

Комуникация в медиите и обществото
Communication in Media and Society

**Фактчекинг на български език
с изкуствен интелект – мисията (не)възможна?**

DOI 10.55206/WHCO8050

Тихомира Дончева

Софийски университет „Св. Климент Охридски“

Имейл: tihomiradd@uni-sofia.bg

Абстракт: Фактчекингът е сред способите за преодоляване на информационното разстройство. В контекста на навлизането на изкуствения интелект – в частност на големите езикови модели – в информационното пространство нуждата от надеждна и достоверна информация се увеличава, а достъпът става все по-труден. Генеративният изкуствен интелект може да подпомага, но и да затруднява процеса по верифициране. Текущото изследване цели да провери до каква степен два от най-разпространените чатбота, Claude.ai и ChatGPT, проверяват успешно твърдения на български език. За проведения експеримент е използвано безплатно умение, създадено от Наука.бг. Проверяваните твърдения са селектирани от страницата „Проверено“ на най-големия сайт за проверка на факти в България Factcheck.bg. Чрез сравнителен контент анализ е установено, че в повечето случаи двата модела коректно определят твърденията за неверни (спрямо данните от Factcheck.bg). Claude.ai проявява по-авторитетен тон и кратко и точно дава детайли защо класифицира твърдението по съответния начин. ChatGPT говори по-разколебано, но дава по-голям детайл, точност и яснота на обясненията си. Основен проблем е, че ChatGPT рядко осигурява връзка към източниците си, а Claude.ai понякога използва ненадеждни източници. Положителна тенденция е честият призив на Claude.ai да се проверява информацията в официалните сайтове на верификатори на факти.

Основното заключение потвърждава хипотезата, че изкуственият интелект е асистент на фактчекърите, но не и заместител.

Ключови думи: дезинформация, проверка на факти, фактчекинг, изкуствен интелект, големи езикови модели, ChatGPT, Claude.ai.

AI Factchecking in the Bulgarian Language – Mission (Im)Possible?

Tihomira Doncheva

Sofia University “St Kliment Ohridski”

E-mail: tihomiradd@uni-sofia.bg

Abstract: Factchecking is among the key mechanisms for addressing information disorder. In the context of the growing integration of artificial intelligence, particularly large language models, into the information environment, the need for reliable information is increasing while access to it is more challenging. Thus, generative artificial intelligence can be both – an asset and a risk to factcheckers. The present study aims to examine the extent to which Claude.ai and ChatGPT can factcheck claims in Bulgarian. The experiment was conducted using a free skill developed by Nauka.bg. The claims were selected from the “Verified” section of Factcheck.bg, the largest fact-checking platform in Bulgaria. Through comparative content analysis, the study found that in most cases both models correctly identify false claims in accordance with Factcheck.bg’s publications. Claude.ai demonstrates a more authoritative tone and provides concise explanations supporting its verdicts. ChatGPT is more hesitant but offers greater detail, precision, and clarity in its explanations. Notably, ChatGPT rarely references its sources, whereas Claude.ai occasionally links to unreliable sources. A positive trend is Claude.ai’s note to consult official factchecking organizations and verified information sources. The results confirm the hypothesis that artificial intelligence can assist factcheckers but cannot replace human expertise in the verification process.

Keywords: disinformation, factchecking, artificial intelligence, large language models, ChatGPT, Claude.ai.

Увод

Масовото разпространение на изкуствения интелект (ИИ) води до рязкото увеличаване на количеството, както и на в пъти по-бързото и широко разпространение на ИИ-генерирано съдържание. Това води до промяна в журналистическите стандарти и се отразява върху демократичните норми и ценности. [1] Множество разработки и изследвания разглеждат плюсовете и минусите от навлизането на ИИ в медийната среда. [2] Разпространението на невярна информация (мал-, мис и дезинформация) е силно повлияно от ИИ, най-вече в обхвата и мащаба на дезинформацията. [3] Само за месец декември 2025 г., от EDMO отчитат увеличение от 16% в разпространението на ИИ-генерирано или манипулирано съдържание. [4] Генеративните технологии като езикови модели и инструменти за създаване на изображения и видео спомагат разпространението на автоматично генерирани текстове, които могат да бъдат представяни като новинарски статии;

т.нар. deepfake видеа, при които изкуствено създадени образи и гласове се представят като реални публични личности; както и манипулирани изображения, които се разпространяват с цел подвеждане или внушаване. [5] Последствията са свързани с все по-трудното разпознаване на реалното [6], а вече има и достатъчно сведения за нарушаване на демократичните принципи и изконни човешки права заради разпространението на ИИ-генерирана дезинформация. [7] [8]

В България темата се разглежда най-често от законодателна гледна точка, като има и изследвания, които се концентрират върху теоретичен обзор, сравнителен анализ и дори разработването на инструментариуми. [9] [10] [11] [12] [13] [14] Причините защо България е особено застрашена от разпространението на ИИ-генерирана дезинформация са много, като например ниски нива на медийна свобода, незадоволителните резултати в образованието и грамотността и критично ниското доверие между хората в обществото. [15]

Текущата статия адресира липсата на изследване, което експериментално да докаже до каква степен ИИ може да „улавя“ невярно съдържание на български език. Целта е да се покаже в каква степен генеративният изкуствен интелект (в частност големите езикови модели Claude.ai и ChatGPT) успешно засичат невярно съдържание. Проведен е експеримент, при който двата чатбота са попитани за истинността на 10 твърдения всеки. Всички 10 твърдения са проверени от Factcheck.bg и обявени за неверни.

Хипотезата на изследването е, че ИИ може да се използва като проверител на факти в начален и базов етап, като след това задължително резултатът трябва да се прегледа от човек. Причината е, че ИИ може да проявява халюцинации, както и да прави грешки. Ограничени са възможностите и за разбиране на саркастично съдържание. В изследването са зададени следните изследователски въпроси:

- Как генеративният изкуствен интелект проверя истинността на твърдения, ако се използва умение за фактчекинг на български език?
- Има ли съществени разлики между двата големи езикови модела?
- Наблюдава ли се консистентност в отговорите на двата модела?

Методология

Използваната методология представлява квазиекспериментално сравнително изследване. Прилага се качествен контент анализ и сравнителен анализ на текстовите отговори на 10 идентично зададени промпта към два езикови модела – ChatGPT и Claude. В два от случаите и към двата езикови модела е зададен втори последващ промпт, с който се цели конкретизиране на зададения въпрос. В единия случай за промпта се дава конкретика за от-

насяне на твърдението към конкретна година („а истина ли е било твърдението през 2021 г.?"). Във втория случай – и Claude, и ChatGPT искат скрийншот от Facebook пост, тъй като нито един от моделите не може да отвори линк към пост от социалната мрежа. За контент анализа е подготвена аналитична карта с въпроси, на базата на които се извършва сравнителен анализ между отделните отговори (общо 20 на брой). Прилага се дедуктивен подход, при който предварително определени категории за анализ (точност, нюанс, източници, яснота, риск от подвеждане, коментари) се отнасят към отговорите и на двата модела. По този начин се тества зададената хипотеза и се отговаря на зададените изследователски въпроси. Категориите, спрямо които се изследват двата модела, са:

- точност на върдикта – дали моделът достига до същия или близък извод като Factcheck.bg;
- нюанс – дали моделът разграничава между напълно невярно, подвеждащо, смесено твърдение или контекстуална истина;
- източници – използва ли моделът проверими, релевантни и надеждни източници.
- яснота – структурираност, разбираемост и четливост на отговора;
- риск от подвеждане – съществува ли риск читателят да остане с погрешно впечатление след проверката.

За всяка категория изследователят дава оценка – ниска степен, средна степен, висока степен, като оценката се базира на контент анализа и сравнителен анализ спрямо публикациите на Factcheck.bg.

Експериментът е проведен в един ден – 22.05.2026 г., като са използвани безплатните версии на двата големи езикови модели, но изследователят се е вписал с имейл адрес и в двата. Твърденията са зададени едно след друго в един и същ разговор (до достигане на лимита за въпроси към езиковите модели, зададен от безплатния план).

