



Yazar/Author
Mustafa DEMİR*

Makale Adı/Article Name

**Yapay Zekâ Kopyası (AI Cheating) ve Büyük Dil Modellerinin
Çevrimiçi Sınavlarda Kullanımı**

*AI Cheating and the Use of Large Language Models (LLM) in Online
Examinations*

ÖZ

2020 yılında ortaya çıkan küresel salgın ve sonrasında klasik yaşam tarzı, çalışma hayatı ve eğitim süreçleri kökten değişmeye başlamıştır. Bu dönemde eğitim kurumları uzaktan eğitime yönelmiş; eğitim süreci ile birlikte ölçme ve değerlendirme işlemleri uzaktan eğitim platformları yoluyla yapılmaya çalışılmıştır. Uzaktan eğitim ve online ölçme değerlendirme uygulamaları, pandeminin bitişi ile de devam etmiştir. Bu durum her ne kadar eğitmen ve öğrenciler açısından bazı kolaylıklar sağlasa da online sınavların geçerliliği ile ilgili soru işaretlerini beraberinde getirmiştir. Online sınavlarda yapay zekâ algoritmalarının kullanımına dair veriler, bu sınav modelinin geçerliliği ve güvenilirliğinin sorgulanmasına yol açmıştır. Yapay zekâ uygulamaları arasında, eğitim süreci ile en fazla ilinti kurulabilecek büyük dil modellerinin, çevrimiçi sınav sorularının cevaplanmasında kullanılabileceği tartışılmaya başlanmıştır. Çalışma bu bağlamda büyük dil modellerinin çoğunlukla çoktan seçmeli (test) olarak tasarlanan çevrimiçi sınavlardaki performansını ve günümüzde önemli bir sorun haline gelen yapay zekâ kopyasına (ai cheating) karşı alınabilecek önlemleri araştırmaktadır. Bu amaçla daha önce çevrimiçi sınavlarda kullanılmış sorulardan bir set oluşturulmuş; sorular, büyük dil modellerine yöneltilmiş; bu modellerin yöneltilen soruları doğru/yanlış çözüme potansiyeli ve hangi tip soruların doğru yanıtlandığı sorularının cevabı aranmıştır. Çalışmada, büyük dil modellerinin yöneltilen sorulara çoğunlukla doğru yanıt verdiği görülmüştür. Elde edilen veriler ışığında yapay zekâ kopyasına yönelik önlemler tartışılmıştır.

Anahtar Kelimeler: Yapay Zekâ, Uzaktan Eğitim, Yapay Zekâ Kopyası, Büyük Dil Modelleri

ABSTRACT

In 2020, after the global pandemic traditional lifestyle, working life and educational processes began to change radically. During this period, educational institutions turned to distance learning; attempts were made to carry out the measurement and assessment processes together with the educational process through distance learning platforms. Distance learning and online assessment and evaluation processes continued after the end of the pandemic. Although this situation brought some convenience for lecturers and students, it raised questions about the validity of online examinations. Data on the use of artificial intelligence algorithms in online exams has led to questions about the validity and reliability of this exam model. Among the applications of artificial intelligence, it has been discussed that large language models (LLM), which are most likely to be associated with the educational process, can be used to answer online exam questions. In this context, this study investigates the performance of LLMs in online exams, which are mostly designed as multiple-choice tests, and the measures that can be taken against AI cheating, which has become an important problem. For this purpose, a set of questions was created that had previously been used in online exams. The questions were posed to LLMs and answers to the questions were sought, e.g. the potential of these models to get the questions right/wrong and which types of questions are more likely to be answered correctly. The study found that LLMs answered most of the questions correctly. In the light of the data obtained, preventive measures for AI cheating are discussed.

