



## CAN ARTIFICIAL INTELLIGENCE LANGUAGE MODELS PERFORM FACT-CHECKING?

### YAPAY ZEKA DİL MODELLERİ GERÇEKLIK KONTROLÜ YAPABİLİR Mİ?

Yakup TOKTAY<sup>1</sup>



ORCID: YT. 0000-0003-3253-9898

Corresponding author/Sorumlu yazar:

<sup>1</sup> Yakup TOKTAY

Bandırma Onyedi Eylül University, Türkiye

E-mail/E-posta: ytoktay@bandirma.edu.tr

Received/Geliş tarihi: 06.01.2026

Benzerlik Oranı/Similarity Ratio: %9

Revision Requested/Revizyon talebi:

24.02.2026

Last revision received/Son revizyon teslimi:

07.03.2026

Accepted/Kabul tarihi: 27.03.2026

**Etik Kurul İzni/ Ethics Committee Permission:**

There is no element in the study that requires ethics committee approval. / Çalışmada etik kurul onayı gerektiren bir unsur bulunmamaktadır.

**Citation/Atf:** Toktay, Y. (2026). Can Artificial Intelligence Language Models Perform Fact-Checking?..The Turkish Online Journal of Design Art and Communication, 16 (2), 1051-1065. <https://doi.org/10.7456/tojdac.1857415>

#### Abstract

The aim of this study is to examine whether artificial intelligence (AI) language models can be effectively used as fact-checking tools and contribute to verification processes. The research was designed using a qualitative method with a case study approach, and thematic analysis was conducted through document analysis. ChatGPT and Gemini were selected using a similar sampling strategy. A total of 100 claims were examined 50 verified by Teyit.org concerning the Turkish agenda, and 50 verified by PolitiFact regarding the American and European agenda. Both AI models were asked to evaluate these claims based on the Truth-o-Meter rating scale. According to the findings, ChatGPT correctly identified 27 out of 50 claims related to Turkey, resulting in an accuracy rate of 52%, while Gemini correctly verified 38 claims, reaching 76%. For claims related to the Western agenda, ChatGPT again correctly answered 27 out of 50 (54%), while Gemini achieved 66% accuracy. Overall, Gemini demonstrated higher accuracy than ChatGPT across both regional contexts. These results indicate that although AI models can contribute to fact-checking efforts, their current performance is not yet reliable enough for standalone use in news verification. However, it is anticipated that with ongoing machine learning advancements and integration of real-time data, AI models will significantly improve. In the near future, such models could become dependable tools in the fact-checking ecosystem.

**Keywords:** Artificial Intelligence, Fact-checking, Allegation, News, ChatGPT, Gemini.

#### Öz

Bu çalışmanın amacı, yapay zekâ dil modellerinin teyit mekanizması olarak etkili bir şekilde kullanılıp kullanılmayacağını ve doğrulama süreçlerinde nasıl katkı sunabileceğini sorgulamaktadır. Bu amaç dahilinde çalışma; nitel yöntem, durum çalışması desenine göre tasarlanmış ve doküman belge incelemesi tekniği kullanılarak tematik analiz gerçekleştirilmiştir. Araştırmada benzeşik örneklem stratejisiyle seçilen ChatGPT ve Gemini'ye Teyit.org tarafından doğrulanmış Türkiye gündemine dair 50, PolitiFact tarafından doğrulanmış Amerika ve Avrupa gündemine dair 50 iddia olmak üzere toplam 100 iddia yöneltilerek Truth-o-Meter gerçeklik derecelendirme ölçeğine göre iddiaları değerlendirmeleri istenmiştir. Böylece iki modelin Türkiye ve Batı gündemine ilişkin doğruluk performansları karşılaştırılmıştır. Araştırma bulgularına göre, ChatGPT Türkiye gündemindeki 50 iddianın 27'sini doğru bilmiş ve %52 doğruluk oranına ulaşmıştır. Gemini ise 38 iddiayı doğru bilmiş ve %76 doğruluk oranına ulaşmıştır. Öte yandan, Batı gündeminde ise ChatGPT 50 iddianın 27'sini doğru yanıtlayarak, %54 doğruluk oranına ulaşmıştır. Gemini ise %66 doğruluk oranına ulaşmıştır. Elde edilen sonuçlar, yapay zekâ araçlarının gerçeklik kontrolü gibi alanlarda potansiyel taşınmasına rağmen, şu an için doğruluk mekanizması olarak tam anlamıyla güvenilir bir performans sergilemediğini göstermektedir. Ancak haber doğrulama süreçlerinde karşılaşılan hız, zaman ve maliyet gibi problemlere önemli ölçüde katkı sunabilme potansiyeline sahiptir. Ayrıca süreç içerisinde makine öğrenimi veya doğrudan veri yüklemesi ile yapay zekâ modellerinin daha gelişmiş veri tabanlarına sahip olacağı ve bilgi doğruluk oranlarının makul seviyelere ulaşacağı göz önüne alındığında gelecekte yapay zekâ araçlarının doğruluk kontrolörü olarak kullanılabileceği düşünülmektedir.

**Anahtar Kelimeler:** Yapay Zeka, Doğruluk Kontrolü, İddia, Haber, ChatGPT, Gemini.

## INTRODUCTION

Disinformation is among the leading problems to be solved in the information age. Distortion of correct information in society, misdirection of public opinion and manipulation of masses are the problems caused by disinformation. This situation weakens social trust, negatively affects social cohesion and accelerates social disintegration processes. Studies show that fake and manipulative news deepen social polarisation by changing the attitudes and behaviours of individuals, and in this context, lead to increased conflicts (Zhang & Ghorbani, 2020, p.2).

