

Проблеми обробки інформації судочинства України та NLP-технології штучного інтелекту

С. С. Мірко¹, П. П. Маслянко¹

¹*КПІ ім. Ігоря Сікорського, Київ, Україна*

sergymirko@gmail.com

Анотація

Сучасне судочинство України стикається з рядом проблем, зокрема залежністю від людського чинника, обмеженою гнучкістю, значними витратами часу та ресурсів. Обробка правової інформації ускладнена її масштабністю, нестандартною структурою та необхідністю постійного оновлення. Використання NLP-технологій та штучного інтелекту дозволяє автоматизувати аналіз юридичних текстів, покращити розуміння правових норм і пришвидшити судові процеси. Запропонований розмірково-пошуковий алгоритм забезпечує ефективну обробку та формування правових позицій для підготовки судових рішень.

Ключові слова: судочинство; штучний інтелект; NLP; автоматизація; правова інформація.

1 Вступ

Традиційні методи обробки інформації судочинства мають ряд важливих недоліків, таких як залежність від людських ресурсів, відсутність гнучкості та значне споживання ресурсів. Ці проблеми включають:

- **Залежність від людського прийняття рішень.** Судочинство традиційно спирається на рішення суддів, прокурорів та адвокатів, що включає аналіз справи, тлумачення норм і винесення рішень [1].
- **Використання прецедентів.** Судді використовують попередні рішення як орієнтири для ухвалення нових рішень, що забезпечує узгодженість правозастосування [1].
- **Обмежена гнучкість.** У випадках правової невизначеності судді мають враховувати контекст для прийняття рішення [1].

- **Затрати часу та ресурсів.** Судові процеси вимагають значних людських, фінансових та інформаційно-комунікаційних ресурсів, що затягує їх розгляд [1].

2 Особливості юридичних даних

Юридична інформація має такі характерні риси:

- **Неструктурованість.** Правова інформація часто представлена у вигляді тексту, що містить закони, судові рішення, юридичні коментарі тощо, зазвичай у нестандартизованому форматі. Для трансформації таких даних у структурований формат застосовуються методи обробки природної мови (NLP) [1].
- **Масштабність і складність.** Правова інформація зазвичай має великий обсяг і складну структуру. Наприклад, законодавчі акти або судові рішення можуть налічувати сотні сторінок. Крім того, різні правові сфери вимагають спеціалізованих знань і підходів до аналізу. [1].
- **Актуальність.** Право є динамічною галуззю, де законодавство та судова практика постійно змінюються. Тому правові бази даних потребують регулярного оновлення, щоб відображати останні зміни [1].
- **Приватність і безпека.** Правова інформація часто містить чутливі дані. Це вимагає суворих заходів із забезпечення конфіденційності, таких як анонімізація даних і контроль доступу [1].

3 Застосування NLP та ШІ в судочинстві

Встановлені проблеми та особливості юридичних інформаційних ресурсів спонукають правників, математиків та інженерів до пошуку методів, моделей та інструментальних засобів для вирішення цих проблем. Одним із таких підходів до вирішення вказаних проблем можуть бути методи, моделі та інструментальні засоби штучного інтелекту [2]. Штучний інтелект, зокрема великі мовні моделі, мають великий потенціал покращити ефективність і якість судових процесів, що стає особливо важливим у умовах зростаючого навантаження на суди.

- **Аналіз юридичних документів.** LLM можуть перевіряти граматику, структуру речень і виділяти ключові елементи (спірні питання, норми права, сторони, докази), спрощуючи підготовку до розгляду.
- **Генерація юридичного контенту.** ШІ створює позови, контракти, судові звіти на основі базових даних, що прискорює підготовку документів [1].

- **Розпізнавання мови.** Автоматизує створення протоколів судових засідань, зменшуючи неточності ручного запису [1].
- **Юридичні консультації.** ШІ надає поради на основі правових баз, знижуючи навантаження на адвокатів [1].
- **Пошук оптимальних рішень.** Аналізуючи бази рішень, ШІ пропонує найкращі варіанти, сприяючи справедливості [1].
- **Логічний аналіз справ.** ШІ оцінює релевантність доказів і формує обґрунтування для судових рішень [1].
- **Автоматизація судових процедур.** Зменшує адміністративне навантаження автоматизуючи рутинні задачі, підвищуючи швидкість розгляду справ [1].

Розмірково-пошуковий алгоритм спрямований на формування та обробку правових позицій, а не повних текстів судових рішень. Такий підхід дозволяє зменшити вплив нерелевантного контенту та збільшити кількість розглянутих ситуацій. Створення правової позиції здійснюється в декілька етапів.

- З кожного судового рішення витягуються посилання на правові норми, що використовуються (статті кодексів або законів).
- Для коректної роботи LLM створюється база даних структурованих правових норм. Наприклад, якщо згадується “ст. 234 ч. 2 КПК”, модель отримує лише текст, що стосується другої частини статті 234 КПК, а не всю статтю.
- Формується правова підказка судового рішення, яка складається з текстів використаних правових норм у судовому рішенні.
- Використовуючи підказки, LLM визначає послідовність думок щодо правових норм, що складається з таких елементів:
- На основі послідовності думок створюється висновок про коректне застосування правових норм та їх взаємодію. Цей висновок являє собою правову позицію.

Розмірково-пошуковий алгоритм використовує три основних компоненти.

1. Компонент пошуку:

- Пошук релевантних правових позицій у зовнішній базі знань за пошуковим запитом користувача.

2. Компонент формування альтернатив:

- Ранжування правових позицій для усунення схожих позицій і забезпечення різноманітності дерева альтернатив.
- Генерація дерева альтернатив LLM на основі релевантних позицій. Дерево зберігається між ітераціями для подальшого розширення.
- Агрегація дерева за допомогою LLM, що вибирає найкращу альтернативу для вирішення задачі.

3. Компонент генерації відповіді:

- Валідація дерева альтернатив LLM для видалення нерелевантних гілок.
- Генерація остаточної відповіді на основі валідованого дерева.
- Агент вирішує, чи завершувати роботу алгоритму, чи продовжувати генерацію.

Алгоритм циклічно створює, обрізає та розширює дерево альтернатив, самостійно обираючи інструменти чи завершуючи розгляд, що забезпечує йому саморефлексію та адаптивність у вирішенні правових задач.

4 Висновки

- Визначено основні проблеми судочинства України, зокрема залежність від людських рішень, високі витрати часу та складність обробки юридичних даних.
- Обґрунтовано доцільність використання NLP та штучного інтелекту для автоматизації аналізу судових документів і пошуку релевантних рішень.
- Запропоновано алгоритм, що будує дерево альтернатив для ефективного аналізу судових рішень, підвищуючи точність правових позицій.
- Порівняльний аналіз підтвердив конкурентоспроможність запропонованого алгоритму та його переваги у автоматизації пошуку судових рішень.

Перелік посилань

- [1] J. Lai, W. Gan, J. Wu, Z. Qi, and P. S. Yu, "Large Language Models in Law: A Survey," *arXiv preprint arXiv:2312.03718*, 2023.
- [2] П.П. Маслянюк, С.С. Мірко, "Концептуальна модель NLP-системи пошуку релевантної правової позиції на множині судових рішень судочинства України", *Наукові вісті КПП*, No 1–4, с. 71–85, 2023. <https://doi.org/10.20535/kpissn.2023.1-4.310302>