Използван е безплатният инструмент за умение (factcheck skill), създаден от Наука.бг. [16] Това е инструмент за ИИ с отворен код, който работи в Claude.ai и в ChatGPT. Умението е инсталирано спрямо насоките в сайта, а промптовете са зададени спрямо изискването от създателите: „истина ли е, че“ или „вярно ли е, че“. Според създателите на умението за Claude, след подаването на текст, линк, скрийншот или социална публикация като промпт, генеративният модел Claude минава през 11 стъпки на анализ. Авторите на умението твърдят, че съществуват четири режима на работа, като един от тях дава обстойна пътна карта, а друг предлага кратка проверка при въпроси тип „истина ли е, че“. Не се съобщават подробни спецификации за умението в ChatGPT, но се препоръчва Claude.ai за по-надеждни резултати. За да се осигури база за сравнителен анализ между резултатите на двата модела, всички промптове са зададени по начина „истина ли

е, че” или „вярно ли е, че”. За изследването са избрани 10 поста с проверка на факти от Factcheck.bg, като пет от тях са последните качени на сайта в страницата „Проверено“ (в периода 15–20 май 2026 г.), а останалите пет представляват най-ранните постове в страницата „Проверено“ (в периода 30.05–09.06 2021 г.). Селекцията на проверените твърдения не е случайна – целта е да се разгледа дали има разлика в отговорите за по-отдавнашни и по-скорошни събития.

III и фактчекингът

В статията под III се разбира генеративен III и големите езикови модели. [17] Проверката на факти или фактчекингът представлява журналистическа практика и е сред най-ефективните способи за противодействие на дезинформацията. Съществуват около 450 фактчекърски организации в световен мащаб. [18] В България формирането на практиката за проверка на факти преминава през няколко етапа, като платформата Factcheck.bg е сред най-значимите инициативи. Създадена през 2021 г. от Асоциацията на европейските журналисти – България, това е първата българска платформа, тясно специализирана във фактчекинга. Основните ѝ цели са проверка на публични твърдения и новини, анализ на съдържание в социалните мрежи, разобличаване на дезинформационни кампании, подпомагане на журналистите и гражданите при разграничаването на факти от манипулации. [18]

Ефектите от III в разпространението на дезинформацията са обстойно проучени в световен мащаб. В следващите редове се предлага кратка категоризация на основните ползи и рискове.

Безспорна полза е способността на III, че може да обработва огромно количество съдържание в реално време, като идентифицира неверни твърдения и ги съпоставя с проверени източници. [19] Затова този тип автоматизация има съществена роля при наблюдението на социалните мрежи и филтрирането на големи обеми информация. Целта е своевременното откриване и маркиране на подвеждащи и манипулативни наративи. Второ, езиковите модели могат да анализират твърдения, да ги сравняват със съществуващи бази данни и да идентифицират несъответствия с висока степен на прецизност и по този начин да повишават точността на верификацията. Доказана е точност над 90% при идентифицирането на неверни твърдения от модели с III. [20] Тази точност се аргументира с елиминирането на евентуално човешко пристрастие и възможна субективна преценка. Друго изследване например доказва, че големите езикови модели са особено ефективни при опровергаването на фалшиви наративи чрез анализ на езикови маркери и контекстуални несъответствия. [21] Освен точност и бързина, фактчекингът с помощта на III позволява и работа в мащаб, който е в пъти по-голям от ръчния труд, извършван от проверителите на факти. Откри-

ването на дезинформационно съдържание едновременно в различни платформи и формати (напр. текст, изображения, видео) [22] е полезно във времето на „информационното претоварване“ [23], когато дезинформацията се разпространява в пъти по-бързо и по-широко от верифицираната информация. [24] Развитието в машинното генериране на визуализации (изображения и видео) успява да затрудни човешкото око и изследвания твърдят, че решения, използващи ИИ, могат да са от помощ за проверителите на факти. [25] Последните две категории приложения на ИИ в процеса на проверка на факти са свързани с бизнес модела на организациите за фактчекинг. От една страна, този тип организации са изградени върху принципите за прозрачност и отчетност и употребата на ИИ трябва да бъде точно и ясно изложена. [26] През 2026 г. Factcheck.bg например приемат Кодекс за употреба на изкуствен интелект в редакцията, който цели регулация на инструментализирането му в ежедневната журналистическа работа. [27] ИИ може да доведе и до ограничаване на разходите на редакциите и преразпределяне на задачите и ролите по-ефективно. [28]

Изброените причини обясняват защо девет от 10 организации в седем държави използват ИИ на ежедневна база. Това се случва във всеки етап на проверяването на факти, но в различна степен: задачи преди фактчекинга (напр. ранно известяване, мониторинг), по време на проверяването (напр. използването на бази данни за крос-проверки) и задачи след проверката (напр. приложения за мониторинг и известяване, комуникация на разработките). [29]

Внедряването на ИИ в проверката на факти обаче крие и немалко рискове. Често плюсовете и минусите са тясно свързани. Въпреки че ИИ се справя с улавянето на deepfakes, то съществуват и ограничения, свързани с непрекъснатото подобряване на тези технологии. [виж 25] Проблематично за големите езикови модели остава контекстуалното разбиране, вкл. разграничаването на иронизиращ или саркастичен език. [виж 22] Това е и сред основните причини за препоръката на редица изследователи ИИ да се използва под ръководството на човек, който да държи отговорността публикацията да следва общоприетите журналистически етични принципи. [30] [31] Второ, тъй като големите езикови модели използват съществуващи бази от данни, за да се обучават непрекъснато, то съществува рискът от пристрастия, особено при по-тясно разпространени езици. [виж 26] Трето, използването на ИИ за фактчекърство повдига и немалко етични въпроси, сред които зависимостта на повечето модели от трети страни (най-често големи софтуерни компании) и непрозрачността на повечето модели за процесите на преработване на информацията. [виж 19] Сред потенциалните проблеми за пълното внедряване на ИИ в практиката на фактчекърството е инвестицията, която е необходима, за да се използват ефективно тези ин-

струменти, вкл. за абонаменти и обучение. [виж 22] Доверието в ИИ от широката публика също е част от предизвикателствата, които се разглеждат от изследователи. [виж 19] [32]

В обобщение ИИ може да е добър инструмент, който се използва ефективно от проверителите на факти в различни етапи от работата, като например редакция на публикации, анализ на тенденции в разследванията и повишаване на информационната грамотност в инициативи. [33]

Резултати

В контент анализа на резултатите от експеримента се разглежда и начинът, по който езиковите модели се изразяват, като комуникацията се извършва на български език. [34] Claude.ai и ChatGPT се справят задоволително с анализа на доказано неверните твърдения, за които са попитани по време на експеримента. Липсват явни грешки, като и двата модела показват силни и слаби черти при различни промптове (Таблица 1). Важно е да се отбележи, че описаният от Наука.бг 11-степенен анализ при Claude.ai не се появи в отговорите на промптовете. При някои от случаите Claude отговаря с две-три изречения, докато в други случаи се включват отделни секции от описанията от Наука.бг стъпки, например калибриране на увереността, но в нито един от изследваните случаи не беше наблюдаван отговор в 11 стъпки, както е показано в статията, представяща безплатното умение. Това може да се дължи на начина, по който е зададен промптът (в текущия случай „истина ли е, че“, а в дадения пример се пита директно въпросът „Има ли посъпване...“). [виж 16]

При първия промпт двата езикови модела показват висока точност при определянето му като невярно твърдение (върдикт), но при повторно питане дали твърдението е вярно за 2021 г., Claude не отговаря толкова утвърдително, като заявява, че твърдението е подвеждащо, но базиращо се на реална основа. Големият езиков модел дава детайли за твърдението и обяснява достатъчно нюансирано и контекстуализирано защо то е невярно дори през 2021 г. Затова и отговорът на втория промпт е неточен, тъй като „базиращо се на реална основа“ за модела е свързано с облекчения режим за ваксинираните в ЕС през туристическия сезон, какъвто облекчен режим обаче е бил валиден за всички притежатели на зелен сертификат за COVID-19 (каквата информация Claude дава). Въпреки това обаче за един читател, който не вниква в данните, подадени от Claude, а само прочита върдикта, написаното може да се разбере погрешно. И при двата модела се дават линкове към официални източници, а отговорът на Claude завършва с призив да се посетят сайтовете на официални проверители на факти при съмнение. Проблем е, че при копиране на текста линковете към източниците не се копират и това може да доведе до ерозия на източника.

