классифицировать эти признаки из текстовых описаний патентов, что значительно облегчает и ускоряет процесс анализа.

Анализ признаков патентов, связанных с патентоспособностью, является ключевым шагом в использовании искусственного интеллекта для прогнозирования вероятности успешного получения патента для конкретного изобретения. Для этого необходимо выделить релевантные признаки из текстовых описаний патентов и структурированных данных, которые могут оказывать влияние на решение патентного офиса.

Искусственный интеллект применяет методы обработки естественного языка (Natural Language Processing, NLP) для анализа текстовых описаний патентов.

#### **3.3.2 Построение и валидация модели прогнозирования**

Собранные признаки и данные используются для построения модели прогнозирования. Это может быть модель машинного обучения, такая как логистическая регрессия, случайный лес, нейронные сети и другие. Модель обучается на исторических данных, которые содержат информацию о патентах, выданных или отклоненных офисом по интеллектуальной собственности. Модель ищет закономерности и зависимости между признаками и исходом получения патента для создания прогноза для нового изобретения [11].

После построения модели необходимо провести ее валидацию на новых данных, чтобы убедиться в ее точности и надежности. Модель проверяется

на тестовой выборке, которая не использовалась при обучении, для оценки ее производительности и способности обобщать знания на новые данные.

Анализ признаков патентов, связанных с патентоспособностью, с помощью искусственного интеллекта позволяет автоматизировать и улучшить процесс оценки вероятности успешного получения патента для новых изобретений. Это способствует более обоснованным решениям в области инноваций и интеллектуальной собственности и повышает эффективность патентных исследований.

### **3.3.3 Методы машинного обучения для прогнозирования вероятности патентоспособности**

Для прогнозирования вероятности патентоспособности конкретного изобретения искусственный интеллект использует различные методы машинного обучения [2]. Эти методы позволяют анализировать обширные объемы патентных данных и выявлять закономерности, которые сложно обнаружить вручную. Ниже представлены основные методы машинного обучения, применяемые для этой цели:

**Логистическая регрессия** является одним из наиболее распространенных методов машинного обучения для задач классификации. Она используется для предсказания вероятности принадлежности объекта к одному из двух классов – в данном случае, к патентоспособному или непатентоспособному классу. Логистическая регрессия строит линейную модель, которая оценивает влияние различных признаков на вероятность патентоспособности.

**Случайный лес (Random Forest)** является ансамблевым методом машинного обучения, который объединяет несколько деревьев решений для принятия окончательного решения. Каждое дерево строится на случайной подвыборке данных и случайных признаках, что позволяет избежать переобучения модели. Случайный лес обладает высокой точностью и устойчив к шумам в данных, что делает его эффективным методом для прогнозирования патентоспособности.

**Нейронные сети** – это глубокие модели машинного обучения, которые имитируют работу человеческого мозга. Они состоят из множества связанных между собой нейронов, которые обрабатывают данные и выдают результат. Нейронные сети обладают высокой гибкостью и способностью обучаться на больших объемах данных, что позволяет достичь высокой точности прогнозирования патентоспособности.

**Градиентный бустинг** – это еще один ансамблевый метод машинного обучения, который объединяет несколько слабых моделей в одну сильную. Он работает поэтапно, последовательно обучая модели и исправляя ошибки предыдущих моделей. Градиентный бустинг способен эффективно моделировать сложные зависимости в данных и обеспечивает высокую точность прогнозирования.

**Метод опорных векторов (Support Vector Machines, SVM)** используется для задач классификации и регрессии. Он находит оптимальное разделение классов в пространстве признаков с помощью опорных векторов, что позволяет достичь высокой разделяющей способности. SVM подходит для прогнозирования вероятности патентоспособности на основе множества признаков из патентных данных.

Использование указанных методов машинного обучения позволяет с высокой точностью прогнозировать вероятность патентоспособности конкретного изобретения.