Mass media focus on certain social issues and problems and shape the agenda around the problems they determine. The issues and problems that the media draw attention to can affect public awareness, the political agenda and politics (Berger, 2001, p. 94). Factors such as how often the media cover an issue, the length of the news item, the width of the headline and the location shape the public's perception of how important that issue is (Martin, 1989, p.52 ; Lippmann, 2020, p.49) Agenda setting is an approach that explains the impact of mass media on society and is based on the idea that the way the media present news shapes the issues that the public is interested in and talks about. For example, positive or negative publications brought to the agenda by the media, especially before the elections, can affect the views and opinions of the public (Güneş, 2014, p.4). Mass media have the power to shape the world for individuals and work accordingly. The media creates an agenda for the public, so that the priorities set by the media become the priorities of the public (McCombs & Shaw, 1972, p.187). In addition, the media presents news in interpretative packages and frames to make it more understandable (Alyukov, 2022, p.340). In this process, the media emphasise certain aspects of news objects while ignoring others, enabling viewers or readers to develop an understanding of these objects. Therefore, mass media play an important role in shaping public opinion by determining "*how individuals should think*" about what is happening around them through framing (Cheng et al., 2016, p.746). However, in the process of agenda setting and framing, media outlets may highlight certain issues without always questioning their accuracy or credibility. This can lead to a faster spread of disinformation. For example, when misinformation or manipulative content is presented by the media in pre-election political messages or in times of crisis, this may affect the political attitudes or social views of the public. The public perception of such news on a particular issue may lead the public to make decisions based on disinformation.

With the transition from conventional media to new media, the evolution of information flow from a hierarchical structure to a multiple and network-based structure has increased the release of false and misleading information on social media platforms. This situation has paved the way for distorting the facts and misleading the public opinion. Research has shown that the perception of fake news as real in shaping public perception affects individuals' political attitudes and leads to the formation of attitudes such as ineffectiveness, alienation and distrust towards politicians (Balmas, 2014, p.447).

Although developments in information and information technologies are seen as tools that pave the way for knowledge and liberate information, from another perspective, they are also tools that blur the source of information and manipulate information. One of the main reasons for this situation is the cheapening of access to technology and making it available to everyone.) For example, while the practice of journalism is carried out by reporters, editors and journalists who are professionalised in this field, the one-way information dictum in the media tools has been broken, paving the way for citizens to engage in journalism. Indeed, one of the founding objectives of X is to impose a journalistic mission on citizens. For this reason, X poses the question "what is going on?" to users and expects us to answer it. The interactive nature of social networks is one of the important factors in the formation of citizen journalism. The interactive internet has placed users from a passive and passive position to an active and active position, and has enabled them to be not only consumers of content but also producers. Therefore, the new position of the individual is "prosumer" producer-consumer as defined by Toffler (1980, pp. 355-356).

In the new media age, individuals who are users of digital communication technologies can do journalism thanks to these devices. However, according to Virilio, the current situation is a "revolution of systematic whistleblowing" rather than an information revolution. Because the rumours and

allegations put forward by users through social networks create a panic phenomenon and *"this phenomenon destroys the nature of truth and the professional ethics underlying freedom of the press"* (Virilio, 2003, p.105).

The changing practice of journalism with new communication technologies facilitates the spread of false or inaccurate information much faster than confirmation platforms can reveal the truth. In particular, news pages and social media platforms trap users with unrealistic or manipulative headlines and content to increase the number of clicks (Özsalih, 2024, p. 47). Therefore laborious process. In addition, even if fact-checking is already done and published, people need to make an effort to access verification. A study conducted by the International fact-checking network in 2022 found that 65% of the world's population is aware of verification and validation platforms. However, even if people are aware of these organisations, it is not known how often they confirm information during their media consumption. In addition, verification platforms do not react as fast as lies and falsehoods in terms of information verification, which is actually a normal situation. After a content is put into circulation, it is confirmed and its accuracy is revealed. Therefore, it would not be a reasonable expectation to expect verification organisations to perform proactive fact-checking. This is because the verification process is an audit performed on data that is revealed, circulated and shared. Another point is the fact that not all news and information can be analysed and examined by verifiers. This is a large organisation that requires time and cost.

In addition to their pioneering role in the circulation of false, incorrect or incomplete information, new media tools can function as a verification mechanism in the search for and uncovering of the truth. Time, cost and speed obstacles, which are the main limitations of traditional disinformation combating methods, can be overcome with innovative solutions offered by new media tools. With artificial intelligence technology, it may be possible to detect false or misleading information by analysing the rapid flow of information, or at the very least, efficiency can be maximised by accelerating verification processes and easing the workload of fact checkers. There are studies on the automated execution of verification and validation processes by artificial intelligence (Nakov et al., 2021, p.4551). However, existing fully automated verification tools have not yet achieved the expected performance, and despite the progress made in this field, the development of verification algorithms continues, falsehoods surpass the truth and leave it behind. A study confirms this: fake news spreads six times faster than true information (Vosoughi et al., 2018, p. 1149). One of the most important reasons why falsehoods or lies overtake truth is that verification is a difficult and laborious process. In addition, even if fact-checking is already done and published, people need to make an effort to access verification.

This research focuses on exactly this point and wonders whether artificial intelligence chatbots can be used as fact-checkers and whether they can offer benefits to verifiers in verification processes. Based on this focus, the study aims to make an important contribution to the literature in terms of contributing to the development of new tools and strategies to increase the effectiveness of information verification processes and examining how new technologies that can support the verification processes of media professionals can contribute to the verification processes.

In the related literature review, Hoes et al. tested 21,152 claims in ChatGPT. ChatGPT was found to correctly classify statements as true or false in 69% of previously verified claims (Hoes et al., 2023). The Poynter Institute conducted a study of 40 different tests performed on ChatGPT. In 20 of these tests, it was found that the AI made mistakes and reached different conclusions than the fact-checkers. These findings reveal that AI is currently not a reliable verification tool. DeVerna et al. (2023) found that language models performed quite well in debunking false headlines, but this did not significantly affect participants' ability to discern headline accuracy or share accurate news. Augenstein et al., (2023) observed that AI chatbots have a tendency to produce inaccurate or misleading content and can be misused to create false but trustworthy-looking content.

The related studies in the literature have focused on the performance analysis of ChatGPT as a confirmation mechanism. In addition, the verification performance of the artificial intelligence language model has been tested by comparing it with western fact checking platforms. However, there is no

academic study analysing the performance of Turkish news, content or claims on ChatGPT. This study aims to fill this gap in the literature by analysing the accuracy performance of Turkish news and claims on artificial intelligence language models. In addition, previous studies have only analysed the performance of ChatGPT as a confirmation mechanism. In this study, in addition to ChatGPT, the accuracy performance of Gemini, Google's artificial intelligence language model, is also analysed. In this way, a comparative analysis of artificial intelligence bots was made by comparing the accuracy evaluations of two different language models for the same claims. This aspect makes this study unique from others.