Keywords: Artificial Intelligence, Distance Education, AI Cheating, Large Language Models



Extended Abstract

With the global pandemic, educational institutions have turned to distance learning in particular; examinations have also been conducted via distance learning systems. Today, many educational institutions run their distance learning systems via Learning Management Systems (LMS). The digitalization of education and the implementation of both education and assessment and evaluation through distance learning have brought some problems. The most important of these is the validity of online exams. It is known that among the applications of artificial intelligence that have entered people's lives with the process, large language models (LLM), which are most likely to be associated with education, can be used to answer exam questions. In this context, this study examines the effectiveness of Chatbot (CB) interfaces and LLMs in online exams designed as multiple-choice test, the measures that can be taken against Artificial Intelligence Cheating (AI cheating) which has become a problem today, and the changes that can occur in question and answer methods. To this end, the questions used in online exams were directed to large language models, the potential of these models to get the questions right and wrong, and which types of questions were more likely to be answered correctly. The study also discusses the measures that can be taken against applications that copy artificial intelligence.

The study consists of two phases. First, the literature was consulted to explain the concept of Artificial Intelligence Replication. As this is a new field and the concept has only recently been discussed, it is not possible to speak of the existence of a broad academic framework. It can be said that the existing academic studies (Abdelaal et al., 2019; Oravec, 2023) mostly focus on problems related to the use of AI for academic purposes. In the second phase, attempts were made to measure the success of artificial intelligence-based large-scale language models in online examinations at associate and undergraduate levels. In this direction, under the purposive sampling model, exam questions used in the online/online assessment and evaluation process were requested from academics teaching in faculties of communication and vocational schools of social and technical sciences. The content of the exams for 17 courses was obtained digitally from academics who agreed to share their data. 173 questions were selected from these exams. Open-ended questions were not used in the study, only multiple-choice questions were analyzed. The selected questions were analyzed individually and divided into two categories. The first category consists of questions that are based purely on knowledge and whose answers can be found with a simple web search. The second category are questions that require interpretation, where more than one topic or piece of information needs to be compared. The categorized questions were asked to chatbots from ChatGPT-3.5, Microsoft Bing (GPT-4), Google Bard, Opera Aria, LLaMA-2-7b, LLaMA-2-13b, LLaMA-2-70b, GPT-3.5 Turbo, which were widely used in January-February 2024. In the further course of the study, the Google Bard model was given the name "Gemini" (February 9, 2024) and the Microsoft Bing chatbot the name "Copilot". The answers given to the questions that the language models were asked to answer through chatbots were compared with the answers from the trainers, and the correct and incorrect answers to the language models were recorded. The recorded answers were classified into two categories, and a situational analysis was conducted. Although chatbots and large language models (LLMs) are new technologies that continue to develop, they have the potential to provide correct answers to most questions asked or to be asked in distance education at universities. In the current study, the assistant named Bing or Copilot answered more than 80% of the questions correctly with the support of the GPT-4 language model. Bots with active Internet search features provide more accurate results. It is thought that the incorrect answers were due to conflicts in the information obtained from different sources.

According to the results of the study, the potential of bots to solve questions with high accuracy with the support of LLMs raises questions regarding the validity and reliability of online exams. The use of artificial intelligence in online exams held in an uncertain area (location, space) without supervision leads to unfair competition. Some measures can be taken to prevent competition. First, it was concluded that the questions asked in the education process should not be based solely on knowledge. It may be considered to restrict the use of questions that arise as a result of basic searches via the Internet, and instead recommend question and answer types that the student can answer verbally or in writing. On the other hand, in the case of online exams, different educational institutions resort to methods such as "turning on the camera" and "environment control". These methods are considered inadequate and primitive. In online exams, advanced artificial intelligence tools can provide real-time assistance by solving questions or providing direct answers. Some of these AI tools can also manage screen sharing or remote surveillance, making it more difficult to detect AI cheating. AI cheating can be prevented in online exams using more sophisticated systems, mostly artificial intelligence. At this point, "online proctoring" applications have come to the fore. Among these applications, Digiexam, Ai-proctor, Mercer-Mettl, ProctorEdu, Respondus Monitor, and other proctoring software are prominent. These software programs aim to stop all potential methods that a

participant may use to cheat on an online exam (screen sharing, enlisting the help of friends or relatives, smartphones, large language models, chat bots) through online proctoring. Here, many parameters come into play, such as tracking a person's eye body movements, controlling tabs, listening to the environment, and monitoring the environment.