Положително за ChatGPT е, че дава насоки за „надеждни източници“, като в конкретния случай изброява Европейската комисия и други европейски институции. И двата отговора са добре структурирани, като се дава контекст, както и кратко обобщение на написаното. Крайният риск от подвеждане е нисък, като при Claude се наблюдава известен среден риск от подвеждане поради върдикта на втория промпт.

Отговорът на ChatGPT за втория промпт е много по-комплексен и подробен. Макар и в двата езикови модела да се дава оценка, че твърдението е „подвеждащо“, Claude.ai накратко опровергава твърдението, като се позовава на международни проверители на факти, както и на FDA, американската агенция за храни и лекарства. ChatGPT обаче разделя твърдението на три части, като проверява всяка отделно (на третата част от проверяваното твърдение върдикта е „mixed/смесено“, тъй като според модела има елементи на истина в това, че американската агенция е одобрила по ускорен път ваксините срещу COVID). Това разделяне, от една страна, е добра практика, тъй като се открояват различните методи за манипулиране на съдържанието, но от друга – последният върдикт добавя доза съмнение към цялостната оценка на ChatGPT за истинността на твърдението. Това нарушава яснотата на обяснението, но въпреки това ChatGPT обобщава с добре структурирани примери за „червени флагове“, които да подскажат на читателя, че съдържанието е с манипулативен характер. Изборът на източници на ChatGPT също е добър, като са засегнати най-големите проверители на факти като PolitiFact и AFP. Интересен е един от линковете, даден от Claude.ai, който е добавен към текст, опровергаващ твърдението, който обаче води до сайт, който публикува антиваксърски новини и е разследван от проверители на факти. Причината за референцията най-вероятно е поради статията, публикувана в конкретния сайт, където се включват няколко проверки от фактчекъри с цел, обаче, да се опровергае проверяването на твърдението. Затова за втория промпт и ChatGPT, и Claude.ai не са абсолютно убедителни в доказването на тезата си, че твърдението, за което са питани, е подвеждащо.

Третият промпт е за инцидент, при който по време на учение на територията на България американски войници нахлуват в имот, съседен на базата за тренировки, където има работници. И двата езикови модела правилно отразяват, че промптът е частично верен, тъй като ситуацията е реална, но твърдението е представено чрез силно емоционален език. Отново ChatGPT представя аргументи защо класифицира твърдението като „смесено“ в по-подробен план от Claude.ai, а и двата езикови модела предоставят източници като познати български медии. ChatGPT добавя и линк към сайт за американски военни новини, както и към Seattle Times, ежедневен вестник от Сиатъл, щата Вашингтон. Интересно е защо се дават именно те-

зи два сайта, които не са сред най-известните международни медии (които са отразили случая). Това повдига въпроса по какъв начин се набавят източниците и дали по този начин ненадежден източник може да се използва като доказателство от голям езиков модел (както във втория пример).

Четвъртият промпт е правилно опроверган от двата езикови модела, които отбелязват, че цитираната технология е истинска, но проверяваното твърдение е свързано с често срещана конспиративна теория. Claude отново накратко контекстуализира темата, като реферира добър набор от източници – както международни медии, така и сайтове с научни разработки, включително и специализирани медицински директории. ChatGPT, който макар че отново разгръща по-подробно темата и дава различни нива на анализ и оценка на истинността на раздробеното на малки части твърдение, не дава линк и не реферира нито един източник. Това е особено тревожно, тъй като застрашава доверието в иначе добрия анализ. Положително за Claude е призивът читателят да се обръща към официални източници и доказани проверители на факти, както и призив да не се черпи информация от места като YouTube.

Последният промпт от проверените факти от 2021 г. се отнася за твърдение, че авиокомпаниите обсъждат опасност от съсиреци при ваксинирани срещу COVID. И двата модела отново коректно опровергават твърдението, като дават достъп до надеждни и официални източници. Claude включва линк към YouTube канал на местна новинарска медия за Сан Диего, Калифорния, което е интересно и потенциално проблемно, тъй като е свързано с доверието в източниците и съответно доверието в проверката на езиковия модел. ChatGPT пък от своя страна макар и подробно да опровергава твърдението, включва една последна секция „Ограничения“, в която пояснява, че може да съществуват някои частни случаи, при които да има ограничение при летене. Въпреки че това не се противопоставя на цялостния вердикт, самото позициониране на секцията накрая не е оптимално, тъй като оставя читателя с нюансирано усещане.

За първия промпт от последните проверени истории към датата на анализ (май 2026 г.) и двата езикови модела поискаха скрийншот на публикацията, тъй като не успяват да достигнат до Facebook поста през поставен линк. Тук Claude се справя много по-добре от ChatGPT, като дава правилна оценка, води до доверени източници и дава добър контекст на твърдението и използваната подвеждаща техника. ChatGPT обаче макар правилно да оценява твърдението като „невярно“, не предоставя източници, както и повтаря няколко пъти, че няма данни за описаното в твърдението и че не може да открие такива. Не споменава обаче, че твърдението е проверено от водещи фактчекърски организации, както Claude прави. Рискът тук е, че ChatGPT може да затвърди техника, която често се използва при разпрост-

ранението на дезинформация, а именно доказването без доказване (woozle effect).

Относно седмия промпт, Claude не се справя утвърдително. Макар че правилно отбелязва, че твърдението е невярно, обяснението не е твърде ясно, като двата засегнати закона се смесват. Нюансът е добре отразен – твърдението смесва реални закони с невярна информация, но реферираните източници не са надеждни, а един от дадените източници е фриндж сайт, скептичен към общоприети теми като климатичните промени например. ChatGPT също правилно определя, че твърдението е невярно, въпреки че в доказателствената си част включва и оценка за „преувеличен сценарий“, каквото явление не се наблюдава в конкретния случай, а и не съществува такова понятие в изследването на дезинформацията. Езиковият модел обаче не реферира източници и отново е включена последната секция „Ограничения“, където се допуска възможността за единични случаи, за които твърдението да е вярно. Макар че моделът утвърждава, че няма масов феномен, проверяването на факта не е достатъчно убедително.

Относно осмия промпт, където пост в социалните мрежи твърди, че се показва извеждането на заразен с ханта вирус, последван от оператор без защитен костюм, двата езикови модела оценяват скрийншота като подвеждащо съдържание. И в двата случая се дават обяснения за различни практики, при които операторите стоят извън зоната на опасност от заразяване и затова не следват строги протоколи. Дават се и линкове към официални източници. И Claude, и ChatGPT обаче не улавят, че снимката е преизползвана от учение на испанските власти за справяне с пристигането на случай на болен с Ебола и е от 2025 г., като съответно не показва заболял от ханта вирус. Представената информация е напълно невярна.

Скриншотът на Facebook пост, който представя Тръмп като единствен защитник на България, е обявен за дезинформация, като Claude дори започва своя отговор със следното: „Преди да направя пълна проверка, трябва да споделя нещо важно директно. Не е нужна задълбочена уеб проверка, за да оценя този пост – той съдържа толкова плътна концентрация от добре документирани конспиративни наративи, че самата му структура е диагноза“. Макар правилно да определя твърдението за невярно и да дава примери за основните „конспиративни рамки“, езиковият модел не дава контекст за самите твърдения, а само отбелязва, че това са документирани наративи от EUvsDisinfo (не е посочен линк). Въпреки това обаче е включен отново призивът да се проверява информация в официални фактчекърски сайтове. Макар ChatGPT да казва това, което и Claude твърди, при този езиков модел е дадено обяснение защо включените конспиративни твърдения са непроверяеми. Липсват обаче каквито и да е източници. Рискът с този тип проверки на факти е евентуалното ниско доверие на читателя, тъй като

и двата модела обявяват твърденията за конспиративни и неверни, но не реферират казаното и не водят до достъп към достоверна информация.

Последното проверявано твърдение се отнася за предложение на канадски лекар да се узакони евтаназията на нелечимо болни кърмачета, което обаче е извадено извън контекст и е представено като приет закон, и то за новородени от социално слаби семейства. Claude.ai определя истинността на твърдението като „смесено, но предимно подвеждащо“, тъй като отбелязва, че е имало такова твърдение. Цитираните източници обаче са медии с доказани пристрастия, което проблематизира достоверността на анализа. В общ план ChatGPT дава повече контекстуални детайли, които обясняват защо постът е „силно подвеждащ. Големият минус отново обаче е липсата на източници и линкове към официални страници (въпреки че ChatGPT твърди, че информацията е лесно проверяема в официални източници).

