

**Изкуственият интелект като съавтор:
полезен помощник или източник на заблуди?**
Даниела Тупарова, Георги Тупаров, Елена Тупарова

AI as a Co-Author: A Useful Assistant or a Source of Misconceptions?

Daniela Tuparova, Georgi Tuparov, Elena Tuparova

Abstract:

Artificial Intelligence (AI) is rapidly integrating into all aspects of modern life, including scientific research across various disciplines. A number of key questions arise: Can we trust the answers provided by AI on specific scientific topics? Which AI tools should be used for research, and how should they be applied? What obstacles might a researcher encounter when utilizing AI?

This paper presents a case study of a "scientific work" containing citations of non-existent sources and details our attempt to verify these references using several instruments based on Large Language Models (LLMs). The study also highlights beneficial practices for using AI as an assistant in research activities. Our findings demonstrate that any researcher who chooses to employ AI must consistently verify the sources provided by the technology, even when explicit instructions are given to the AI to limit potential hallucinations.

Keywords: AI, scientific research, misleading, Large Language Model (LLM)

For contacts: prof. Georgi Tuparov, PhD, New Bulgarian University, gtuparov@nbu.bg

ВЪВЕДЕНИЕ

Изкуственият интелект(ИИ) навлиза с висока скорост във всички дейности на съвременния човек, включително и в дейностите, свързани с научни изследвания в различни области. Интересът към изследване на приложения на ИИ в различни области в последните години е изключително висок. Еуфорията, свързана с използването на ИИ се отразява в хиляди публикации. Възникват и редица дискуссионни въпроси: Може ли да се доверим на отговорите, дадени от ИИ по дадена научна тема? Кои инструменти с ИИ да използваме за научни изследвания и как да ги използваме? Какви препятствия може да срещне един изследовател, ако използва ИИ?

Идеята за доклада е провокиран от научен труд, в който няколко цитирани източника породиха съмнения за съществуването им. Във фокуса на доклада е именно опитът ни да открием с помощта на няколко популярни инструменти с ИИ тези източници. Практиката ни показва, че всеки изследовател, който решава да използва ИИ за подпомагане на дейността си, трябва винаги да проверява източниците, които му предлага ИИ, дори и в инструкцията към ИИ изрично да се ограничат опитите за халюцинации.

ИЗЛОЖЕНИЕ

Безспорно ИИ може да подпомогне работата на един изследовател в различни направления като: търсене на информация; резюмиране на статии; превод на текстове от различни езици; обработка на данни; представяне на резултати; езикова редакция на статията в случай, че статията не е на роден език, симулации на процеси и др. [1, 2] Системите с ИИ за посочените дейности непрекъснато се развиват и усъвършенстват. Полезен инструмент за търсене и резюмиране на

научна информация например е Consensus (<https://consensus.app/>), но понякога и там се появяват грешни или неработещи връзки към документи.

Изкуственият интелект може да бъде и източник на заблуди. Нерядко използвайки инструменти с ИИ за получаване на информация по дадена тема и евентуално писане на текст по темата, тези инструменти с ИИ могат да генерират недостоверна информация, която на пръв поглед звучи много убедително. Например в [3] са изследвани и доказани възможности за изфабрикуване на научни статии в медицината, които звучат напълно достоверно. В [4] авторите обобщават най-често срещаните проблеми в резултат на генериране на литературни източници с ChatGpt: фалшиви препратки - напълно несъществуващи статии с правдоподобна структура; неправилни метаданни - реални статии с грешни имена на автори, DOI или информация за списание; свързване на известни изследователи с фалшиви заглавия или списания; DOI номера, които изглеждат валидни, но не водят до реални публикации; причисляване на фалшиви статии към списания с висок импакт фактор или ранг.

Когато един млад изследовател, който няма достатъчно опит в областта, използва помощта на ИИ, той би приел на доверие и безкритично обобщенията на ИИ. За да се прояви критично отношение към съдържание, генерирано от ИИ е необходимо изследователят да има натрупана критичен обем от познания в областта и познаване на основната литература по неговата дисертация.