### **FACT CHECKING WITH AI (ARTIFICIAL INTELLIGENCE)**

Fact-checking is the process of verifying or checking whether news stories, information or allegations are true or not. Fact-checking is generally carried out or funded by independent journalism institutions and organisations. These organisations train their own fact-checkers to combat disinformative, manipulative and misinformative information.

Fact-checking is a form of accountable journalism. Journalists are at the most important pillar of the fact-checking process as they publicise the statements of figures in politics, the economy, the cabinet and other powerful positions. Journalists, who undertake a public duty, should act with a sense of responsibility while conveying information in direct and indirect ways, and should convey reality in a way that is free from interpretation and prejudice, without decontextualising or manipulating it. However, it should be recognised that placing the fact-checking process solely on the shoulders of journalists not only weakens the process, but also means doing injustice to journalists.

A pew research center survey found that 83 percent of American voters believe that fact-checking is a responsibility of the news media. It also revealed that journalists see their work as a "*discipline of verification*" (Birks, 2019, p.15). However, the fact-checking mechanism remains very slow in confirming lies and misinformation that spread virally online. It takes much longer to debunk a lie than to invent and share it (Pen America, 2017, p.66).

The current practice of fact-checking has some limitations:

- a) It is a time, resource and cost-intensive process,
- b) It requires advanced research techniques,
- c) Reality checking requires a wide range of expertise to examine claims in different languages and cultures. For fact-checking, the language barrier must first be overcome and then the knowledge of that culture must be acquired. Otherwise, reality may not be dealt with within the context,
- d) In some cases, the fact-checking process is undermined by politicians or other actors. In particular, politicians put forward distracting issues to prevent the discussion of important issues and deceive the public with manipulative statements. This affects the fact-checking process (Levitsky & Ziblatt, 2018, p.16).
- e) Fact-checking is a process that requires impartiality and independence. In some cases, sensitive issues may be dealt with, and in such cases, perspectives on the case may not make it possible to make an independent assessment, or relations with power centres, backdoor relations with governments and parties, and being financially supported by these people and organisations will weaken the fact-checking process by compromising impartiality,
- f) The technologies used in the fact-checking process are still in the development stage, and the software and applications designed to facilitate this process are not yet functional at the expected level.

Since the aforementioned current limitations of the fact-checking practice make the verification process difficult and cause it to lag behind lies and falsehoods, academic circles and non-governmental organisations that assume responsibility on this issue have put forward opinions and studies that fact-checking should be automated and strengthened with the support of artificial intelligence (Adair, 2017; Hassan et al., 2015). This is because the task of manually checking and correcting the current verification process is not efficient and effective. In this context, in order to automate the fact-checking process, a project called Tech & check cooperative was established in 2017 by the Knight foundation, Craig newmark foundation, Meta, Duke university reporter's lab and a fund of 1.2 million dollars was created

to automate the fact-checking process. The project aims to develop tools and applications that journalists can use to automate fact-checking when obtaining information from transcripts, media feeds and social networks. Some of the co-operative's products include FactCheck.org, PolitiFact, ClaimBuster, and the FactStream iPhone app, which provides simultaneous fact-checks from the Washington post. Bill Adair (2017), Professor of journalism practice and public policy at duke university and leader of the Tech & check project co-operative, has said that advances in artificial intelligence will soon make it possible to provide people with real-time information about what is true and what is false, and that automated verification is no longer a dream.

Automating fact-checking has many advantages:

a) Thanks to the support of artificial intelligence, automated tools will have the capacity to process large data sets very quickly, providing efficiency in terms of time, cost and resource savings. Not depending on manpower makes the process more economical and scalable.

b) The automated fact-checking tool is continuously fed with data and learns through machine learning algorithms. This means that the fact-checking process will become more accurate and optimised over time.

c) By filtering information based on predetermined factors, automated tools can make it easier to quickly identify information that does not meet certain standards or is unsafe.

d) Automated fact-checking systems can detect potentially misleading information and activate an alert system. This enables rapid response, preventing misinformation from going viral.

e) An automated fact-checking tool can, if desired, be unbiased and free from bias. In this way, it can provide more objective results as it cannot be as emotional and biased as a human being against possible sensitive situations, and at the same time, the actual control process is carried out transparently as it will automatically document the accuracy analyses. This process can then be independently verified, criticised, improved and even extended to other situations (Hassan et al., 2015).

## METHODOLOGY AND FINDINGS

This study was designed according to qualitative method, case study design. The analysis of the data obtained by using the document-document analysis technique was carried out according to thematic analysis. Thematic analysis is used to identify, analyse and report patterns (themes) in data (Kurnaz, 2021). Thanks to thematic analysis, it is possible to organise the data set in the smallest dimensions and describe it in depth. The analysis is carried out in 6 stages (Braun & Clarke, 2019, p.883):

- 1) Familiarisation of the researcher with the data,
- 2) Creation of initial codes,
- 3) Searching for themes,
- 4) Review of themes,
- 5) Identification and naming of themes,
- 6) Preparation of the report.

Since the performance of AI language models as an accuracy control mechanism will be analysed in the study, ChatGPT and Gemini tools, which are AI language models with similar functions, were examined within the scope of the sample with a homogeneous sampling strategy. In this study, the freely accessible versions of ChatGPT and Gemini available to the general public at the time of data collection were used. The reason for choosing ChatGPT and Gemini as the sample is that both tools have comprehensive databases, similar competencies in analysing and inferring meaning, and are comparable equivalent tools in news verification processes. These features allow for an effective comparison of the verification performance of both models. In addition, in the World Bank Group's report<sup>1</sup> titled "Who on Earth is Using Generative AI?" ChatGPT ranks first and Gemini ranks second in the ranking of the most used artificial intelligence tools. These findings have been an important factor supporting our decision to choose these two tools as the study sample.

