

ПЛАГИАТСТВОТО И АВТОМАТИЗИРАНИТЕ СИСТЕМИ ЗА ВЕРИФИКАЦИЯ НА ФАКТИ, ИЗПОЛЗВАЩИ ИИ¹

Доц. д-р Валентина Александрова, Катедра ППН, Правно-исторически факултет, ЮЗУ Неофит Рилски, E-mail: aleksandrovalentina@abv.bg

Резюме: Плагиатството е сериозен етичен и правен проблем, изискващ внимателно разглеждане и стриктни мерки за предотвратяване и санкциониране. Етичните аспекти включват нарушаване на академичната и професионалната честност, както и неуважение към интелектуалната собственост на оригиналния автор. Правните аспекти включват нарушения на авторските права, които могат да доведат до правни действия и санкции. Антиплагиат системите, използващи изкуствен интелект (ИИ), са ефективни инструменти за откриване и предотвратяване на плагиатство. Те използват технологии като обработка на естествен език, алгоритми за сравнение на текстове, машинно обучение и дълбоко обучение за идентифициране на сходства и перифразирани текстове. Системите като Turnitin, Grammarly и Unicheck предлагат контекстуален анализ и интеграция с различни платформи за улесняване на процеса на проверка и редактиране на текстове.

Ключови думи: Плагиатство; Авторски права; Етичен проблеми; Антиплагиат системи; Изкуствен интелект

PLAGIARISM AND AUTOMATED FACT-VERIFICATION SYSTEMS USING AI

Assoc. Prof. Valentina Aleksandrova, PhD, Department of Public Law Sciences, History Faculty of Law and History, South-West University "Neofit Rilski", E-mail: aleksandrovalentina@abv.bg

Abstract: Plagiarism is a serious ethical and legal issue that requires careful consideration and strict measures for prevention and sanctioning. Ethical aspects include violations of academic and professional integrity, as well as disrespect for the intellectual property of the original author. Legal aspects include copyright infringements, which can lead to legal actions and sanctions. Anti-plagiarism systems using artificial intelligence (AI) are effective tools for detecting and preventing plagiarism. They employ technologies such as natural language processing, text comparison algorithms, machine learning, and deep learning to identify similarities and

¹ Докладът е представен на кръгла маса, проведена на 22 май 2024 г., на тема: „ИЗКУСТВЕН ИНТЕЛЕКТ, СИГУРНОСТ И ДЕЗИНФОРМАЦИЯ - СЪВРЕМЕННИ ТЕНДЕНЦИИ И ПРЕДИЗВИКАТЕЛСТВА“, която се организира от Правно-историческия факултет на Югозападния университет „Неофит Рилски“, гр. Благоевград в рамките на научноизследователски проект № RP-C1/24 на тема: „Изкуствен интелект, сигурност и дезинформация - съвременни тенденции и предизвикателства“.

paraphrased texts. Systems like Turnitin, Grammarly, and Unicheck offer contextual analysis and integration with various platforms to facilitate the process of checking and editing texts.

Keywords: *Plagiarism; Copyright; Ethical issue; Anti-plagiarism systems; Artificial intelligence*

Плагиатството е сериозен етичен и правен проблем, който изисква внимателно разглеждане и строги мерки за предотвратяване и санкциониране. Разбирането на етичните и правните аспекти на плагиатството е от съществено значение за всеки, който се занимава с академична или професионална дейност. Спазването на етичните норми и законодателството за авторските права² е ключово за поддържане на честност, почтеност и уважение към интелектуалния труд.

Този феномен – плагиатство, представлява използването на чужди идеи, текстове или други произведения като свои, без да се даде надлежното признание на оригиналния автор чрез неговото посочване.

Етичните аспекти на плагиатството могат да се формулират по следния начин:

- Плагиатството нарушава основните принципи на академичната и професионална честност. То е форма на измама, при която лицето се представя за автор на труд, който не е създаден от него.
- Когато се използват чужди материали без разрешение и признание, това показва неуважение към труда и интелектуалната собственост на оригиналния автор.