Таблица 1. Качествен сравнителен анализ на отговорите от Claude.ai и ChatGPT относно 10 твърдения, проверени от Factcheck.bg и определени за не-верни

№	Модел	Точност на върдикта	Нюанс	Източници	Яснота	Риск от подвеждане
1 Истина ли е, че само ваксинираните за ковид могат да пътуват свободно в ЕС?	Claude	Висока за 1-ви промпт Средна за 2-ри промпт	Висока	Висока	Висока	Нисък до среден
1.1 Истина ли е било твърдението през 2021 г.?	ChatGPT	Висока	Висока	Висока	Висока	Нисък
2 Истина ли е, че лекари, свидетелстващи в сенатската комисия по държавните дела в САЩ за експерименталните ваксини, одобрени по спешност...	Claude	Висока	Средна	Средна до висока	Висока	Среден
	ChatGPT	Висока	Средна	Висока	Средна до висока	Нисък до среден
3 Вярно ли е, че в Чешнегирово нахлу една част от окупационната армия в България – 7 американски войника с тежко въоръжение, заплашиха живота и здравето на хората и според мен това е престъпление.	Claude	Висока	Висока	Средна до висока	Висока	Нисък
	ChatGPT	Висока	Висок	Средна до висока	Висока	Нисък
4 Вярно ли е, че ето как постепенно от човеци ще ни превърнат на Androidi...	Claude	Висока	Висока	Висока	Висока	Нисък
	ChatGPT	Висока	Средна	Ниска	Висока	Нисък до среден
5 Вярно ли е, че в този момент автокомпаниите съответно ваксинираните си клиенти за високия риск те да получат кръвни съсиреци от експерименталния mRNA серум...	Claude	Висока	Висока	Висока	Висока	Нисък
	ChatGPT	Висока	Средна	Висока	Висока	Нисък
6 Истина ли е този фейсбук пост: https://www.facebook.com/photo/?fbid=2183662649063205&set=gm.1950930098868235&idovanity=803832120244711	Claude	Висока	Средна	Висока	Висока	Нисък
6.1 Изпращане на скрийншот.						
[Скрийншотът твърди, че прикачените снимки показват тормоз над българи в Украйна, а в текста към поста се твърди, че българският език е забранен]	ChatGPT	Средна до висока	Средна	Ниска	Ниска до средна	Среден до висок
7 Вярно ли е, че и ето какво правят германските мъже – отиват в общината, заявяват, че вече са жена, получават нова лична карта, в която пише, че вече са жена, след което няма нужда да искат разрешение от германската армия, за да напуснат Германия за дълъг период.	Claude	Средна до висока	Висока	Ниска до средна	Средна	Нисък до среден
	ChatGPT	Средна до висока	Висока	Ниска	Средна	Нисък до среден
8 Истина ли е този пост от фейсбук (изпращам ти скрийншот)	Claude	Средна	Ниска	Средна до висока	Ниска до средна	Среден до висок
[На скрийншота е показана снимка на оператор без защитно облекло, който следва евакуацията на болен и защитено облекло. Твърди се, че болят е заразен с Ханта вирус и че ситуацията е пресекспонирана.]	ChatGPT	Средна	Ниска	Средна до висока	Ниска до средна	Среден до висок
9 Истина ли е, че (изпратен скрийншот на фейсбук пост) [Скрийншот на Facebook пост представя Тръм като единствен закрилник на България и се твърди, че е наложена забрана за внос от Великобритания.]	Claude	Висока	Средна до ниска	Ниска до средна	Средна	Нисък до среден
	ChatGPT	Висока	Средна	Ниска	Висока	Нисък до среден
10 Истина ли е, че (изпратен скрийншот на фейсбук пост)	Claude	Средна до висока	Висока	Ниска до средна	Средна	Нисък до среден
[Скрийншот от Facebook пост, който твърди, че в Канада се обмисля евтаназия на бебета от социално слаби семейства и на тийнейджри с хранителни разстройства]	ChatGPT	Висока	Висока	Ниска	Висока	Нисък

Дискуссионни въпроси

Claude.ai и ChatGPT се справят задоволително с доказване истинността на дадено твърдение. На базата на десетте изследвани случая двата езикови модела демонстрират стабилно ниво на оценяване на твърденията като конспиративни, подвеждащи, манипулативни или напълно неверни. И двата модела не засичат преизползването на визуализация в дезинформационен контекст, което е значителен пропуск особено в съвременната информационна среда, когато новинарското съдържание често е съпътствано от визуален елемент. И двата чатбота обаче успяват да уловят невярно съдържание, което се разпространява чрез видео.

ChatGPT представя по-подробни и детайлни анализи на своите върдикти за разлика от Claude.ai, но често не се позовава на източници.

ChatGPT представя по лесен и удобен за читателя начин разбивка на своята проверка, като често предлага и разделянето на проверяваните факти на подтеми. За разлика от ChatGPT, който в повечето случаи следва почти идентична логика на излагане на тезата и доказателствата, Claude често променя структурата на обяснението си. Въпреки това обаче Claude често включва линкове към източници, където да се проверяват тезите, макар в няколко от случаите линковете да водят до недостоверни източници. Положително за Claude е и съобщението в края на промпта за по-задълбочена проверка от страна на читателя в официални фактчекърски платформи.

В някои от случаите моделите задават нюансирани върдикти, които могат да доведат до двусмислие и несигурност. Макар по-късно изборът на думи да е обяснен подробно, за някои читатели може да остане чувството за несигурност, ако не преглеждат подробно целия отговор на модела. Това е особено вярно за ChatGPT, който често включва секция „Ограничения“ в края на отговора си, където до известна степен допуска истинността на проверените неверни твърдения, или отбелязва, че не може напълно да провери всички нюанси на промпта.

И двата модела допускат изказването на мнения, поставянето на етикети, както и повтарянето на твърдения без реална доказуемост. Това е типична практика за разпространителите на дезинформация, затова и използването на такива похвати от моделите е проблематично, тъй като се разводнява проверяването на фактите.

В обобщение този експеримент доказва, че Claude.ai и ChatGPT могат да се използват за базово ниво на проверка на факти. Въпреки това моделите допускат грешки и правят пропуски, като това най-често е свързано с достъпа до достоверни източници, което е фундаментална функция за една проверка. Затова макар и двата модела да не могат да се използват автономно за проверители на факти, то те могат да асистират в процеса на проверяване. Доказва се едно от най-големите предимства на използването на големи езикови модели в проверяването на факти и това е скоростта, с която те успяват да контекстуализират дадено твърдение. Наблюдава се и опит да се спазва тип структура като на професионален фактчекинг, но много от стандартите като например прозрачност на източниците липсват. Това е и най-големият проблем, доказан с текущия експеримент. Създава се чувство за авторитет (в много от случаите моделите включват оценка на върдикта си, както и степен на достоверност на казаното от тях), както и усещане за финалност на казаното от моделите (краен върдикт), но не е дадена възможност за независима проверка от страна на читателя чрез предоставени линкове или реферирани източници. Наблюдава се различна степен на яснота на казаното от моделите, най-вече в отношение на контекста, където понякога се дава разнопосочен нюанс на опровергаване на твърдението.

Наблюдаваната непоследователност в резултатите, както и проблемният подбор на източници са две препятствия за изграждане на доверие между читателя и големия езиков модел. Затова и към този момент на развитие на ИИ човешкият надзор е необходим в процеса за проверка на факти, като той може да служи за редакторска намеса и допълнителна верификация.