### **3.4 Автоматизация процесса создания структурированных отчетов о патентном исследовании на основе полученных результатов помощью ИИ**

Создание структурированных отчетов является важным этапом в патентных исследованиях, поскольку предоставление информации в удобной и понятной форме является ключевым для принятия обоснованных решений. Использование искусственного интеллекта (ИИ) в этом процессе позволяет значительно повысить эффективность и точность создания отчетов, а также сделать их более информативными и полезными для пользователей.

Полученные результаты анализа патентных данных структурируются и организуются в соответствии с predetermined шаблоном отчета. Используя методы семантического анализа и машинного обучения, система ИИ определяет связи и зависимости между различными элементами информации, что позволяет создать целостный и логически связанный отчет. Здесь следует отметить, что на данный момент нет конкретных систем или программных решений, которые обеспечивающий автоматизацию процесса создания отчетов по патентным исследованиям в соответствии с ГОСТом.

Одним из важных аспектов создания структурированных отчетов является **визуализация данных**. Система ИИ использует различные методы визуализации, такие как графики, диаграммы и таблицы, чтобы наглядно представить полученные результаты. Визуализация помогает сделать отчет

более понятным и удобным для восприятия, а также выявить тенденции и закономерности в патентных данных.

После завершения процесса анализа и структурирования данных, система ИИ генерирует **пользовательский отчет**. В отчете содержатся все полученные результаты, представленные в понятной и удобной для пользователя форме. Отчет может включать краткое описание найденных патентов, ключевые слова, классификацию по МПК, связанные патентные документы и другую важную информацию.

### **3.5 Анализ больших объемов патентных данных с помощью искусственного интеллекта (ИИ)**

Анализ больших объемов патентных данных является критическим аспектом патентных исследований, поскольку позволяет выявить тенденции развития технологий, обнаружить новые технические решения и оценить конкурентное положение в определенной области.

**Обнаружение тенденций развития технологий и выявление новых технических решений** с помощью искусственного интеллекта (ИИ) является одним из ключевых преимуществ его применения в анализе патентных данных. ИИ обладает уникальными возможностями обработки и анализа больших объемов информации, что позволяет исследователям и бизнес-аналитикам выявлять важные тенденции и инновационные направления в различных отраслях. В данном разделе рассмотрим, как искусственный интеллект способствует обнаружению тенденций развития технологий и выявлению новых технических решений.

Используя методы машинного обучения, ИИ способен **выявлять инновационные направления** в патентных данных [12]. Алгоритмы кластеризации и классификации помогают группировать похожие патенты и выделить наиболее перспективные технологические решения. Это позволяет исследователям определить, какие области технологий находятся в стадии активного развития и являются приоритетными для дальнейшего изучения.

С помощью искусственного интеллекта можно осуществлять **прогнозирование будущих технологических трендов** на основе анализа патентных данных (см., например, [13]). ИИ способен определить, какие технологии и инновации находятся в стадии активного развития, и предсказать их будущее развитие. Это позволяет компаниям и организациям принимать информированные решения о направлении своих исследовательских и разработочных проектов.

Искусственный интеллект помогает **выявить потенциальных партнеров** для сотрудничества и инвестиционных возможностей на основе анализа патентных данных [14]. Алгоритмы анализа позволяют выделить компании и организации, которые активно разрабатывают новые технологии и имеют перспективные патенты. Это позволяет исследователям и бизнес-аналитикам находить потенциальных партнеров для сотрудничества, а инвесторам – инноваторов

#### **4. Обзор и сравнение основных систем патентных исследований с функциями искусственного интеллекта**

На рынке существует несколько платных сервисов для патентных исследований, которые используют искусственный интеллект и машинное обучение для обработки и анализа патентных данных. Вот наиболее известные из них, условно ранжированные по популярности: **PatSnap, Clarivate Analytics (ранее Derwent Innovation), Questel Orbit, PatentSight, CIPHER, PatSeer, IP.com.**