² ЗАКОН ЗА АВТОРСКОТО ПРАВО И СРОДНИТЕ МУ ПРАВА В сила от 01.08.1993 г. Отражена деноминацията от 05.07.1999 г. Обн. ДВ. бр.56 от 29 Юни 1993г., ... изм. и доп. ДВ. бр.100 от 1 Декември 2023г.

Изключения

Чл. 4. (Доп. - ДВ, бр. 21 от 2014 г.; доп., бр. 100 от 2023 г., в сила от 01.12.2023 г.) Не са обект на авторското право:

1. (доп. - ДВ, бр. 21 от 2014 г.) нормативни и индивидуални актове на държавни органи за управление, актовете на съдилищата, както и официалните им преводи;

Предположение за авторство

Чл. 6. (1) (Предишен текст на чл. 6, изм. и доп. - ДВ, бр. 99 от 2005 г., в сила от 11.01.2006 г.) До доказване на противното за автор на произведението се смята лицето, чието име или друг идентифициращ знак са посочени по обичайния за това начин върху оригинала на произведението, копия или екземпляри от него и/или техните опаковки.

Чл. 24. (Изм. - ДВ, бр. 77 от 2002 г., в сила от 01.01.2003 г.; изм., бр. 25 от 2011 г., в сила от 25.03.2011 г.; изм., бр. 28 от 2018 г., в сила от 29.03.2018 г.; изм., бр. 94 от 2018 г.; изм. и доп., бр. 100 от 2023 г., в сила от 01.12.2023 г.) (1) Без съгласието на носителя на авторското право и без заплащане на възнаграждение е допустимо: [...].

2. (изм. - ДВ, бр. 100 от 2023 г., в сила от 01.12.2023 г.) използването на цитати от вече разгласени произведения на други лица за цели като критика или обзор при посочване на източника и името на автора, освен ако това е невъзможно; цитирането трябва да съответства на обичайната практика и да е в обем, оправдан от целта;

- Плагиатството може да навреди сериозно на репутацията на възползвалият се от чужд труд в академичните среди и на работното място, като това може да доведе до загуба на доверие, авторитет и уволнение.
- Ако плагиатството не бъде санкционирано по надлежния ред, то може да стимулира подобно неетично поведение сред колеги, студенти и професионалисти, създавайки култура на неетичност.

От правна гледна точка плагиатството често представлява нарушение на авторските права, което може да доведе до правни действия от страна на оригиналния автор. Авторските права защитават произведенията на интелектуалната дейност и дават на автора изключителното право да използва и разпространява своето произведение. В тази връзка, много академични и професионални институции и организации изискват от своите членове и служители, да се придържат към определени стандарти за честност и оригиналност. Нарушаването на тези договорни задължения може да доведе до дисциплинарни мерки, включително уволнение или отстраняване от организацията.

В зависимост от конкретната юрисдикция, плагиатството може да доведе до налагането на глоби и санкции. В някои страни има строго законодателство, което предвиждат наказателна отговорност за нарушение на авторските права, включително финансови санкции и дори затвор.

Публикуването на чужд труд или обект на авторско право от свое име, може да доведе до правни действия от страна на засегнатите лица и институции, което може да нанесе сериозни репутационни щети и правни последици за лицето, което е извършило плагиатството.

Какво представляват антиплагиат системите, които все по-често се използват за установяване или за предотвратяване на плагиатството?

Антиплагиат системите, известни също като системи за откриване на плагиатство, работят на принципа на сравнение на текстовете и анализ на сходствата между тях. Основните технологии и методи, използвани от тези системи, включват:

1. Обработка на естествен език (NLP). Антиплагиат системите използват техники за обработка на естествен език за анализа на текстовете, като изследват граматиката, синтаксиса и семантиката на изреченията. Това им позволява да откриват не само

идентични фрази, но и перифразирани изрази, които могат да бъдат индикатор за плагиатство.

2. Алгоритми за сравнение на текстове. Антиплагиат системите използват различни алгоритми за сравнение на текстове, които включват:

- Търсене на точни съвпадения чрез откриване на буквални съвпадения на фрази и изречения между документи.
- Шинглинг (Shingling) техника за разделяне на текста на последователности от думи (наречени "шингли") и сравняване на тези последователности между различни документи. Приложения на Shingling
 - ✓ Откриване на плагиатство: Техниката се използва за идентифициране на сходства между документи и откриване на дублирано или плагиатствано съдържание.
 - ✓ Информационно извличане: За подобряване на резултатите от търсенето чрез откриване на документи със сходно съдържание.
 - ✓ Клъстериране на документи: За групиране на документи по тематична близост.