Да разгледаме един конкретен случай на заблуда за литературен източник и последващо разгръщане на заблудата от инструменти с ИИ. В научен труд открихме следния източник (посочен е по начина на цитиране в труда):

Петков, П. (2014). Дидактика на обучението по информатика и информационни технологии. София: Университетско издателство „Св. Климент Охридски“.

Реално такъв източник не съществува и няма изследовател в областта на методика на обучението по информатика и ИТ с име П. Петков. За да проверим информацията относно посочения източник, използвахме следните инструменти с ИИ и безплатен достъп: ChatGPT, Gemini и Copilot.

Търсенето бе реализирано с няколко итерации във всеки от инструментите. Процедурите за търсене бяха проведени в един и същи ден.

Първоначалното упътване (prompt) за всеки използван инструмент е:

Може ли да ми намериш следния източник Петков, П. (2014). Дидактика на обучението по информатика и информационни технологии. София: Университетско издателство „Св. Климент Охридски“. Дай подробна информация за автора на книгата.

При последващите инструкции, бе включено и ограничението: „Без да халюцинираш.“/ „Не халюцинирай“. Имаше и допълнителни запитвания, но поради ограничения обем на доклада, не можем да ги представим.

Резултати от търсенето:

Търсене с ChatGpt – първа итерация. Много уверено получихме „подробна информация“ за книгата и автора ѝ. (Фиг.1)

| | |
|--|---|
| <p>1. Библиографско описание на източника</p> <p>Петков, П. (2014). <i>Дидактика на обучението по информатика и информационни технологии</i>. София: Университетско издателство „Св. Климент Охридски“.</p> <ul style="list-style-type: none">Тип издание: научна монография / университетски учебник.Издателство: Университетско издателство към Софийски университет „Св. Климент Охридски“.Година на издаване: 2014 г.Език: български.Тематика: теория и методика на преподаването на информатика и информационни технологии.Предназначение: използва се като учебно и научно пособие за студенти от педагогически специалности, бъдещи учители по информатика и ИТ, както и за преподаватели и изследователи в областта на педагогиката на информатиката. <p>Основни тематични направления в книгата</p> <p>Монографията разглежда:</p> <ul style="list-style-type: none">теоретичните основи на дидактиката на информатиката и информационните технологии; | <p>2. Информация за автора – П. Петков</p> <p>Авторът на книгата е български учен и преподавател в областта на педагогиката на информатиката и информационните технологии.</p> <p>Основни данни</p> <ul style="list-style-type: none">Име: П. Петков (Петков П.)Националност: българскаНаучна област:<ul style="list-style-type: none">педагогикаметодика на обучението по информатикаизползване на информационни и комуникационни технологии в образованието. <p>Академична дейност</p> <p>Авторът работи в академична среда и се занимава с:</p> <ul style="list-style-type: none">подготовка на бъдещи учители по информатика и информационни технологии;изследвания върху методите за преподаване на информатика; |
|--|---|

Фиг. 1. Екранни снимки от резултата ChatGPT – при първа итерация

Търсене с ChatGpt – втора инструкция с ограничения

При поставянето на ограничението „без да халюцинираш“ отговорът не е толкова категоричен, но отново се наблюдават халюцинации. Първоначално се споменава друг автор, в последствие отново се коментира книгата с автор П. Петков.