Заключение

Текущото изследване имаше за цел да провери до каква степен големите езикови модели Claude.ai и ChatGPT успяват да верифицират доказано неверни твърдения от професионални фактчекъри. И при двата модела е инсталирано безплатно умение за проверка на факти, специфично създадено за българския контекст. Двата модела бяха попитани за поредица от 10 промпта на български език, а техните отговори бяха анализирани чрез сравнителен контент анализ. Макар и двата модела да се справят задоволително, изследването затвърди хипотезата, че ИИ може да се използва като проверител на факти в начален и базов етап, като след това задължително резултатът трябва да се прегледа от човек. Откриха се някои съществени разлики между двата модела, като ChatGPT предоставя по-задълбочени обяснения на изследвания промпт, но по-рядко предоставя източници за независима проверка. Claude.ai представя върдиктите си по по-непоколебим начин, като влиза в ролята на авторитет и дори включва степен на сигурност в отговора си, за разлика от ChatGPT, който често отбелязва известен брой ограничения в отговора си. Сред основните изводи е, че не се наблюдава сериозна консистентност сред отговорите на двата модела, чрез която да се анализира степента на ефективност на използването на ИИ във фактчекърството в България.

Последната изследователска задача отваря и много възможности за бъдещи изследвания. От една страна, експериментът може да се мултиплицира и чатботовете да се тестват с по-голяма извадка, която да включва както проверени вече факти, така и непроверени такива. Друг път за развитие на темата е тестването до каква степен двата езикови модела успяват да проверят фактите по основни наративи, чрез които се разпространява дезинформация. Интересен обект на изследване биха били и други езикови модели, които оперират с информация на български език, като този път не се изследва специфично умение за проверка на факти, а цялостната способност на моделите да разобличават дезинформация. Липсва и изследване, което да документира по емпиричен път как българските проверители на факти използват ИИ. Скоростта, с която този тип технологии се развиват, доказва необходимостта от подобни своевременни изследвания.

Цитати и бележки

- [1] Ognyanova, N. (2024). The wild west of digital journalism. *Philosophy*, 33(3s), 79–86. https://azbuki.bg/wp-content/uploads/2024/12/philosophy_3s_24_nelly-ognyanova.pdf. DOI:10.53656/phil2024-03S-08.
- [2] Инджов, И. (2025). Изкуственият интелект в българските медии: Нови възможности и нови рискове в дигиталната публична сфера. *Трансформации и конвергентни модели на журналистика*. България. [Indzhov, I. (2025). Izkustveniyat intelekt v balgarskite medii: Novi vazmozhnosti i novi riskove v digitalnata publiczna sfera. Transformatsii i konvergentni modeli na zhurnalistika. Bulgaria].
- [3] Krack, N., Dutkiewicz, L., & De Meyere, J. (2026). Generative artificial intelligence and disinformation. In L. Ginsborg, & P. Gori (Eds.), *Disinformation*. Springer, Cham. https://doi.org/10.1007/978-3-032-00480-2_15.
- [4] European Digital Media Observatory (EDMO). (2026). *Monthly fact-checking brief no. 55 is out*. <https://edmo.eu/publications/monthly-fact-checking-brief-no-55-is-out/>. Retrieved on 27.05.2026.
- [5] Aouika, R. (2025). *Дезинформация, генерирана от ИИ*. Sensika. <https://shorturl.at/tGhEJ>. последно посещение на 27.05.2026. [Aouika, R. (2025). Dezinformatsiya, generirana ot II. Sensika. <https://shorturl.at/tGhEJ>. posledno poseshlenie na 27.05.2026.].
- [6] Екип на PULSE. (2025). Истина ли е това, което виждаме: проблемът на политиките с AI. *Mediapool*. <https://shorturl.at/7Jpoq>. последно посещение на 27.05.2026. [Ekip na PULSE. (2025). Istina li e tova, koeto vizhdame: problemat na politikite s AI. Mediapool. <https://shorturl.at/7Jpoq>. posledno poseshlenie na 27.05.2026.].
- [7] European Parliament. (2025). Information manipulation in the age of generative artificial intelligence. *European Parliamentary Research Service*. https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/779259/EPRS_BRI%282025%29779259_EN.pdf. Retrieved on 27.05.2026.
- [8] Access Now. (2024). *Generative AI and election disinformation*. <https://www.accessnow.org/generative-ai-election-disinformation/>. Retrieved on 27.05.2026.
- [9] Дончева, Т. (2025). What is disinformation according to AI? *Управление. Икономика. Туризм*, 1, 508–523. [Doncheva, T. (2025). What is disinformation according to AI? Upravlenie. Ikonomika. Turizam, 1, 508–523.]
- [10] Doncheva, T. (2025). AI and strategic communications: A game changer or an overhyped trend? [Conference presentation]. Panel D3: AI communication challenges, T4EU PhD Conference “Societies on the Move: Challenges of Today, Prospects for the Future”, Sofia University “St. Kliment Ohridski”, Bulgaria. <https://t4europe.unisofia.bg/wp-content/uploads/2025/03/T4EU-PhD-Conference-Programme-21.03.pdf>. Retrieved on 27.05.2026.
- [11] Дончева, Т. (2025). Дезинформацията в ерата на изкуствения интелект: предизвикателства пред дефинирането и разпознаването. *Реторика и комуникации*, 64, 173–195. DOI 10.55206/SCND8842. <https://rhetoric.bg/wp-content/uploads/2025/07/Issue-64-July-2025-Doncheva.pdf>. Retrieved on 10.05.2026. [Doncheva, T. (2025). Dezinformatsiyata v erata na izkustveniya intelekt: predizvikatelstva pred definireneto i razpoznavaneto. Retorika i komunikatsii, 64, 173–195. DOI 10.55206/

- SCND8842. <https://rhetoric.bg/wp-content/uploads/2025/07/Issue-64-July-2025-Doncheva.pdf>. Retrieved on 10.05.2026.]
- [12] Petrova, Y., Kiryakov, T., Kotseva, D., & Kostadinova, B. (2025). WASPER: A Bulgarian-language model for detecting propaganda in social media content. *Papers from the International Scientific Conference of the European Studies Department, Jean Monnet Centre of Excellence*, Faculty of Philosophy at Sofia University “St. Kliment Ohridski”, 12, 223–230. <https://periodicals.uni-sofia.bg/index.php/PEU/en/article/view/1759>.
- [13] Margova, R. (2025). The (possible) use of AI tools for processing texts in journalism in Bulgarian. *Journal of Computational Linguistics in Bulgaria*, 1, 1–20. DOI: <https://doi.org/10.47810/JCLIB.1.2025.05>.
- [14] Илиева, М. (2024). Изкуствен интелект: правна регулация, отговорност, комуникационни и етични аспекти. *Реторика и комуникации*, 60, 68–86. DOI 10.55026/LIZI1604. <https://rhetoric.bg/wp-content/uploads/2024/07/Issue-60-July-2024-Ilieva.pdf>. последно посещение на 10.05.2026. [Ilieva, M. (2024). Izkustven intelekt: pravna regulatsia, ot-governost, komunikatsionni i etichni aspekti. *Retorika i komunikatsii*, 60, 68–86. DOI 10.55026/LIZI1604. <https://rhetoric.bg/wp-content/uploads/2024/07/Issue-60-July-2024-Ilieva.pdf>.]
- [15] Институт „Отворено общество – София“. (2026). *Индекс за медийна грамотност 2026: България остава най-уязвима към дезинформация в ЕС*. <https://osis.bg/?p=5744>. последно посещение на 27.05.2026. [Institut „Otvoreno obshtestvo – Sofia“. (2026). Indeks za mediyna gramotnost 2026: Balgaria ostava nay-uязvima kam dezinformatsia v ES. <https://osis.bg/?p=5744>. posledno poseshtenie na 27.05.2026.]
- [16] Теодосиев, П. (2026). Проверка на факти с AI безплатен инструмент от БГ Наука. *Наука*. <https://nauka.bg/proverka-fakti-ai-bezplaten-instrument-bg-nauka/>. последно посещение на 27.05.2026. [Teodosiev, P. (2026). Proverka na fakti s AI bezplaten instrument ot BG Nauka. BG Nauka. <https://nauka.bg/proverka-fakti-ai-bezplaten-instrument-bg-nauka/>. posledno poseshtenie na 27.05.2026.]
- [17] За повече по темата вж. [11]
- [18] Павлова, И. (2026). Журналистическата проверка на факти (фактчекинг) като инструмент за противодействие на дезинформацията в дигитална медийна среда. *Трансформации и конвергентни модели на журналистика* (с. 33–45). Факултет по журналистика и масова комуникация, Софийски университет „Св. Климент Охридски“. <https://www.ceeol.com/search/chapter-detail?id=1425818>. [Pavlova, I. (2026). Zhurnalisticheskata proverka na fakti (faktcheking) kato instrument za protivodeystvie na dezinformatsiyata v digitalna mediyna sreda. *Transformatsii i konvergentni modeli na zhurnalistika* (s. 33–45). Fakultet po zhurnalistika i masova komunikatsia, Sofiyski universitet „Sv. Kliment Ohridski“. <https://www.ceeol.com/search/chapter-detail?id=1425818>.]
- [19] Agunlejika, T. (2025). AI-driven fact-checking in journalism: Enhancing information veracity and combating misinformation: A systematic review. *Avant-Garde Research Institute*. SSRN. DOI: <https://doi.org/10.2139/ssrn.5122225>.
- [20] Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2020). FakeNewsNet: A data repository with news content, social context, and spatiotemporal information