Техниката Shingling е ефективен начин за структуриране на текста и сравняване на съдържанието на документи, което я прави незаменим инструмент в обработката на естествен език и анализ на текстове.

Алгоритми за сравнение на документи: Алгоритми като Jaccard index, Cosine similarity и други³, които изчисляват степента на сходство между два текста на база на честотата и разпределението на думите.

³ Jaccard Index

Описание: Jaccard индексът измерва сходството между два набора, като изчислява съотношението между броя на общите елементи и общия брой уникални елементи в двата набора.

Формула:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

където A и B са множествата от думи (или шингли) на двата текста. $|A \cap B|$ е броят на общите думи, а $|A \cup B|$ е броят на уникалните думи в двата текста.

Пример: Текст 1: {"котка", "куче", "мишка"} Текст 2: {"котка", "мишка", "зайче"}

$|A \cap B| = 2$ ($|A \cap B| = 2$ (общи думи: "котка", "мишка")) $|A \cup B| = 4$ ($|A \cup B| = 4$ (уникални думи: "котка", "куче", "мишка", "зайче"))

Jaccard индекс = $24/42=0.571$

Cosine Similarity

Описание: Cosine similarity измерва косинуса на ъгъла между два векторни представяния на текстове в n-мерно пространство. Този метод е особено полезен при анализ на честота и разпределение на думи в документи.

Формула:

$$\cos(\vec{A}, \vec{B}) = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\| \|\vec{B}\|} \cos(A, B) = \frac{A \cdot B}{\|A\| \|B\|}$$

където $\vec{A} \cdot \vec{B}$ е скаларното произведение на векторите, а $\|\vec{A}\|$ и $\|\vec{B}\|$ са нормите (дължините) на векторите.

Пример: Да предположим, че имаме два текста с векторни представяния: Текст 1: [1, 1, 0, 1] (думите: "котка", "куче", "мишка", "зайче") Текст 2: [1, 0, 1, 1]

$$\vec{A} \cdot \vec{B} = (1 \times 1) + (1 \times 0) + (0 \times 1) + (1 \times 1) = 2 \quad A \cdot B = (1 \times 1) + (1 \times 0) + (0 \times 1) + (1 \times 1) = 2$$

$$\|\vec{A}\| = \sqrt{1^2 + 1^2 + 0^2 + 1^2} = 2 \quad \|A\| = \sqrt{1^2 + 1^2 + 0^2 + 1^2} = 2$$

$$\|\vec{B}\| = \sqrt{1^2 + 0^2 + 1^2 + 1^2} = 2 \quad \|B\| = \sqrt{1^2 + 0^2 + 1^2 + 1^2} = 2$$

$$\cos(\vec{A}, \vec{B}) = \frac{2}{2 \times 2} = 0.5 \quad \cos(A, B) = \frac{2}{2 \times 2} = 0.5$$

Dice's Coefficient

Описание: Dice's coefficient е подобен на Jaccard индекса, но дава двоен акцент на броя на общите елементи.

Формула:

$$D(A, B) = \frac{2|A \cap B|}{|A| + |B|} \quad D(A, B) = \frac{2|A \cap B|}{|A| + |B|}$$

Пример: Текст 1: {"котка", "куче", "мишка"} Текст 2: {"котка", "мишка", "зайче"}

$$|A \cap B| = 2 \quad |A \cap B| = 2 \quad |A| = 3 \quad |B| = 3$$

$$\text{Dice's coefficient} = \frac{2 \times 2}{3 + 3} = 0.67$$

Euclidean Distance

Описание: Euclidean distance измерва разстоянието между два вектора в n-мерно пространство.