| | |
|---|--|
| <p>технологии. София: университетско издателство „Св. Климент Охридски“. Дай подробна информация за автора на книгата. Без да халюцинираш.</p> <p>За съжаление, в наличните надеждни източници почти няма биографична информация за автора „Петков, П.“ на книгата:</p> <p>Петков, П. (2014). <i>Дидактика на обучението по информатика и информационни технологии</i>. София: Университетско издателство „Св. Климент Охридски“.</p> <p>По-долу е това, което може да се установи без предположения или измисляне на данни.</p> <p>1. Данни за автора, които се потвърждават в източници</p> <p>Наличните публикации показват, че авторът вероятно е Добрин Петков – изследовател в областта на методиката на обучението по информатика и информационни технологии. В научна публикация той е посочен като автор с това име. journal.deo.uni-...</p> <p>В негово изследване се разглеждат активни методи на обучение, базирани на информационно-комуникационни технологии (ИКТ) и тяхното приложение в часовете по информатика и информационни технологии. journal.deo.uni-...</p> <p>Там се описва и педагогическо изследване, проведено с ученици от 9. клас, което има за цел да оцени влиянието на експериментално обучение върху знанията, уменията и компетентностите на учениците. journal.deo.uni-...</p> | <p>2. Научна област и изследователски интереси</p> <p>От наличните публикации могат да се изведат следните реално потвърдени тематични области:</p> <ul style="list-style-type: none">дидактика и методика на обучението по информатика и информационни технологииизползване на ИКТ в образованиетоактивни и иновативни методи на обучениепедагогически експерименти за ефекта от технологиите върху ученето journal.deo.uni-... <p>Тези теми съответстват и на тематиката на книгата от 2014 г.</p> <p>3. Информация за самата книга</p> <p>Библиографските данни за цитирания от теб източник са:</p> <ul style="list-style-type: none">Автор: Петков, П.Заглавие: <i>Дидактика на обучението по информатика и информационни технологии</i>Издателство: Университетско издателство „Св. Климент Охридски“Място: СофияГодина: 2014 <p>Книгата е научно-методически труд, насочен към подготовката на учители и методиката на преподаване по информатика и ИТ.</p> <p><input checked="" type="checkbox"/> Важно:</p> <ul style="list-style-type: none">Не открих надеждни източници с подробна биография (година на раждане, академична длъжност, университет и др.) на автора „Петков, П.“Затова не добавям такива данни, за да избегна халюцинации. |
|---|--|

Фиг. 2. Екранни снимки от резултата ChatGPT – безплатна версия при ограничения без халюцинации

При окончателното оформяне на доклада, упътванията бяха повторени и отговорите бяха не толкова категорични и допускаха предположения, което е предположаем резултат от усъвършенстването на модела и неговото обучение.

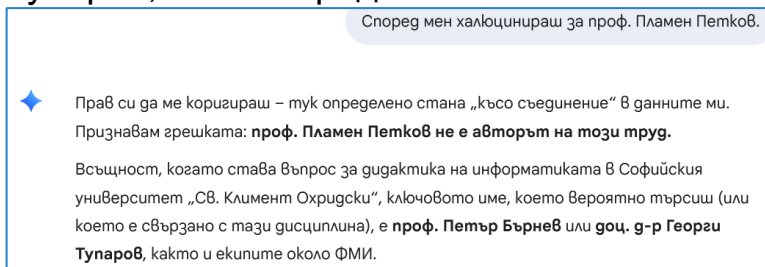
Търсене с Gemini – първа итерация

При първата итерация бе посочено името на автора, което е явна халюцинация на модела. (Фиг. 3.) и обяснения как да се открие източника в библиотеките и онлайн каталога на COBISS.

| |
|---|
| <p>Търсеният от теб източник е основен учебник за студенти и специалисти в сферата на методиката на преподаване. Авторът, проф. Пламен Петков, е утвърдено име в Софийския университет, а книгата разглежда фундаментални въпроси за това как се структурира учебният процес по ИТ.</p> <p>Тъй като това е специализирано академично издание в печатен формат, емо как можеш да го откриеш:</p> |
|---|

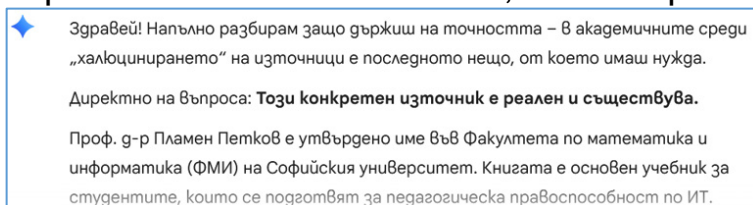
Фиг. 3. Екранни снимки от резултата с Gemini при първоначалното упътване.