There is no doubt that artificial intelligence has brought many innovations into our lives. On the other hand, it is obvious that this development leads to new dilemmas, and that new perspectives need to be developed against these dilemmas. Instead of continuing with the classical education and examination system, it seems inevitable to support the use of artificial intelligence and create more rational education systems.

Giriş

İnsanlık, hem yaşam dinamikleri hem de teknolojik olarak büyük değişimler yaşanan bir dönemden geçmektedir. Öncelikle 2020 yılında başlayan pandemi, bireylerin yaşam alışkanlıklarını kökten değiştirmiştir. Bu süreçte bireyler, uzaktan çalışma modeli ile tanışmış; birçok şirket çalışanlarını evden çalışmaya teşvik etmiştir. Bu model günümüzde de kısmen devam etmektedir. Diğer yandan kamusal alanlardaki popülasyonun azaltılması ve pandeminin kontrol altına alınması için tüm dünyada yüz yüze eğitim öğretim faaliyetlerine ara verilmiştir. Süreç içinde eğitimler ve sınavlar uzaktan eğitim sistemleri yoluyla yapılmıştır. Uzaktan eğitim sürecinde öğrencilerin değerlendirilmesi için çevrimiçi sınavlara başvurulmuştur. Bu sınav sisteminin kullanımı, pandeminin bitişi ile de devam etmektedir. Halihazırda Türkiye’de faaliyet gösteren üniversitelerin büyük kısmı, servis dersi olarak nitelenen Atatürk İlkeleri ve İnkılap Tarihi, İngilizce, Türk Dili ve Edebiyatı gibi dersleri, uzaktan eğitim sistemi üzerinden sürdürmektedir. Eğitimin çevrimiçi olarak sürdürülmesi, ölçme ve değerlendirme işlemlerinin de yine çevrimiçi yapılması gerekliliğini doğurmuştur.

Pandemi ile birlikte insan hayatına dahil olan bir başka yenilik, yapay zekâ algoritmalarına bağlı büyük dil modelleri (BDM) olmuştur. Uzaktan eğitim sisteminin kalıcılık gösterdiği bu süreçte, insanlık daha önce sık sık ifade edilen fakat dijital alanlarda yüzleşmediği yapay zekâ ve yapay zekâ temelli uygulamalarla tanışmıştır. Bunlardan en yaygın bilineni ilk GPT (Generative Pre-trained Transformer / Üretken Önceden Eğitilmiş Dönüştürücü) versiyonu olan GPT-1'dir. GPT-1, OpenAI adlı şirket tarafından 2018'de geliştirilmiştir. Büyük dil modeli olarak adlandırılan bu yapay zekâ (YZ) algoritmasını daha sonra GPT-2 ve GPT-3 gibi gelişmiş versiyonlar takip etmiştir. GPT-2, 2019'da, GPT-3 ise 2020 yılında tanıtılmıştır. GPT-3, 2023 itibarıyla yaygın kullanılan GPT-3.5'in temelini oluşturmuştur. GPT, büyük miktarlardaki metin verisi üzerinde önceden eğitilmiş ve doğal dil işleme görevlerini yerine getirmek için kullanılan bir yapay zekâ modelini ifade etmektedir. Günümüzde GPT-4o'nun tanıtılması ve interaktif özelliklerin gelişmesi ile büyük dil modellerinin hayatın içindeki yeri giderek artmaktadır.