Формула:

$$d(A, B) = \sqrt{\sum_{i=1}^n (A_i - B_i)^2} \quad d(A, B) = \sqrt{\sum_{i=1}^n (A_i - B_i)^2}$$

Пример: Векторите на текстовете: Текст 1: [1, 1, 0, 1] Текст 2: [1, 0, 1, 1]

$$d(A, B) = \sqrt{(1-1)^2 + (1-0)^2 + (0-1)^2 + (1-1)^2} = \sqrt{0 + 1 + 1 + 0} = \sqrt{2} \approx 1.41$$

Manhattan Distance

Описание: Manhattan distance (такситна разстояние) измерва разстоянието между два вектора чрез сумиране на абсолютните разлики между съответстващите им компоненти.

Формула:

$$d_{man}(A, B) = \sum_{i=1}^n |A_i - B_i| \quad d_{man}(A, B) = \sum_{i=1}^n |A_i - B_i|$$

Пример: Векторите на текстовете: Текст 1: [1, 1, 0, 1] Текст 2: [1, 0, 1, 1]

$$d_{man}(A, B) = |1-1| + |1-0| + |0-1| + |1-1| = 0 + 1 + 1 + 0 = 2$$

TF-IDF (Term Frequency-Inverse Document Frequency)

Описание: TF-IDF е статистическа мярка, която оценява значимостта на дадена дума в документ, относително спрямо цялото корпус от документи.

Формули:

- Term Frequency (TF):
 $TF(t, d) = \text{брой на появявания на } t \text{ в } d$
 $TF(t, d) = \text{брой на появявания на } t \text{ в } d$
- Inverse Document Frequency (IDF):
 $IDF(t) = \log\left(\frac{N}{df(t)}\right)$
 $IDF(t) = \log\left(\frac{\text{общ брой документи}}{\text{брой документи, съдържащи } t}\right)$

3. Машинно обучение. Системите могат да използват модели на машинно обучение⁴, обучени върху големи набори от данни, за да разпознават сложни модели на плагиатство.

-
- **TF-IDF:** $TF\text{-}IDF(\Pi, \Pi) = TF(\Pi, \Pi) \times IDF(\Pi)$ $TF\text{-}IDF(t, d) = TF(t, d) \times IDF(t)$

Пример: Ако думата "котка" се появява 3 пъти в документ с 100 думи и се среща в 10 от общо 1000 документи, тогава:

$$TF(\text{"котка"}, \Pi) = 3/100 = 0.03 \quad TF(\text{"котка"}, d) = 100/3 = 0.03$$

$$IDF(\text{"котка"}) = \log\left[\frac{1000}{10}\right] = \log\left[\frac{100}{3}\right] = 2 \quad IDF(\text{"котка"}) = \log\left[\frac{1000}{10}\right] = \log 100 = 2$$

$$TF\text{-}IDF(\text{"котка"}, \Pi) = 0.03 \times 2 = 0.06 \quad TF\text{-}IDF(\text{"котка"}, d) = 0.03 \times 2 = 0.06$$

Тези методи и алгоритми са основни в областта на обработката на естествен език (NLP) и предоставят различни начини за измерване на сходството между текстове въз основа на честотата и разпределението на думите.

⁴ Моделите на машинно обучение, обучени върху големи набори от данни, са способни да разпознават сложни модели на плагиатство. Ето няколко примера за такива модели и подходите, които използват:

1. Трансформъри (Transformers)

Трансформърите, като BERT (Bidirectional Encoder Representations from Transformers) и GPT (Generative Pre-trained Transformer), са мощни модели за обработка на естествен език, които могат да бъдат използвани за разпознаване на плагиатство.

- **BERT:**
 - **Обучение:** BERT е предварително обучен върху голям корпус от текст и след това може да бъде финетюниран върху специфичен набор от данни за плагиатство.
 - **Подход:** Текстовете се разделят на сегменти, които се кодират в контекстуални векторни представяния. Сравняването на тези представяния може да разкрие сходства на дълбоко ниво.
 - **Пример:** Да предположим, че имаме два документа. BERT може да се използва за създаване на векторни представяния за всеки сегмент от текста и след това да се измери сходството между тях чрез cosine similarity или други метрики.

2. Recurrent Neural Networks (RNNs)

RNN и техните вариации, като LSTM (Long Short-Term Memory) и GRU (Gated Recurrent Unit), са подходящи за анализ на последователности от текст, което е полезно при откриване на плагиатство.