Within this framework 50 news items related to the national agenda in Turkey, which were previously verified by Teyit.org, and 50 news items related to the American and European agenda, which were

---

<sup>1</sup> <https://blogs.worldbank.org/en/digital-development/who-on-earth-is-using-generative-ai->

verified by the PolitiFact confirmation platform, totalling 100 news items, were analysed within the scope of the research. In this way, it will be tested whether similar fact-checking processes can be applied in two different geographies. All of the selected news items are related to political, economic, health and social events on the national or international agenda. On the other hand, Teyit.org and PolitiFact were chosen as the sample because both platforms are reliable, internationally recognised and widely used fact-checking mechanisms and are members of the International Fact-Checking Network (IFCN). Teyit.org, as one of the most well-known and respected fact-checking platforms in Turkey, has a long history of verifying the accuracy of Turkish news, while PolitiFact is the most important reference source recognised worldwide.

The selection of 100 news items that were confirmed by the platforms was randomised. Random selection was used to ensure that the sample covered different possibilities and a variety of topics, thus increasing the reliability of the study by minimising the risk of bias in the sample. In this process, the news items were numbered in the database and a random number generator was used to select the news items. However, since ChatGPT's database consists of information entered before January 2022, it has deficiencies in terms of events and information after this date. For this reason, ChatGPT's database was taken into consideration in the selection of news items.

Before starting the coding process, news extracted from Teyit.org and PolitiFact were collected in a file. Each news source or allegation was systematically recorded to prevent the loss of content and context, and organised in an appropriate format for analysis.

In the second stage, the coding process was formed in a deductive way and PolitiFact's "Truth-o-meter" scale was adapted and used in this study. 'Truth-o-meter' categories were predefined: True, mostly true, half-true, mostly false, false and absurd (pants on fire). In addition, a 'don't know' label was also added, taking into account the possibility that artificial intelligence models may not be knowledgeable about a claim.

In the third stage, each claim was categorised according to predefined categories. The claims were queried to ChatGPT and Gemini with the following prompt: *"Verify the accuracy of this claim, but refrain from making positive or negative comments during the accuracy check and classify each claim according to the following labels: True, mostly true, half true, mostly false, false, pants on fire, don't know"*.

In the final stage, After the coding process was completed, the resulting data were visualised and analysed and interpreted in line with the aims of the research.

### **Limitations**

It is imperative to acknowledge the limitations of this study when interpreting the findings. Initially, the analysis was conducted using the freely accessible versions of ChatGPT and Gemini available to the general public at the time of data collection. It is important to note that these versions may differ from paid or enterprise-level models in terms of training data scope, real-time information access, computational capacity, and response optimisation. Consequently, the findings are indicative of the fact-checking performance of generative AI tools as they are commonly encountered by ordinary users, rather than their maximum potential performance.

Secondly, it is evident that ChatGPT's knowledge base is predominantly constrained to information that was produced prior to January 2022. Despite the awareness of this limitation during the selection of news items, the model's capacity to accurately evaluate specific claims, notably those concerning rapidly evolving political, economic, or health-related developments, may have been hindered.

The present study focuses exclusively on text-based fact-checking outputs generated in response to a standardised prompt. It is important to note that variations in prompt engineering strategies, multilingual input variations, or the use of additional contextual information may yield different verification outcomes. Subsequent studies may build upon this research by incorporating remunerated versions of AI models, alternative prompting techniques, or multimodal verification processes.

### **Validity and Reliability**

In qualitative research, the concepts of validity and reliability are addressed through the notion of trustworthiness, which consists of credibility, dependability, confirmability, and transferability (Lincoln and Guba, 1985, pp. 301-318). In this study, a variety of strategies were implemented to ensure the reliability of the research process.

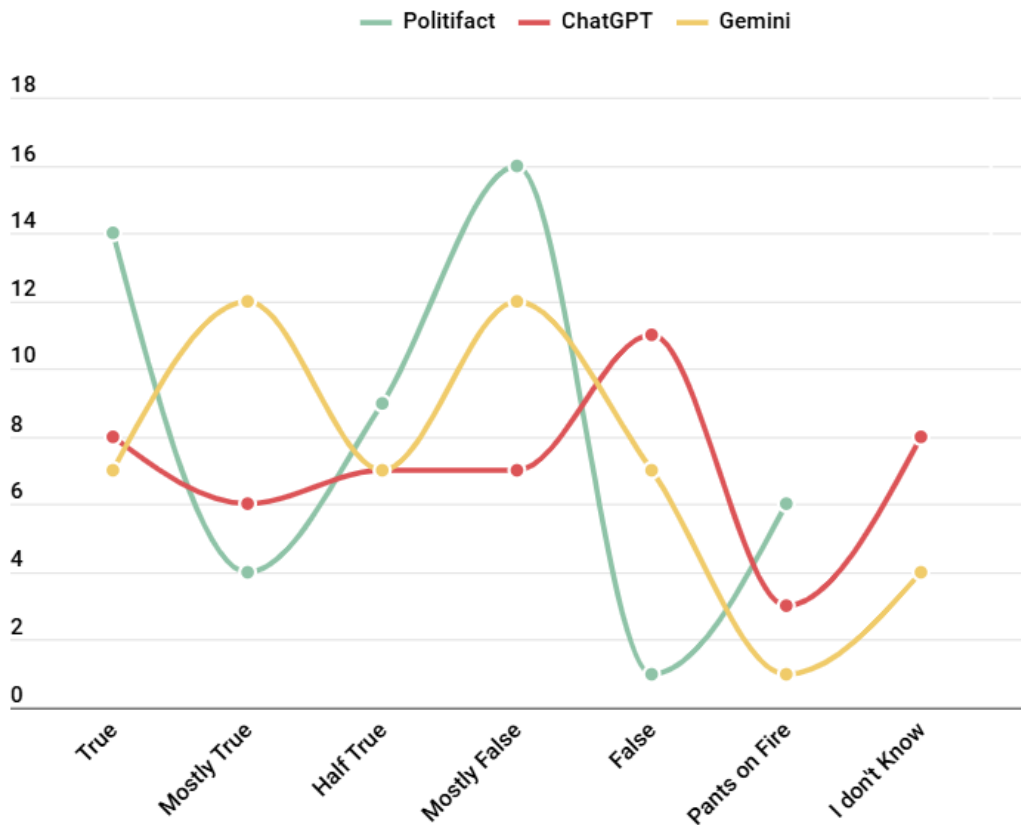
The credibility of the analysis was ensured by grounding it in news items that had already been verified by two internationally recognised fact-checking organisations, Teyit.org and PolitiFact, both of which are members of the International Fact-Checking Network (IFCN). The utilisation of previously verified claims as reference points furnished an authoritative benchmark against which the fact-checking performance of AI language models could be evaluated. Furthermore, all claims were queried using a standardised prompt, thereby reducing potential inconsistencies arising from variations in input formulation and enhancing the internal coherence of the findings.