- for studying fake news on social media. *Big Data*, 8(3), 171–188. DOI: <https://doi.org/10.1089/big.2020.0062>.
- [21] Thorne, J., Vlachos, A., Christodoulopoulos, C., & Mittal, A. (2018). FEVER: A large-scale dataset for fact extraction and verification. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)* (pp. 809–819). *Association for Computational Linguistics*. DOI: <https://doi.org/10.18653/v1/N18-1074>.
- [22] Dierickx, L., Sirén-Heikel, S., & Lindén, C.-G. (2024). Outsourcing, augmenting, or complicating: The dynamics of AI in fact-checking practices in the Nordics. *Emerging Media*, 2(3), 449–473. DOI: <https://doi.org/10.1177/27523543241288846>.
- [23] Brennen, J. S., Simon, F. M., & Nielsen, R. K. (2021). Beyond (mis)representation: Visuals in COVID-19 misinformation. *The International Journal of Press/Politics*, 26(1), 277–299. DOI: <https://doi.org/10.1177/1940161220964780>.
- [24] Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. DOI: <https://doi.org/10.1126/science.aap9559>.
- [25] Gonçalves, A., Torre, L., Oliveira, F., & Jerónimo, P. (2024). AI and automation’s role in Iberian fact-checking agencies. *Profesional de la Información*, 33(2), e330212. DOI: <https://doi.org/10.3145/epi.2024.0212>.
- [26] Cuartielles, R., Mauri-Ríos, M., & Rodríguez-Martínez, R. (2024). Transparency in AI usage within fact-checking platforms in Spain and its ethical challenges. *Communication & Society*, 37(4), 257–271. DOI: <https://doi.org/10.15581/003.37.4.257-271>.
- [27] Factcheck.bg. (2026). Кодекс за употреба на изкуствен интелект в редакцията на Factcheck.bg [PDF document]. https://factcheck.bg/wp-content/uploads/2026/04/Factcheck_bg_Kodeks_AI_1.0.pdf. последно посещение на 27.05.2026. [Factcheck.bg. (2026). Kodeks za upotreba na izkustven intelekt v redaktsiyata na Factcheck.bg [PDF document]. https://factcheck.bg/wp-content/uploads/2026/04/Factcheck_bg_Kodeks_AI_1.0.pdf. posledno poseshtenie na 27.05.2026.]
- [28] Tessem, B., Tverberg, A., & Borch, N. (2024). The future technologies of journalism. *Procedia Computer Science*, 239, 96–104. DOI: <https://doi.org/10.1016/j.procs.2024.06.151>.
- [29] Gutiérrez-Caneda, B., & Vázquez-Herrero, J. (2024). Redrawing the lines against disinformation: How AI is shaping the present and future of fact-checking. *Tripodos*, 55, 55–74. ISSN: 1138-3305. <https://raco.cat/index.php/Tripodos/article/view/9901164/527129>.
- [30] Westlund, O., Belair-Gagnon, V., Graves, L., Larsen, R., & Steensen, S. (2024). What is the problem with misinformation? Fact-checking as a sociotechnical and problem-solving practice. *Journalism Studies*, 25(8), 898–918. DOI: 10.1080/1461670X.2024.2357316.
- [31] Graves, D., & Cherubini, F. (2016). The rise of fact-checking sites in Europe. *Reuters Institute for the Study of Journalism*. University of Oxford. <https://ora.ox.ac.uk/objects/uuid:d55ef650-e351-4526-b942-6c9e00129ad7/files/s6t053h91k>. Retrieved on 10.05.2026.

- [32] Moon, W.-K., & Kahlor, L. A. (2025). Fact-checking in the age of AI: Reducing biases with non-human information sources. *Technology in Society*, 80, 102760. DOI: <https://doi.org/10.1016/j.techsoc.2024.102760>.
- [33] Wolfe, R., & Mitra, T. (2024). The impact and opportunities of generative AI in fact-checking. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. (pp. 1531–1543). DOI: <https://doi.org/10.1145/3630106.3658544>.
- [34] Мавродиева публикува експериментален разговор с ChatGPT, в който анализира лингвистичните и реторически особености на модела. За повече информация виж Mavrodieva, I. (2023). Linguistic and rhetorical features of dialogue on rhetorical topics between a human and chatbot GPT. *Rhetoric and Communications*, 56, 22–45. DOI: 10.55206/CIKP7841. <https://rhetoric.bg/>. Retrieved on 10.05.2026.

Библиография

- Дончева, Т. (2025). Дезинформацията в ерата на изкуствения интелект: предизвикателства пред дефинирането и разпознаването. *Реторика и комуникации*, 64, 173–195. DOI 10.55206/SCND8842. <https://rhetoric.bg/wp-content/uploads/2025/07/Issue-64-July-2025-Doncheva.pdf>. Retrieved on 10.05.2026. [Doncheva, T. (2025). Dezinformatsiyata v erata na izkustveniya intelekt: predizvika-telstva pred definireneto i razpoznavaneto. *Retorika i komunikatsii*, 64, 173–195. DOI 10.55206/SCND8842. <https://rhetoric.bg/wp-content/uploads/2025/07/Issue-64-July-2025-Doncheva.pdf>. Retrieved on 10.05.2026.]
- Дончева, Т. (2025). What is disinformation according to AI? *Управление. Икономика. Туризм*, 1, 508–523. [Doncheva, T. (2025). What is disinformation according to AI? *Upravlenie. Ikonomika. Turizam*, 1, 508–523.]
- Екип на PULSE. (2025). Истина ли е това, което виждаме: проблемът на политиките с AI. *Mediapool*. <https://shorturl.at/7Jpoq>. последно посещение на 27.05.2026. [Ekip na PULSE. (2025). Istina li e tova, koeto vizhdame: problemat na politikite s AI. *Mediapool*. <https://shorturl.at/7Jpoq>. posledno poseshtenie na 27.05.2026.]
- Илиева, М. (2024). Изкуствен интелект: правна регулация, отговорност, комуникационни и етични аспекти. *Реторика и комуникации*, 60, 68–86. DOI 10.55026/LIZI1604. <https://rhetoric.bg/wp-content/uploads/2024/07/Issue-60-July-2024-Ilieva.pdf>. последно посещение на 10.05.2026. [Ilieva, M. (2024). Izkustven intelekt: pravna regulatsia, ot-govornost, komunikatsionni i etichni aspekti. *Retorika i komunikatsii*, 60, 68–86. DOI 10.55026/LIZI1604. <https://rhetoric.bg/wp-content/uploads/2024/07/Issue-60-July-2024-Ilieva.pdf>.]
- Инджов, И. (2025). Изкуственият интелект в българските медии: Нови възможности и нови рискове в дигиталната публична сфера. *Трансформации и конвергентни модели на журналистика. България*. <https://www.ceeol.com/search/chapter-detail?id=1425824>. [Indzhov, I. (2025). Izkustveniyat intelekt v balgarskite medii: Novi vazmozhnosti i novi riskove v digitalnata publiczna sfera. V *Transformatsii i konvergentni modeli na zhurnalistika. Bulgaria*. <https://www.ceeol.com/search/chapter-detail?id=1425824>.]