- **LSTM:**
 - **Обучение:** LSTM моделите се обучават върху последователности от текст, за да предсказват следващите думи или фрази, което помага при разпознаване на пренаписани или перефразирани пасажии.
 - **Пример:** Документите се разделят на речеви последователности и се подават на LSTM модел, който идентифицира сходства между последователностите на думи в различни документи.

3. Convolutional Neural Networks (CNNs)

CNN обикновено се използват за анализ на изображения, но са ефективни и при текстови данни чрез текстови ембединг представяния.

- **CNN:**
 - **Обучение:** Текстовете се представят като матрици (например чрез word embeddings като Word2Vec или GloVe) и се подават на CNN за извличане на характеристики и модели.
 - **Пример:** Два документа се кодират в ембединг представяния и се подават на CNN, който извлича характерни черти и открива модели на плагиатство.

4. Siamese Neural Networks

Siamese мрежите са специализирани в задачите за сравнение на двойки.

- **Siamese Network:**
 - **Обучение:** Обучават се върху двойки текстове, като се минимизира разстоянието между представянията на сходни текстове и се максимизира разстоянието между различни текстове.

Тези модели могат да идентифицират перифразирани текстове и да разграничават случайни съвпадения и съзнателно копиране.

4. Бази данни и индексирание. Антиплагиат системите разчитат на обширни бази данни с академични публикации⁵, уебсайтове, студентски работи и други източници на текстове. Те

-
- **Пример:** Две версии на текст се подават в паралелни мрежи, които споделят същите тегла. Мрежата изчислява сходството между техните векторни представяния, като по този начин открива плагиатство.

5. Ensemble Methods

Комбинация от различни модели и подходи може да даде по-добри резултати при разпознаване на плагиатство.

- **Обучение:** Използват се различни модели (например, трансформъри, RNNs, и CNNs), като се комбинират резултатите им чрез ансамблов метод (например, voting, stacking).
- **Пример:** Изходът на няколко модела се комбинира, за да се вземе крайното решение за това дали даден текст е плагиатствен или не.

6. Feature-based Methods

Използване на класически подходи с характеристики, извлечени от текстовете.

- **Обучение:** Извличат се различни характеристики (например, TF-IDF, n-grams, syntactic features) и се подават на класически модели за машинно обучение като SVM (Support Vector Machines), Random Forest, или Gradient Boosting.
- **Пример:** Създава се характеристично представяне на текстовете, което се използва за обучение на модел, който класифицира текстовете като плагиатствани или не.

Заклучение

Моделите на машинно обучение, обучени върху големи набори от данни, са изключително ефективни при разпознаване на сложни модели на плагиатство. Комбинирането на различни подходи и техники, като трансформъри, RNNs, CNNs, Siamese мрежи и ансамблови методи, позволява създаването на мощни системи за откриване на плагиатство.

⁵ Съществуват множество бази данни с академични публикации, които предоставят достъп до огромен обем научни изследвания и статии. Ето няколко от най-известните и широко използвани бази данни:

1. PubMed

- **Описание:** PubMed е свободно достъпна база данни, която включва над 30 милиона цитати и резюмета на биомедицински литература от MEDLINE, списания по биологични науки и онлайн книги.
- **Предметни области:** Медицина, биология, здравеопазване.
- **Уебсайт:** [PubMed](https://pubmed.ncbi.nlm.nih.gov/)

2. Google Scholar

- **Описание:** Google Scholar е свободно достъпен търсач за академична литература. Той индексира разнообразие от научни дисциплини и формати, включително статии, книги, конференционни доклади и патенти.
- **Предметни области:** Мултидисциплинарна.
- **Уебсайт:** [Google Scholar](https://scholar.google.com/)

3. IEEE Xplore

- **Описание:** IEEE Xplore е цифрова библиотека за изследвания в областта на електротехниката, електрониката, компютърните науки и свързани технологии. Базата данни съдържа статии от списания, конференционни доклади, стандарти и други.
- **Предметни области:** Електротехника, електроника, компютърни науки.
- **Уебсайт:** [IEEE Xplore](https://ieeexplore.ieee.org/)

4. ScienceDirect

- **Описание:** ScienceDirect е водеща пълнотекстова научна база данни, която предоставя достъп до статии и глави от книги в областите на науката, технологиите и медицината.