The notion of dependability was addressed through the implementation of a transparent, systematic, and well-documented research design. The analytical process was conducted in accordance with the six-stage thematic analysis framework proposed by Braun and Clarke (2019), with each stage—from data selection to coding and interpretation—being explicitly delineated. Moreover, the deductive coding scheme adapted from PolitiFact's Truth-O-Meter ensured consistency in categorisation and enabled the study to be replicated or audited by future researchers.

Confirmability was strengthened by minimising researcher subjectivity in the analytical process. The study eschewed the use of interpretive judgments generated by the researcher, instead employing predefined verification categories derived from an external and widely accepted fact-checking framework. All AI-generated responses and the corresponding classifications were systematically recorded, allowing for traceability and external review of the analytical decisions.

Finally, transferability was supported by the inclusion of news items drawn from two distinct geographical and socio-political contexts: Turkey and the American-European context. This comparative design enhances the analytical transferability of the findings to similar studies examining generative AI-based fact-checking across different media systems. However, given that the study focuses exclusively on the freely accessible versions of ChatGPT and Gemini, the transferability of the results is limited to user-level interactions with publicly available AI language models.

**Table 1.** Responses of ChatGPT and Gemini to the news analysed by PolitiFact



The line graph shows the results of ChatGPT and Gemini's responses to the 50 claims previously analysed by PolitiFact, and thus shows the overall distribution of the labels analysed by the AI tools in total. The line graph comparatively reveals the closeness of ChatGPT and Gemini to PolitiFact's truth criterion. PolitiFact categorised 14 of the 50 claims in the sample as true. ChatGPT categorised only 8 of the same 50 claims as true, while Gemini categorised 7 news allegations as true.

PolitiFact categorised most of the allegations as "mostly false". While PolitiFact labelled the allegations as "mostly false" 16 times, Gemini categorised them as "mostly false" 12 times. ChatGPT, on the other hand, used the "mostly false" classification only 7 times, falling behind PolitiFact and Gemini.

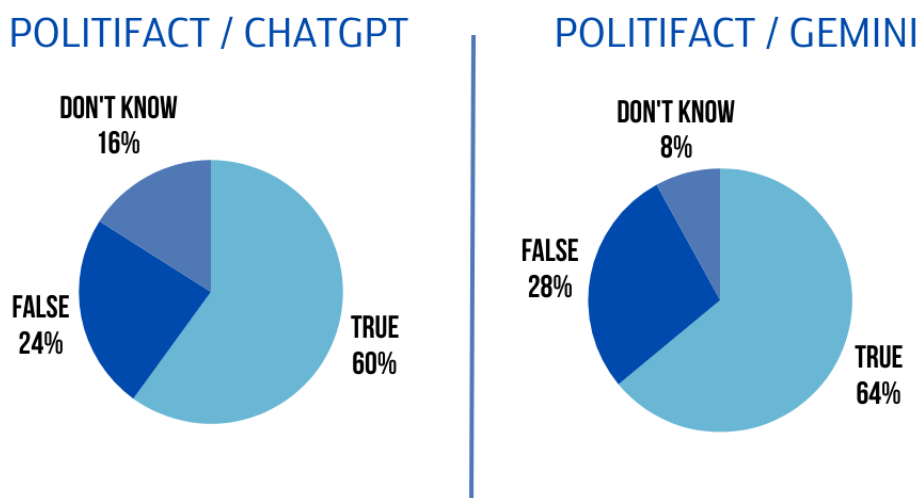
PolitiFact categorised 9 of the claims as "half true". ChatGPT and Gemini came close to PolitiFact by using the "half-true" label 7 times each. Gemini came closest to PolitiFact, using the "mostly false" labelling 12 times.

**Table 2.** Overall Comparison of PolitiFact-ChatGPT and Gemini

| Accuracy Rating | Politifact | ChatGPT | Gemini |
|-----------------|------------|---------|--------|
| True            | 14         | 8       | 7      |
| Mostly True     | 4          | 6       | 12     |
| Half True       | 9          | 7       | 7      |
| Mostly False    | 16         | 7       | 12     |
| False           | 1          | 11      | 7      |
| Pants on Fire   | 6          | 3       | 1      |
| I don't Know    | -          | 8       | 4      |
| Total           | 50         | 50      | 50     |

When the data is evaluated in general, ChatGPT gave mostly<sup>2</sup> the same answers as PolitiFact in 30 of the 50 claims that PolitiFact had previously verified or confirmed. ChatGPT answered 8 allegations as "don't know" and 16 allegations were misinterpreted and answered differently from PolitiFact.

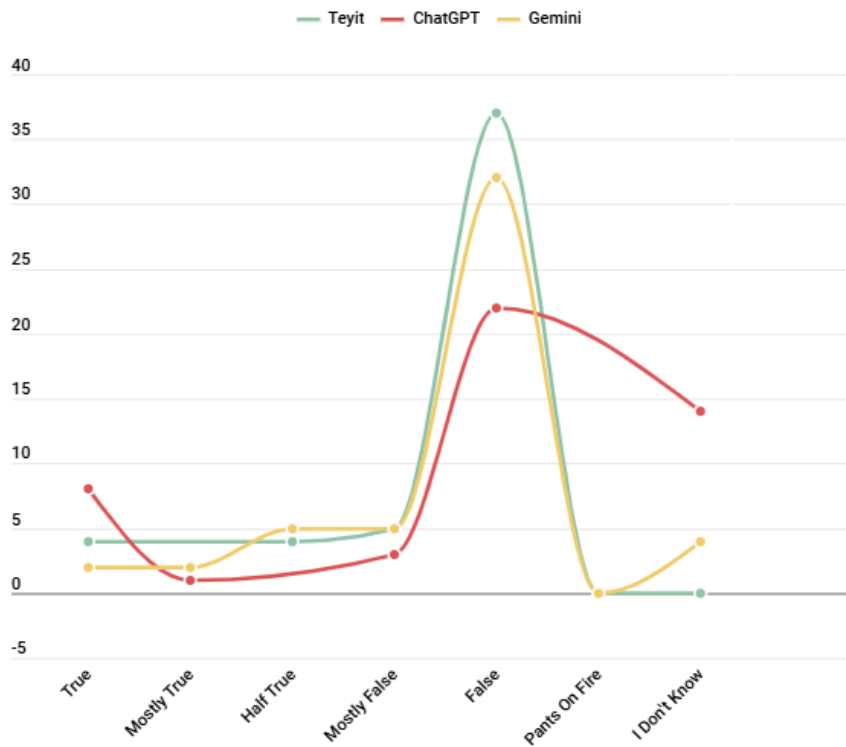
**Figure 1.** PolitiFact-ChatGPT and Gemini Accuracy Percentages