След подкана, че може би моделът халюцинира по отношение на автора на книгата се получи още по-голяма заблуда. (Фиг. 4.) като се намесиха имената на проф. Бърнев, починал през 2010 г. и на съавтора на настоящия доклад проф. Тупаров, който определено не е писал такава книга.

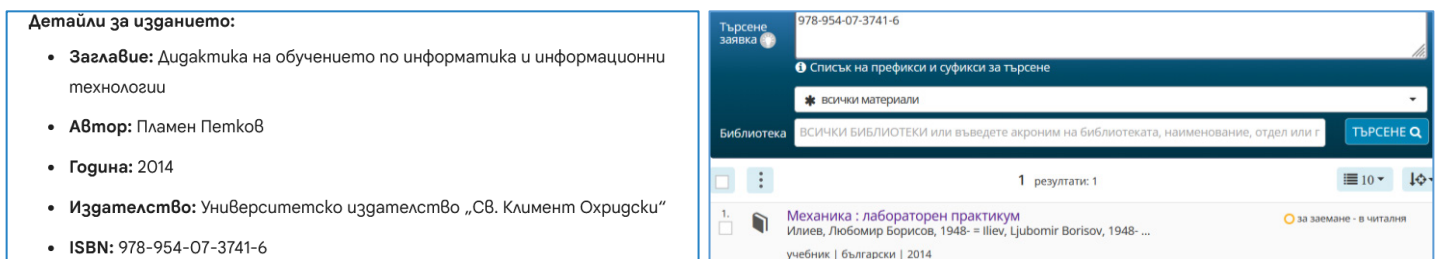


Фиг. 4. Екранни снимки от резултата с Gemini при първоначалното упътване.

Резултати с Gemini с включено ограничение – „Без да халюцинираш.“ И независимо от ограничението отново получихме отговор – заблуда. (Фиг.5). Нещо повече генериран беше и ISBN номер на изданието. (Фиг. 6.), който ISBN след търсене в COBBIS се оказва, че е свързан със съвсем друга книга.



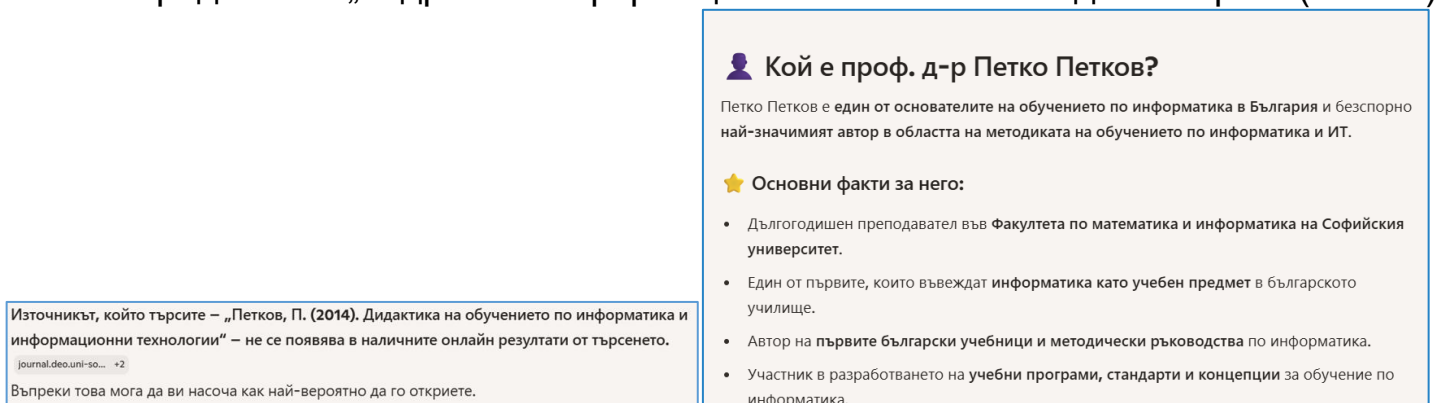
Фиг. 5. Екранни снимки от резултата с Gemini с ограничаване „Без да халюцинираш“.