индексират тези документи, за да могат бързо и ефективно да сравняват нови текстове с вече съществуващите.

5. Проверка на интернет съдържание. Много системи извършват търсения в интернет, за да откриват съвпадения с публично достъпни уебсайтове, блогове и форуми. Те използват уеб обхождане и индексирание за събиране на информация от интернет в реално време.

6. Рапорти за сходство. След извършване на анализа, антиплагиат системите генерират доклади, които показват процентът на сходство между проверявания текст и откритите източници. Те също така подчертават идентичните и сходни пасажки, като предоставят връзки към оригиналните източници.

-
- **Предметни области:** Наука, технологии, медицина.
 - **Уебсайт:** [ScienceDirect](#)

5. SpringerLink

- **Описание:** SpringerLink е платформа, която предоставя достъп до милиони научни документи от списания, книги, протоколи и справочници в различни научни дисциплини.
- **Предметни области:** Мултидисциплинарна.
- **Уебсайт:** [SpringerLink](#)

6. JSTOR

- **Описание:** JSTOR е дигитална библиотека, която предоставя достъп до хиляди академични списания, книги и първични източници в различни области на изследване.
- **Предметни области:** Хуманитарни науки, социални науки, изкуства, наука.
- **Уебсайт:** [JSTOR](#)

7. ACM Digital Library

- **Описание:** ACM Digital Library е база данни, която предоставя достъп до публикации на Association for Computing Machinery, включително списания, конференционни доклади и технически доклади.
- **Предметни области:** Компютърни науки, информационни технологии.
- **Уебсайт:** ACM Digital Library

8. Web of Science

- **Описание:** Web of Science е платформа за научно цитиране, която включва множество бази данни за академични статии, позволявайки на потребителите да откриват цитирания и да анализират връзките между изследванията.
- **Предметни области:** Мултидисциплинарна.
- **Уебсайт:** [Web of Science](#)

9. ProQuest

- **Описание:** ProQuest е платформа, която предлага достъп до множество бази данни, обхващащи различни дисциплини. Тя включва академични списания, дисертации, вестници, книги и други източници.
- **Предметни области:** Мултидисциплинарна.
- **Уебсайт:** [ProQuest](#)

10. arXiv

- **Описание:** arXiv е отворен архив за електронни препринти на научни статии в областите на физиката, математиката, компютърните науки, квантовата биология и други.
- **Предметни области:** Физика, математика, компютърни науки.
- **Уебсайт:** [arXiv](#)

Тези бази данни предоставят богат източник на академична информация, която може да бъде използвана за изследвания, анализи и откриване на плагиатство.

7. Контекстуален анализ. Някои напреднали системи извършват контекстуален анализ на текстовете, за да оценят дали съвпаденията са в рамките на допустимите граници, като цитати или общоизвестни фрази, или представляват неправомерно копиране. Антиплагиат системите, които използват контекстуален анализ, са по-усъвършенствани и могат да идентифицират плагиатство дори когато текстът е преформулиран или парафразиран. Те използват технологии като машинно обучение и изкуствен интелект за анализ на смисъла и структурата на текста, вместо да разчитат само на точно съвпадение на думи. Ето няколко примера за такива системи:

1. Turnitin е една от най-популярните и широко използвани антиплагиат системи в академичните среди. Тя използва контекстуален анализ, за да сравнява текстове и да идентифицира сходства дори когато текстът е преформулиран. Системата анализира структурата, фразите и стилистичните елементи на текста.
2. Grammarly е известен инструмент за проверка на граматиката и стила, но също така предлага и антиплагиат функции. Той използва контекстуален анализ, за да открие плагиатство, като сравнява текста с огромна база данни от академични публикации, уебсайтове и други източници.
3. Unicheck е антиплагиат система, която използва контекстуален анализ, за да проверява документи за уникалност. Тя анализира текстовете на семантично ниво и може да идентифицира сходства, дори когато текстът е преформулиран или парафразиран.
4. Copyleaks използва усъвършенствани алгоритми за контекстуален анализ, за да идентифицира плагиатство. Тази система проверява текста за сходства с други текстове в база данни, която включва милиарди уеб страници, академични публикации и други източници.
5. Plagscan също използва контекстуален анализ за откриване на плагиатство. Системата анализира структурата и стилистичните особености на текста и ги сравнява с други текстове в базата си данни.
6. iThenticate е предназначен предимно за професионалисти и изследователи. Тази система използва контекстуален анализ за сравнение на текстове с огромна база данни от научни и академични източници, за да открие случаи на плагиатство.