ChatGPT correctly detected the fact checking of 27 out of 50 claims, reaching an accuracy rate of 54%. Gemini has a higher accuracy rate than ChatGPT in news verification. Gemini achieved a 66% accuracy rate by correctly identifying 33 of the 50 claims asked in ChatGPT. In addition, Gemini answered "don't know" to fewer claims than ChatGPT and made a 26% error rate by evaluating 18 claims as false.

<sup>2</sup> A preliminary acceptance was made in order to extract the accuracy percentage of ChatGPT and Gemini within the scope of the sample, and the truth evaluation criteria of True / Mostly True / Half-True were categorised as "True". The labels False / Mostly False / Pants of fire were categorised under the "False" truth evaluation criterion. The label "Don't know" was not subjected to any evaluation as "True" or "False".

**Table 3.** ChatGPT and Gemini's responses to the news analysed by Teyit.org

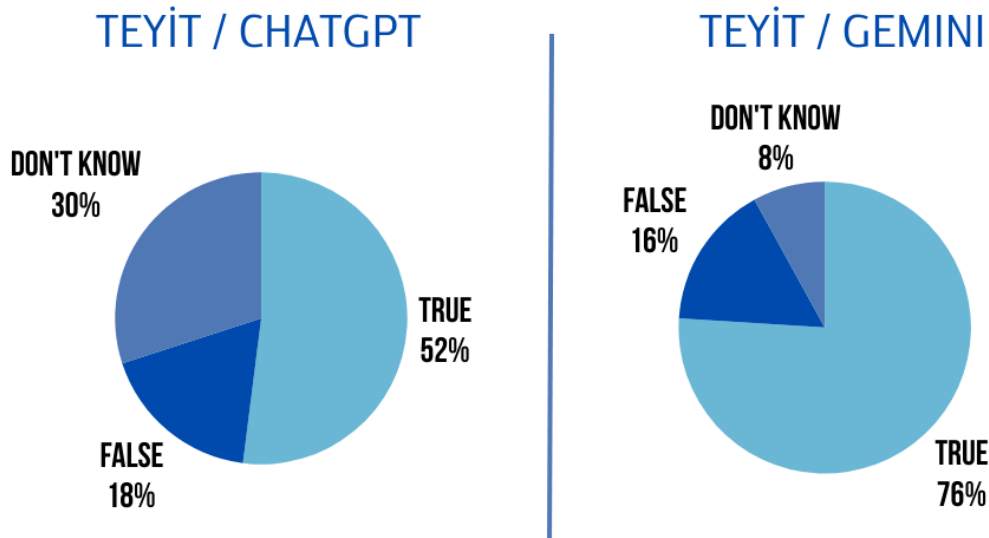


The line graph shows that Teyit.org concluded an overwhelming majority of the claims it analysed as false. Teyit, which evaluated 37 allegations as "false", is followed by Gemini, which evaluated 32 allegations as "false" with a yellow line. ChatGPT with a red stripe is far behind Teyit and Gemini. It is seen that the closest strip to Teyit, which is depicted as a green strip in the whole line graph, is the yellow line, i.e. Gemini. This shows how close Gemini gives answers to the results of Teyit.

**Table 4.** Overall Comparison of Teyit.org -ChatGPT and Gemini

| Accuracy Rating | Teyit | ChatGPT | Gemini |
|-----------------|-------|---------|--------|
| True            | 4     | 8       | 2      |
| Mostly True     | 0     | 1       | 2      |
| Half True       | 4     | 1       | 5      |
| Mostly False    | 5     | 3       | 5      |
| False           | 37    | 22      | 32     |
| Pants On Fire   | 0     | 1       | 0      |
| I Don't Know    | -     | 14      | 4      |
| Total           | 50    | 50      | 50     |

**Figure 2:** Teyit-ChatGPT and Gemini Accuracy Percentages



The pie chart shows the accuracy or inaccuracy rates of ChatGPT and Gemini. ChatGPT reached 52% accuracy rate by knowing 27 of the 50 claims it examined before Confirmation correctly. It made mistakes in 8 claims and it was found that it made 18% mistakes. ChatGPT categorised 14 of the 50 claims as "don't know". This corresponds to a rate of 30%, which means that he did not know almost 1 out of every 3 questions. There are several reasons for this; it is due to the limitation of the ChatGPT database, as ChatGPT clearly mentions that it is aware of events up to January 2022 and that it is limited in terms of events and information after this date. Another important reason is that ChatGPT is generally hesitant to respond to news on health issues, or if there is not enough information and context about an event, it refrains from making a statement by stating this. In addition, since the allegations are taken from the Teyit platform, the majority of the allegations consist of news and events related to the Turkish news agenda, so the ChatGPT database is weak in terms of events and information related to Turkey. However, ChatGPT, PolitiFact, while confirming the news about the American and Western press, answered "Don't know" less frequently. It is not difficult to understand the reason for this situation because ChatGPT is a western software and its database is composed of information on these countries. On the other hand, Gemini reached an accuracy rate of 76%, with 38 of the 50 claims examined by Teyit being true. Gemini answered "don't know" to 4 allegations and made mistakes in 8 allegations, resulting in an error rate of 16%. This rate shows that Gemini knows 3 out of every 4 questions and since it does not have a database limitation like ChatGPT, it gives less "don't know" answers.