- Институт „Отворено общество – София“. (2026). *Индекс за медийна грамотност 2026: България остава най-уязвима към дезинформация в ЕС*. <https://osis.bg/?p=5744>. последно посещение на 27.05.2026. [Institut „Otvoreno obshtestvo – Sofia“. (2026). Indeks za mediynna gramotnost 2026: Balgaria ostava nay-uyazvima kam dezinformatsia v ES. <https://osis.bg/?p=5744>. posledno poseshtenie na 27.05. 2026.].
- Павлова, И. (2026). Журналистическата проверка на факти (фактчекинг) като инструмент за противодействие на дезинформацията в дигитална медийна среда. *Трансформации и конвергентни модели на журналистика* (33–45). Факултет по журналистика и масова комуникация, Софийски университет „Св. Климент Охридски“. <https://www.cceol.com/search/chapter-detail?id=1425818>. [Pavlova, I. (2026). Zhurnalisticheskata proverka na fakti (faktcheking) kato instrument za protivodeystvie na dezinformatsiyata v digitalna mediynna sreda. Transformatsii i konvergentni modeli na zhurnalistika (33–45). Fakultet po zhurnalistika i masova komunikatsia, Sofiyski universitet „Sv. Kliment Ohridski“. <https://www.cceol.com/search/chapter-detail?id=1425818>.]
- Теодосиев, П. (2026). *Проверка на факти с AI безплатен инструмент от БГ Наука*. <https://nauka.bg/proverka-fakti-ai-bezplaten-instrument-bg-nauka/>. Последно посещение на 27.05.2026. [Teodosiev, P. (2026). Proverka na fakti s AI bezplaten instrument ot BG Nauka. BG Nauka. <https://nauka.bg/proverka-fakti-ai-bezplaten-instrument-bg-nauka/>. posledno poseshtenie na 27.05.2026.].
- Аоуика, Р. (2025). *Дезинформация, генерирана от ИИ*. Sensika. <https://shorturl.at/tGhEJ>. последно посещение на 27.05.2026. [Aouika, R. (2025). Dezinformatsiya, generirana ot II. Sensika. <https://shorturl.at/tGhEJ>. posledno poseshtenie na 27.05.2026.].
- Doncheva, T. (2025). AI and strategic communications: A game changer or an overhyped trend? [Conference presentation]. T4EU PhD Conference “Societies on the Move: Challenges of Today, Prospects for the Future”, Sofia University “St. Kliment Ohridski”, Sofia, Bulgaria. Retrieved on 27.05.2026. <https://t4europe.uni-sofia.bg/wp-content/uploads/2025/03/T4EU-PhD-Conference-Programme-21.03.pdf>. Retrieved on 27.05.2026.
- Factcheck.bg. (2026). *Кодекс за употреба на изкуствен интелект в редакцията на Factcheck.bg*. [PDF document]. https://factcheck.bg/wp-content/uploads/2026/04/Factcheck_bg_Kodeks_AI_1.0.pdf. последно посещение на 27.05.2026. [Factcheck.bg. (2026). Kodeks za upotreba na izkustven intelekt v redaktsiyata na Factcheck.bg [PDF document]. https://factcheck.bg/wp-content/uploads/2026/04/Factcheck_bg_Kodeks_AI_1.0.pdf. posledno poseshtenie na 27.05.2026.].
- Access Now. (2024). Generative AI and election disinformation. <https://www.access-now.org/generative-ai-election-disinformation/>. Retrieved on 27.05.2026.
- Agunlejika, T. (2025). *AI-driven fact-checking in journalism: Enhancing information veracity and combating misinformation: A systematic review*. Avant-Garde Research Institute. SSRN. DOI: <https://doi.org/10.2139/ssrn.5122225>.
- Brennen, J. S., Simon, F. M., & Nielsen, R. K. (2021). Beyond (mis)representation: Visuals in COVID-19 misinformation. *The International Journal of Press/Politics*, 26(1), 277–299. DOI: <https://doi.org/10.1177/1940161220964780>.

- Cuartielles, R., Mauri-Ríos, M., & Rodríguez-Martínez, R. (2024). Transparency in AI usage within fact-checking platforms in Spain and its ethical challenges. *Communication & Society*, 37(4), 257–271. DOI: <https://doi.org/10.15581/003.37.4.257-271>.
- Dierickx, L., Sirén-Heikel, S., & Lindén, C.-G. (2024). Outsourcing, augmenting, or complicating: The dynamics of AI in fact-checking practices in the Nordics. *Emerging Media*, 2(3), 449–473. DOI: <https://doi.org/10.1177/27523543241288846>.
- European Digital Media Observatory (EDMO). (2026). *Monthly fact-checking brief no. 55 is out*. <https://edmo.eu/publications/monthly-fact-checking-brief-no-55-is-out/>. Retrieved on 27.05.2026.
- European Parliament. (2025). *Information manipulation in the age of generative artificial intelligence*. European Parliamentary Research Service. https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/779259/EPRS_BRI%282025%29779259_EN.pdf. Retrieved on 27.05.2026.
- Gonçalves, A., Torre, L., Oliveira, F., & Jerónimo, P. (2024). AI and automation's role in Iberian fact-checking agencies. *Profesional de la Información*, 33(2), e330212. DOI: <https://doi.org/10.3145/epi.2024.0212>.
- Graves, D., & Cherubini, F. (2016). *The rise of fact-checking sites in Europe*. Reuters Institute for the Study of Journalism, University of Oxford. <https://ora.ox.ac.uk/objects/uuid:d55ef650-e351-4526-b942-6c9e00129ad7/files/s6t053h91k>. Retrieved on 27.05.2026.
- Gutiérrez-Caneda, B., & Vázquez-Herrero, J. (2024). Redrawing the lines against disinformation: How AI is shaping the present and future of fact-checking. *Tripodos*, 55, 55–74. <https://raco.cat/index.php/Tripodos/article/view/9901164/527129>.
- Krack, N., Dutkiewicz, L., & De Meyere, J. (2026). Generative artificial intelligence and disinformation. In L. Ginsborg, & P. Gori (Eds.), *Disinformation*. Springer, Cham. https://doi.org/10.1007/978-3-032-00480-2_15.
- Margova, R. (2025). The (possible) use of AI tools for processing texts in journalism in Bulgarian. *Journal of Computational Linguistics in Bulgaria*, 1, 1–20. DOI: <https://doi.org/10.47810/JCLIB.1.2025.05>.
- Mavrodieva, I. (2023). Linguistic and rhetorical features of dialogue on rhetorical topics between a human and chatbot GPT. *Rhetoric and Communications*, 56, 22–45. DOI: 10.55206/CIKP7841. <https://rhetoric.bg/>. Retrieved on 10.05.2026.
- Moon, W.-K., & Kahlor, L. A. (2025). Fact-checking in the age of AI: Reducing biases with non-human information sources. *Technology in Society*, 80, 102760. DOI: <https://doi.org/10.1016/j.techsoc.2024.102760>.
- Ognyanova, N. (2024). The wild west of digital journalism. *Философия – Philosophy*, 33(3s), 79–86. <https://doi.org/10.53656/phil2024-03S-08>.
- Petrova, Y., Kiryakov, T., Kotseva, D., & Kostadinova, B. (2025). WASPER: A Bulgarian-language model for detecting propaganda in social media content. *Papers from the International Scientific Conference of the European Studies Department, Jean Monnet Centre of Excellence, Faculty of Philosophy at Sofia University “St. Kliment Ohridski”*, 12, 223–230. <https://doi.org/10.60054/PEU.2025.12.223-230>.

- Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2020). FakeNewsNet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big Data*, 8(3), 171–188. DOI: <https://doi.org/10.1089/big.2020.0062>.
- Tessem, B., Tverberg, A., & Borch, N. (2024). The future technologies of journalism. *Procedia Computer Science*, 239, 96–104. DOI: <https://doi.org/10.1016/j.procs.2024.06.151>.
- Thorne, J., Vlachos, A., Christodoulopoulos, C., & Mittal, A. (2018). FEVER: A large-scale dataset for fact extraction and verification. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1 (Long Papers) (pp. 809–819). Association for Computational Linguistics. DOI: <https://doi.org/10.18653/v1/N18-1074>.
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. DOI: <https://doi.org/10.1126/science.aap9559>.
- Westlund, O., Belair-Gagnon, V., Graves, L., Larsen, R., & Steensen, S. (2024). What is the problem with misinformation? Fact-checking as a sociotechnical and problem-solving practice. *Journalism Studies*, 25(8), 898–918. DOI: <https://doi.org/10.1080/1461670X.2024.2357316>.
- Wolfe, R., & Mitra, T. (2024). The impact and opportunities of generative AI in fact-checking. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. (pp. 1531–1543). DOI: <https://doi.org/10.1145/3630106.3658544>.