Много от съвременните антиплагиат системи използват изкуствен интелект (ИИ) и машинно обучение, за да подобрят точността и ефективността на откриването на плагиатство.

Как се използва изкуствен интелект в антиплагиат системите?

- Обработка на естествен език (NLP). ИИ моделите анализират текстовете на дълбоко ниво, разбирайки граматиката, синтаксиса и семантиката. Това позволява откриването на парафразирани текстове и скрити форми на плагиатство.
- Машинно обучение. Моделите на машинно обучение се обучават върху големи набори от данни с примери за плагиатство и оригинално съдържание. Те се използват за разпознаване на сложни модели на плагиатство, които не могат да бъдат открити само чрез традиционни методи за съвпадение на текстове.
- Дълбоко обучение. Дълбоките невронни мрежи могат да анализират текстовете на ниво на абстракция, което надхвърля простите съвпадения на думи. Те могат да откриват смислови и контекстуални сходства между текстове.
- Семантично търсене. ИИ системите използват семантични търсения, които разпознават значението на думите и фразите, а не само точните съвпадения. Това помага за откриване на парафразирани изрази и синоними.

Примери за антиплагиат системи, използващи изкуствен интелект: Turnitin и Grammarly са сред най-популярните.; Copyleaks използва ИИ и машинно обучение за разпознаване на плагиатство в различни езици и формати. Системата предлага семантичен анализ и поддържа проверка на съдържание в реално време.; Unicheck използва технологии за машинно обучение и NLP за откриване на плагиатство в академични текстове и онлайн съдържание. Системата предлага интеграция с различни платформи за управление на обучение (LMS).; Quetext е онлайн инструмент за откриване на плагиатство, който използва ИИ и машинно обучение за анализ на текстовете.; Scribbr предлага инструмент за проверка на плагиатство, който използва технологията на Turnitin. Той е предназначен за студенти и академични автори.

Антиплагиат системите играят важна роля в подпомагането на авторите на текстове по различни начини, включително защита на интелектуалната собственост, подобряване на

качеството на текста и осигуряване на етично поведение. Ето някои конкретни начини, по които тези системи помагат на авторите:

1. Защита на интелектуалната собственост. Антиплагиат системите помагат на авторите да открият случаи, когато техните работи са били използвани без разрешение или надлежно признание. Тези системи предоставят документирани доказателства за плагиатство, които могат да бъдат използвани при правни спорове или при подаване на жалби.
2. Подобряване на качеството на текстовете. Антиплагиат системите помагат на авторите да открият неволно дублиране на текст или идеи, като по този начин им дават възможност да направят необходимите корекции преди публикуване. Като предоставят обратна връзка за сходствата с други текстове, системите насърчават авторите да създават по-оригинални и иновативни материали.
3. Осигуряване на етично поведение. Антиплагиат системите помагат на авторите да спазват етичните стандарти и правилата за цитиране в академични и професионални среди. Използването на тези системи насърчава авторите да се придържат към добри практики при създаването на текстове, като правилно цитиране и признаване на чуждия труд.
4. Оптимизиране на процеса на писане. Антиплагиат системите предлагат бързи и точни проверки, което позволява на авторите да оценят оригиналността на своите текстове без значителни забавяния. Много системи се интегрират с текстообработващи програми и платформи за управление на документи, улеснявайки процеса на проверка.
5. Подобряване на научните изследвания. Антиплагиат системите помагат на авторите да гарантират, че техните научни изследвания са оригинални и надеждни, като намаляват риска от публикуване на плагиатствани материали. При използване в процеса на рецензия, антиплагиат системите помагат на рецензентите да оценят оригиналността на изследванията, които преглеждат.
6. Подкрепа за преподаватели и студенти. За преподавателите и студентите, антиплагиат системите служат като образователен инструмент, който помага на студентите да научат правилата за цитиране и важността на оригиналността. За преподавателите, тези системи осигуряват справедлива и обективна оценка на