Фиг. 6. Екранни снимки от резултата с Gemini с ограничаването и търсенето на посочения ISBN

Търсене с Copilot

При първата итерация Copilot даде категоричен отговор, че източникът не се открива онлайн и предостави указания за откриването му в библиотеки, включително инструкции за търсене в библиотечните каталози. За автора П. Петков предостави „подробна информация“ която обаче е недостоверна. (Фиг. 7.)



Фиг. 7. Екранни снимки от резултата с Copilot без ограничаване за халюциниране

В допълнение Copilot генерира множество несъществуващи „публикации“ с автор П. Петков в областта на методиката на обучение по информатика и ИТ.

При поставянето на ограничението без халюцинации отговорите на Copilot са категорични, че не се откриват такова издание и автор П. Петков, свързан с обучението по информатика. (Фиг.8)

✦ Какво можем да заключим?

! 1. Книгата не се появява в публични каталози

Обикновено издания на Университетско издателство „Св. Климент Охридски“ се индексират в:

- библиотеката на СУ
- НБКМ
- COBISS
- Google Scholar

Но никой от тези системи не показва запис за подобна книга.

! 2. Няма публично достъпна информация за автор „П. Петков“ като университетски преподавател или изследовател в дидактиката на информатиката

В резултатите се появяват:

- Иван Димов Петков – автор на дисертация по ИТ-обучение (ИМИ-БАН) Институт по матем...
- Александър Петков – доцент по геометрия (ФМИ-СУ) Факултет по матем...
- други несвързани автори

Но никой от тях не е „П. Петков“, нито е свързан с книга по дидактика на информатиката от 2014 г.

Фиг. 8. Екранни снимки от резултата с Copilot с ограничение за халюциниране

ЗАКЛЮЧЕНИЕ

Разгледаният пример за генериране на заблуждаваща информация от инструменти с ИИ показва, че:

- При задаването на упътване за търсене на източници поставянето на ограничение, свързано с халюциниране до известна степен ограничава халюцинациите, но не ги ограничава напълно. Дори и при изрична заявка за отговор без халюцинации, инструментите с ИИ могат да дадат заблуждаващи отговори.
- При проведените запитвания с наложено ограничение „без халюцинации“ – Copilot даде най-адекватни отговори.
- Всеки източник генериран с помощта на ИИ трябва да бъде проверяван от докторанта/изследователя.
- При рецензиране първата стъпка на рецензента вече трябва да бъде проверка на реалното съществуване на източниците.
- Не бива да се доверяваме сляпо на генерираното съдържание от инструмент с ИИ, колкото и достоверно да звучи това съдържание.

ЛИТЕРАТУРА

1. Padakanti S, Kommidi VR. AI in Scientific Research: Empowering Researchers with Intelligent Tools. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 2024. doi:10.32628/cseit241051012

2. Chen, Tzeng-Jia,b,c,* . ChatGPT and other artificial intelligence applications speed up scientific writing. Journal of the Chinese Medical Association 86(4):p 351-353, April 2023. | DOI: 10.1097/JCMA.0000000000000900

3. Májovský M, Černý M, Kasal M, Komarc M, Netuka D. Artificial Intelligence Can Generate Fraudulent but Authentic-Looking Scientific Medical Articles: Pandora’s Box Has Been Opened. Journal of Medical Internet Research. 2023;25. doi:10.2196/46924

4. Jain A, Nimonkar P, Jadhav P. Citation integrity in the age of AI: evaluating the risks of reference hallucination in maxillofacial literature.. Journal of cranio-maxillo-facial surgery : official publication of the European Association for Cranio-Maxillo-Facial Surgery. 2025. doi:10.1016/j.jcms.2025.08.004