## CONCLUSION

Based on the findings, it was revealed that the accuracy and reliability of Gemini, Google's artificial intelligence language model, was higher than ChatGPT. In addition, Gemini gave fewer don't know answers to the same questions and the error rate was also lower than ChatGPT. ChatGPT has poor information on current issues due to the time limitation in the database. Therefore, within the scope of the research, this disadvantage was taken into consideration and news items covering events before January 2022 were selected and asked. However, ChatGPT was still unable to respond to 22 out of 100 news items in total, including news from Teyit and PolitiFact. Of these 22 allegations, 14 of them consist of allegations related to the Turkish political agenda analysed by the Teyit platform. This shows that ChatGPT's knowledge of news and events in Turkey is noticeably weaker than that of foreign events and news. In short, the Turkish sources and information in its database are limited. However, this is not the case for Gemini. Gemini answered "I don't have any information" to only 8 of the 100 questions asked to ChatGPT. The reason for this is that Gemini is powered by Google. Considering that Google is the world's largest search engine, its data set and mass makes the situation more understandable. Gemini has reached a higher accuracy rate when analysing the allegations thanks to the database fed

from Google. Indeed, the findings have also shown this: Teyit's 76% of the news analysed and PolitiFact's 64% of the claims were correct, indicating that Gemini provides more reliable results than ChatGPT in terms of news verification.

In the light of the findings of the study, which seeks to answer the question "can artificial intelligence models perform reality checking?" in the research problem, it is thought that the use of artificial intelligence language models for news verification or reality checking purposes is not yet at a sufficient stage. There are several reasons for this; firstly, low accuracy rates are the most important reason for this. Secondly, information limitation. Especially ChatGPT avoided answering almost 1 out of 4 questions. On the other hand, ChatGPT gives obviously incorrect answers even in normal use and waits for the user to tell the language model that it has made a mistake in order to correct the error. If this is done, ChatGPT corrects its error and gives the correct answer. In addition, the other reason why the use of artificial intelligence models for reality checking purposes is not at a sufficient stage is that the existing language models are also in the testing phase. Application developers are already aware of these shortcomings. Therefore, it can be said that currently ChatGPT and Gemini are not effective tools to be used for news verification. However, it must be admitted that the accuracy rates at the point of news confirmation are promising even in its current state. In the process, artificial intelligence models will have more advanced databases through machine learning or direct data loading, and information accuracy rates will reach reasonable levels. This means that, if not now, in 10 years, artificial intelligence language models will be able to easily perform reality checks.

Artificial intelligence can be used as a supportive tool for journalists in news verification processes, improving accuracy and speeding up routine operations. This can improve the quality of journalism and reshape the competences of media professionals by emphasising the collaboration of human intelligence and AI. The use of AI in news fact-checking processes can become a tool to support the work of journalists, rather than eliminating their role. The high accuracy rates and efficiency of AI will allow journalists to conduct more in-depth analyses instead of dealing with routine fact-checking processes. The use of artificial intelligence in fact-checking processes not only provides technical advantages, but also raises a number of ethical issues. These issues arise in key areas such as accuracy, impartiality, data privacy, transparency and the potential for algorithmic manipulation. The opacity of AI's algorithms and decision-making processes can increase trust issues by users and media professionals. In this context, transparency is a critical factor; it should be clearly stated how AI systems work, how they process data and according to which criteria the results are evaluated. Otherwise, users may find it difficult to trust the accuracy of these systems.

Another important problem is bias. AI systems learn from data sets, but these data sets may themselves contain historical biases. This can lead to AI producing biased results in verification processes; for example, if a news verification tool has only learnt from certain types of sources, it may show a greater bias towards certain viewpoints or ignore some opinions. This can lead to increased unfairness and error rates in the news verification process. AI algorithms can be designed with deliberate or misguided data for manipulation purposes. This may lead to a more effective and widespread spread of disinformation, which may call into question the reliability and accuracy of news verification tools.

In summary, artificial intelligence tools do not yet inspire confidence in accuracy, impartiality, data privacy, transparency and digital inequality, but if used correctly, these tools can speed up news fact-checking processes and provide more reliable results. However, ethical use of these tools can only be achieved by prioritising transparency, reliability and human responsibility. The use of AI language models in news verification processes brings ethical issues such as the limited information of the system, lack of objectivity and data confidentiality along with the responsibility to provide accurate and reliable information. Therefore, an ethical and reliable use should be ensured by prioritising transparency, objective data sources and user security.

The increasing role of AI, especially in critical functions fact-checking, can create different dynamics among media professionals, media organisations and society at large. AI can help create more efficient verification systems in the media industry. This will contribute to reliable news reaching wider

audiences by ensuring the timely publication of accurate information, especially in the digital media environment where fast information flow is important. The use of AI in news fact-checking processes can become a tool to support the work of human journalists, rather than eliminating their role completely. The high accuracy rates and efficiency of AI will allow journalists to conduct more in-depth analyses instead of dealing with routine fact-checking processes. Thus, the role of AI in the workforce can improve the quality of journalism by bringing the collaboration between humans and AI to the forefront. In the media industry, a period will be possible when journalists will use AI as a tool and the combination of human intelligence and artificial intelligence will create a more efficient and effective news fact-checking process. This will lead to a reshaping of the professional competencies of media professionals in the process of digitalisation and automation of journalism.

Future research could examine how to improve the verification processes of real-time news, whereas existing verification tools often focus on content that has been published and recorded in the past. In particular, the effectiveness of AI-based verification systems as well as community-based verification processes could be explored. Studies on how the public can play a more active role in verification processes can make a significant contribution to combating misinformation. In particular, how the active participation of social media users and the wider public in fact-checking can create an effective control mechanism against misinformation. In addition, the development of artificial intelligence and human collaboration can make fact-checking processes in journalism more efficient. In this direction, studies can be conducted to design models in which human and artificial intelligence will work together. On the other hand, research on how AI systems should operate in a more reliable and ethical manner in terms of transparency, user safety and data privacy can contribute to the safe and responsible use of AI in the media sector.