Также некоторые функции ИИ постепенно интегрируются в патентные сервисы стран и организаций. Например, в системе поиска SearchPlatform Роспатента реализован «Поиск похожих документов с использованием искусственного интеллекта» ([searchplatform.rospatent.gov.ru/equal\\_docs](http://searchplatform.rospatent.gov.ru/equal_docs)), который принимает на вход не поисковый запрос, а текстовое описание изобретения. В сервисе Google Patents поиск патентов осуществляется по обычным запросам, однако далее система позволяет использовать ИИ для анализа уровня техники для найденных патентов и поиска их аналогов.

#### **5. Список литературы**

- [1] Achille Souili, Denis Cavallucci, François Rousselot, Natural Language Processing (NLP) – A Solution for Knowledge Extraction from Patent Unstructured Data, *Procedia Engineering*, Volume 131, 2015, 635-643
- [2] Hongxun Jiang, Shaokun Fan, Nan Zhang, Bin Zhu, Deep learning for predicting patent application outcome: The fusion of text and network embeddings, *Journal of Informetrics*, Volume 17, Issue 2, May 2023, 101402
- [3] Cascini, Gaetano, and Federico Neri, Natural Language Processing for patents analysis and classification. In *ETRIA World Conf., TRIZ Future*, 199-212. 2004.
- [4] Arts, Sam, Jianan Hou, and Juan Carlos Gomez. "Natural language processing to identify the creation and impact of new technologies in patent text: Code, data, and new measures." *Research Policy*, 50, no. 2 (2021): 104144.

- [5] Wang, B., Liu, S., Ding, K. et al. Identifying technological topics and institution-topic distribution probability for patent competitive intelligence analysis: a case study in LTE technology. *Scientometrics*, 101, 685–704 (2014).
- [6] Giovanni Puccetti, Vito Giordano, Irene Spada, Filippo Chiarello, Gualtiero Fantoni, Technology identification from patent texts: A novel named entity recognition method, *Technological Forecasting and Social Change*, Volume 186, Part B, January 2023, 122160
- [7] Anna Maria Villa, Manuel Wirz, A sequential patent search approach combining semantics and artificial intelligence to identify initial State-of-the-Art documents, *World Patent Information*, Volume 68, March 2022, 102096
- [8] Trappey AJC, Liang C-P, Lin H-J. Using Machine Learning Language Models to Generate Innovation Knowledge Graphs for Patent Mining. *Applied Sciences*. 2022; 12(19):9818.
- [9] L. Lei, J. Qi and K. Zheng, "Patent Analytics Based on Feature Vector Space Model: A Case of IoT," in *IEEE Access*, vol. 7, pp. 45705-45715, 2019
- [10] Helen Niemann, Martin G. Moehrle, Jonas Frischkorn, Use of a new patent text-mining and visualization method for identifying patenting patterns over time: Concept, method and test application, *Technological Forecasting and Social Change*, Volume 115, February 2017, 210-220
- [11] Lee C-W, Tao F, Ma Y-Y, Lin H-L. Development of Patent Technology Prediction Model Based on Machine Learning. *Axioms*. 2022; 11(6):253.
- [12] Kwon, U., Geum, Y. Identification of promising inventions considering the quality of knowledge accumulation: a machine learning approach. *Scientometrics* 125, 1877–1897 (2020).
- [13] M. Shibata, Y. Ohtsuka, M. Takahashi and K. Okamoto, "Advanced FPGA technology trend based on patent analysis with link mining," 2018 International Conference on Electronics Packaging and iMAPS All Asia Conference (ICEP-IAAC), Mie, Japan, 2018, pp. 147-151
- [14] Aristodemou, L., Tietze, F., 2018. The state-of-the-art on Intellectual Property Analytics (IPA): A literature review on artificial intelligence, machine learning and deep learning methods for analysing intellectual property (IP) data. *World Patent Information (WPI)* 55, 37–51.