студентските работи, като гарантират, че всички студенти се оценяват по еднакви стандарти. Съществуват антиплагиат системи, които не само откриват плагиатство, но и предлагат редактиране на текстове. Тези системи често предоставят препоръки за подобряване на уникалността и качеството на текста. Ето някои от най-известните антиплагиат системи, които включват функции за редактиране:

- ✓ **Grammarly** - Открива плагиатство чрез сравнение с база данни от милиарди веб страници; Предлага граматически и стилистични препоръки; Подобрения в структурата на изреченията, правописа и пунктуацията. **Уебсайт:** [Grammarly](https://www.grammarly.com)
- ✓ **Turnitin** - Разпознава плагиатство чрез сравнение с академични и онлайн източници. Предлага инструменти за обратна връзка и оценка на текстове. Инструментът GradeMark включва възможности за коментиране и редактиране. **Уебсайт:** [Turnitin](https://www.turnitin.com)
- ✓ **ProWritingAid** - Открива плагиатство чрез сравнение с огромна база данни от академични и публични текстове.; Предлага подробни стилови и граматически предложения.; Анализ на структурата на текста, яснота и четивност. **Уебсайт:** [ProWritingAid](https://www.prowritingaid.com)
- ✓ **Quetex t-** Плагиатство проверка с помощта на DeepSearch™ технология.; Интегрирани инструменти за писане и редактиране, които предлагат подобрения в текста.; Изготвя подробни отчети за плагиатството и предоставя препоръки за редакция. **Уебсайт:** [Quetext](https://www.quetext.com)
- ✓ **Scribbr**- Проверка за плагиатство чрез сравнение с академични публикации и интернет източници.; Предлага услуги за редактиране от професионални редактори, които включват граматика, стил и структура на текста.; Персонализирани коментари и предложения за подобрения. **Уебсайт:** [Scribbr](https://www.scribbr.com)
- ✓ **Copyscape**- Открива плагиатство чрез сравнение с онлайн съдържание.; Не предлага директни инструменти за редактиране, но предоставя подробни доклади за намерените съвпадения, които помагат на потребителите да преработят текста си. **Уебсайт:** [Copyscape](https://www.copyscape.com)
- ✓ **Unicheck**-Проверка за плагиатство чрез сравнение с множество източници, включително интернет, библиотеки и архиви.; Интеграция с платформи за

обучение и текстови процесори за лесно редактиране.; Подробни отчети и препоръки за подобрене на текста. **Уебсайт:** [Unicheck](https://unicheck.com/)

- ✓ **PlagScan-** Проверка за плагиатство чрез сравнение с различни бази данни и интернет източници.; Възможности за редактиране на текста директно в системата, с предложения за подобрения.; Интеграция с учебни платформи и инструменти за преподаватели. **Уебсайт:** [PlagScan](https://plagscan.com/)

Тези системи не само откриват плагиатство, но също така предоставят ценни инструменти и препоръки за редактиране на текстове, което ги прави изключително полезни за академични и професионални цели.

Заклучение

Антиплагиат системите предоставят значими ползи за авторите на научни и др. текстове, като например защита на интелектуалната собственост, подобряване на качеството и оригиналността на текстовете, спазване на етични стандарти и оптимизиране на процеса на писане. Тези системи са ценен инструмент за всеки, който създава, публикува или оценява текстове.

Антиплагиат системите, които използват изкуствен интелект, предоставят по-висока точност и ефективност при откриването на индикатори за плагиатство. Те не само идентифицират буквални съвпадения, но и разпознават перефрази, синоними и контекстуални сходства, като по този начин предлагат по-цялостен подход към предотвратяване на неправомерно копиране на съдържание.

Цитирана литература:

<https://www.plagscan.com/en/>
<https://unicheck.com/>
<https://www.copyscape.com/>
<https://www.scribbr.com/>
<https://www.quetext.com/>
<https://prowritingaid.com/>
<https://www.turnitin.com/>
<https://www.grammarly.com/>