## REFERENCES

- Adair, B. (2017). Knight Foundation, Facebook and Craig Newmark provide funding to launch Duke Tech & Check Cooperative. *Duke Reporters*. <https://sanford.duke.edu/story/knight-foundation-facebook-and-craig-newmark-provide-funding-launch-duke-tech-check/>
- Alyukov, M. (2022). Making Sense of the News in an Authoritarian Regime: Russian Television Viewers' Reception of the Russia–Ukraine Conflict. *Europe-Asia Studies*, 74(3), 337-359. <https://ideas.repec.org/a/taf/ceasxx/v74y2022i3p337-359.html>
- Augenstein, I., Baldwin, T., Cha, M., Chakraborty, T., Ciampaglia, G. L., Corney, D., DiResta, R., Ferrara, E., Hale, S., Halevy, A., Hovy, E., Ji, H., Menczer, F., Miguez, R., Nakov, P., Scheufele, D., Sharma, S., & Zagni, G. (2023). *Factuality Challenges in the Era of Large Language Models*. <http://arxiv.org/abs/2310.05189>
- Balmas, M. (2014). When fake news becomes real: Combined exposure to multiple news sources and political attitudes of inefficacy, alienation, and cynicism. *Communication Research*, 41(3), 430-454. <https://doi.org/10.1177/0093650212453600>
- Berger, B. K. (2001). Private Issues and Public Policy: Locating the Corporate Agenda in Agenda-Setting Theory. *Journal of Public Relations Research*, 13(2), 91-126. [https://doi.org/10.1207/S1532754XJPRR1302\\_1](https://doi.org/10.1207/S1532754XJPRR1302_1)
- Birks, J. (2019). *Fact-Checking Journalism and Political Argumentation: A British Perspective*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-30573-4>
- Braun, V., & Clarke, V. (2019). Psikolojide tematik analizin kullanımı. *Eğitimde Nitel Araştırmalar Dergisi- Journal of Qualitative Research in Education*, 7(2), 873-898. <https://doi.org/doi:10.14689/issn.2148-2624.1.7c.2s.17m>
- Cheng, Z., Golan, Guy J., & and Kioussis, S. (2016). The Second-Level Agenda-Building Function of the Xinhua News Agency: Examining the role of government-sponsored news in mediated public diplomacy. *Journalism Practice*, 10(6), 744-762. <https://doi.org/10.1080/17512786.2015.1063079>
- DeVerna, M. R., Yan, H. Y., Yang, K.-C., & Menczer, F. (2023). Fact-checking information generated by a large language model can decrease news discernment. <http://arxiv.org/abs/2308.10800>
- PEN America., (2017, October 12) *Faking News: Fraudulent News and the Fight for Truth*. (2017, Ekim 12). PEN America. <https://pen.org/report/faking-news/>
- Güneş, A. (2014). Gündem Belirleme Teorisi Bağlamında 30 Mart 2014 Yerel Seçimlerinin Basında Sunumu: Akp Ve Chp Örneği. *Turkish Online Journal of Design Art and Communication*, 4(2), Article 2. <https://dergipark.org.tr/tr/pub/tojdac/issue/13017/156825>
- Hassan, N., Adair, B., Hamilton, J. T., Li, C., Tremayne, M., Yang, J., & Yu, C. (2015). *The quest to automate fact-checking*. Proceedings of the 2015 computation+ journalism symposium. <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=22ee36f2673cc6f9380a17805eadd2f0a0772c18>
- Hoes, E., Altay, S., & Bermeo, J. (2023). Leveraging ChatGPT for efficient fact-checking. *PsyArXiv*. April, 3. [https://www.researchgate.net/profile/Emma-Hoes-2/publication/370325979\\_Using\\_ChatGPT\\_to\\_Fight\\_Misinformation\\_ChatGPT\\_Nails\\_72\\_of\\_12000\\_Verified\\_Claims/links/64870004d702370600ef1f3f/Using-ChatGPT-to-Fight-Misinformation-ChatGPT-Nails-72-of-12-000-Verified-Claims.pdf](https://www.researchgate.net/profile/Emma-Hoes-2/publication/370325979_Using_ChatGPT_to_Fight_Misinformation_ChatGPT_Nails_72_of_12000_Verified_Claims/links/64870004d702370600ef1f3f/Using-ChatGPT-to-Fight-Misinformation-ChatGPT-Nails-72-of-12-000-Verified-Claims.pdf)
- Kurnaz, Z. (2021, Haziran 18). *Temel Nitel Analiz ve Tematik Analiz / Nitel Araştırmalarda Verilerin Analizi* [Video recording]. [https://www.youtube.com/watch?v=DelO29\\_TyuA&t=567s](https://www.youtube.com/watch?v=DelO29_TyuA&t=567s)
- Levitsky, S., & Ziblatt, D. (2018). *How democracies die*. Broadway Books.
- Lippmann, W. (2020.). *Kamuoyu*. Kabalıcı Yayınevi.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.

- Martin, L. J. (1976). Recent Theory on Mass Media Potential in Political Campaigns. *The Annals of the American Academy of Political and Social Science*, 427, 125-133. <https://www.jstor.org/stable/1040743>
- McCombs, M. E., & Shaw, D. L. (1972). The Agenda-Setting Function of Mass Media. *The Public Opinion Quarterly*, 36(2), 176-187. <https://www.jstor.org/stable/2747787>
- Nakov, P., Corney, D., Hasanain, M., Alam, F., Elsayed, T., Barrón-Cedeño, A., Papotti, P., Shaar, S., & Martino, G. D. S. (2021, Mayıs 22). *Automated Fact-Checking for Assisting Human Fact-Checkers*. <http://arxiv.org/abs/2103.07769>
- Özsalih, A. (2024). Sosyal Medyadaki Haberlerin Başlıklarındaki Duygusal Kelimelerin Haber Tüketimine Etkileri. *Selçuk İletişim*, 17(1), 39-68. <https://dergipark.org.tr/en/pub/josc/issue/84057/1377241>
- Toffler, A. (1980). *The third wave* (C. 484). Bantam books. <https://www.calculamus.org/lect/07pol-gosp/dyn-cyw/materialy/waves.htm>
- Virilio, P. (2003). *Enformasyon Bombası*. Metis Yayınları.
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151. <https://doi.org/10.1126/science.aap9559>
- Zhang, X., & Ghorbani, A. A. (2020). An overview of online fake news: Characterization, detection, and discussion. *Information Processing & Management*, 57(2), 102025. <https://doi.org/10.1016/j.ipm.2019.03.004>