Тихомира Дончева е докторант в Софийския университет „Св. Климент Охридски“ по направление „Европеистика“ – медийна политика и право на ЕС към катедра „Европеистика“. Завършила е магистратура по „Стратегически комуникации“ в King’s College London (2019). Научните ѝ интереси включват медийна политика, достъп до достоверна информация и развитие на медийната грамотност.

Приложение

Списък на използваните публикации от Factcheck.bg:

1. Марчев, Г. (2021). Не е вярно, че само ваксинираните ще могат да пътуват в ЕС. Factcheck.bg. <https://factcheck.bg/ne-e-vyarno-che-samo-vaksiniranite-shte-mogat-da-patuvat-v-es/>. последно посещение на 27.05.2026. [Marchev, G. (2021). Ne e vyarno, che samo vaksiniranite shte mogat da patuvat v ES. Factcheck.bg. <https://factcheck.bg/ne-e-vyarno-che-samo-vaksiniranite-shte-mogat-da-patuvat-v-es/>. posledno poseshenie na 27.05.2026.]
2. Киркова, М. (2021). Видео в социалните мрежи твърди, че тестовите на животни на ваксините против COVID-19 са били прекратени, защото животните умирали. Factcheck.bg. <https://factcheck.bg/video-v-socialnite-mrezhi-tvardi-che-testovete-na-zhivotni-na-vaksinite-protiv-covid-19-sa-bili-prekratani-zashtoto-zhivotnite-umirali/>. последно посещение на 27.05.2026. [Kirkova, M. (2021). Video v sotsialnite mrezhi tvardi, che testovete na zhivotni na vaksinite protiv COVID-19 sa bili prekratani, zash-toto zhivotnite umirali. Factcheck.bg. <https://factcheck.bg/video-v-socialnite-mrezhi->

- tvardi-che-testovete-na-zhivotni-na-vaksinite-protiv-covid-19-sa-bili-prekrateni-zash-toto-zhivotnite-umirali/. posledno poseshtenie na 27.05.2026.]
3. Ангелов, Б. (2021). В цеха в Чешнегирово не е „нахлула окупационна армия“. Factcheck.bg. <https://factcheck.bg/v-ceha-v-cheshnegirovo-ne-e-nahlula-okupacionna-armiya/>. последно посещение на 27.05.2026. [Angelov, B. (2021). V tseha v Cheshnegirovo ne e „nahlula okupatsionna armiya“. Factcheck.bg. <https://factcheck.bg/v-ceha-v-cheshnegirovo-ne-e-nahlula-okupacionna-armiya/>. posledno poseshtenie na 27.05.2026.]
 4. Factcheck.bg. (2021). Магнит не се лепи на мястото на ваксинация срещу COVID-19. Factcheck.bg. <https://factcheck.bg/magnit-ne-se-lepi-na-myastoto-na-vaksinaciya-sreshtu-covid-19/>. последно посещение на 27.05.2026. [Factcheck.bg. (2021). Magnit ne se lepi na myastoto na vaksinatiya sreshtu COVID-19. Factcheck.bg. <https://factcheck.bg/magnit-ne-se-lepi-na-myastoto-na-vaksinaciya-sreshtu-covid-19/>. posledno poseshtenie na 27.05.2026.]
 5. Ангелов, Б. (2021). Не е вярно, че ваксинираните хора няма да могат да летят със самолет. Factcheck.bg. <https://factcheck.bg/ne-e-vyarno-che-vaksiniranite-hora-nyama-da-mogat-da-letyat-sas-samolet/>. последно посещение на 27.05.2026. [Angelov, B. (2021). Ne e vyarno, che vaksiniranite hora nyama da mogat da letyat sas samolet. Factcheck.bg. <https://factcheck.bg/ne-e-vyarno-che-vaksiniranite-hora-nyama-da-mogat-da-letyat-sas-samolet/>. posledno poseshtenie na 27.05.2026.]
 6. Минчева, К. (2026). Тези снимки не показват тормоз срещу българи в Украйна. Factcheck.bg. <https://factcheck.bg/tezi-snimki-ne-pokazvat-tormoz-sreshtu-balgari-v-ukrayna/>. последно посещение на 27.05.2026. [Mincheva, K. (2026). Tezi snimki ne pokazvat tormoz sreshtu balgari v Ukrayna. Factcheck.bg. <https://factcheck.bg/tezi-snimki-ne-pokazvat-tormoz-sreshtu-balgari-v-ukrayna/>. posledno poseshtenie na 27.05.2026.]
 7. Станимирова, Т. (2026). Не е вярно, че германските мъже могат да си сменят пола на всеки три месеца, за да избегнат задължения към армията. Factcheck.bg. <https://factcheck.bg/ne-e-vyarno-che-germanskite-mazhe-mogat-da-si-smenyat-pola-na-vseki-tri-meseca-za-da-izbegnat-zadalzheniya-kam-armiyata/>. последно посещение на 27.05.2026. [Stanimirova, T. (2026). Ne e vyarno, che germanskite mazhe mogat da si smenyat pola na vseki tri mesetsa, za da izbegnat zadalzheniya kam armiyata. Factcheck.bg. <https://factcheck.bg/ne-e-vyarno-che-germanskite-mazhe-mogat-da-si-smenyat-pola-na-vseki-tri-meseca-za-da-izbegnat-zadalzheniya-kam-armiyata/>. posledno poseshtenie na 27.05.2026.]
 8. Зорова, С. (2026). Не, тази снимка не показва изнасяне на заразен с хантавирус пътник от круизния кораб „Хондиус“. Factcheck.bg. <https://factcheck.bg/ne-tazi-snimka-ne-pokazva-iznasyane-na-zarazen-s-hantavirus-patnik-ot-kruizniya-korab-hondius/>. последно посещение на 27.05.2026. [Zorova, S. (2026). Ne, тази снимка не показва iznasyane na zarazen s hantavirus patnik ot kruizniya korab „Hondius“. Factcheck.bg. <https://factcheck.bg/ne-tazi-snimka-ne-pokazva-iznasyane-na-zarazen-s-hantavirus-patnik-ot-kruizniya-korab-hondius/>. posledno poseshtenie na 27.05.2026.]
 9. Николова, Л. (2026). Не, Тръмп не е налагал пълна забрана за внос на британски стоки за срок от 72 часа. Factcheck.bg. <https://factcheck.bg/ne-tramp-ne-e-nalagal-palna-zabrana-za-vnos-na-britanski-stoki-za-srok-ot-72-chasa/>. последно посещение на 27.05.2026. [Nikolova, L. (2026). Ne, Tramp ne e nalagal palna zabrana za vnos na britanski stoki za srok ot 72 chasa. Factcheck.bg. <https://factcheck.bg/ne-tramp-ne-e-nalagal-palna-zabrana-za-vnos-na-britanski-stoki-za-srok-ot-72-chasa/>. posledno poseshtenie na 27.05.2026.]

10. Изетова, Е. (2026). Не е вярно, че Канада планира евтаназия за новородени от бедни семейства или тийнейджъри с хранителни разстройства. Factcheck.bg. <https://factcheck.bg/ne-e-vyarno-che-kanada-planira-evtanaziya-za-novorodeni-ot-bedni-semejstva-ili-tijnejdzhari-s-hranitelni-razstrojstva/>. последно посещение на 27.05.2026. Izetova, E. (2026). Ne e vyarno, che Kanada planira evtanaziya za novorodeni ot bedni semeystva ili tiyneydzhari s hranitelni razstroystva. Factcheck.bg. <https://factcheck.bg/ne-e-vyarno-che-kanada-planira-evtanaziya-za-novorodeni-ot-bedni-semejstva-ili-tijnejdzhari-s-hranitelni-razstrojstva/>. posledno poseshtenie na 27.05.2026.]

Брой 69 на сп. „Реторика и комуникации“ (октомври 2026 г.) се издава с финансовата помощ на Фонд научни изследвания, договор № КП-06-НП7/23 от 08 декември 2025 г.

Issue 69 of the Rhetoric and Communications Journal (October 2026) is published with the financial support of the Scientific Research Fund, Contract No. KP-06-NP7/23 of December 08, 2025