Büyük dil modellerinin pandemi sonrası bireylerin kullanımına açılması ile birlikte, çevrimiçi ölçme değerlendirme uygulamalarında yapay zekâ algoritmalarının kullanılabilmesi; bu durumun haksız sonuçlar doğurabileceği görülmüştür. Yapay zekânın ve büyük dil modellerinin çevrimiçi sınavlarda kullanılması, AI cheating (yapay zekâ kopyası) olarak adlandırılmaktadır (Smith, 2024). Bu süreç sadece çevrimiçi sınavlarda değil özellikle ödev, proje yapımı veya akademik içeriklerin oluşturulması aşamasında da bir sorun olarak karşımıza çıkmaktadır. Bireylerin yapay zekâ kullanarak sınav sorularını cevaplaması, ödevlerini hazırlaması veya makale yazdırması günümüzde yaygın hale gelmiştir.

Konunun etik ve ahlaki boyutlarının ötesinde, yapay zekânın sınavlarda bir kopya unsuru olarak kullanılıp kullanılmayacağına dair çalışmalar kısıtlıdır. Çalışmada, yapay zekâ temelli büyük dil modellerinin çevrimiçi sınavlardaki başarısı ve yapay zekâ yoluyla gerçekleşen sahtecilik uygulamalarına yönelik önlemler detaylıca incelenmiştir.

1. Kavramsal Çerçeve

1.1. Yapay Zekâ

Yapay zekâ kavramı, öğrenme ve belirli bir çerçevede düşünme eyleminin makineler yoluyla yapılabileceği varsayımı üzerine şekillenmiştir (McCarthy vd., 1955). Daha güncel bir yaklaşımla

Russell ve Norvig (2021: 7-8) yapay zekâyı, bulunduğu çevreden komut alan ve eyleme döken unsurlar olarak tanımlamaktadır. Bu aşamada en önemli unsur, makine öğrenmesi olarak gösterilmektedir. Campa'ya göre, makine öğrenmesinin temel bileşenleri olan bilgilerin işlenmesi, anlık karar mekanizması ve geometrik modellemeler gibi özellikler, ilerleyen süreçte duygusal zekânın ortaya çıkışını sağlamıştır (Campa, 2020). Campa'dan yola çıkarak, duygusal zekânın gelişimi ile birlikte yapay zekâ algoritmalarının bireylerin hayatında önemli değişikliklere yol açacağı düşünülmektedir. Johnson (2017), yapay zekânın gelişimi ile ilgili olarak birçok tahminde bulunmaktadır. Johnson'a göre ilerleyen yıllarda yapay zekâ algoritmalarının kullanımı ve 3D yazıcılar ile, canlı dokular dahil birçok materyal üretilebilir olacaktır. Yine bireyler, günlük hayatlarında yardımcı olabilecek asistanlara kavuşacaktır. Bir sonraki aşamada bu asistanların robotikle birleşimi dahi söz konusu olabilecektir. Johnson tarafından ortaya konulan bu öngöründe yer alan zeki asistanlar (Siri, Google Asistan, Alexa vb.), 2024 yılı itibarıyla insan hayatının vazgeçilmez enstrümanlarından biri haline gelmiştir. Bu enstrümanların, robotiğin gelişimi ile birlikte fiziksel varlıklar haline dönüşmesi kaçınılmaz olarak düşünülebilir. Yapay zekânın potansiyeli, bu bağlamda büyük bir güç (Andrade ve Tumelero, 2022) olarak görülmektedir. Andrade ve Tumelero'ya göre (2022), artırılmış zekâ, otonom sistemler, big datanın işlenmesi gibi eylemler, yapay zekânın gerçekleştirebildiği işlemler arasındadır.

Pool'a (1997) göre, yapay zekânın önündeki en büyük engel öğrenme sürecidir ve bu sürecin anlaşılabilmesi için insan beyninin nasıl çalıştığının anlaşılması gerekmektedir. Özellikle öğrenmenin nasıl gerçekleştiğinin detaylı anlaşılması, yapay zekânın gelişimine destek verecektir. Yapay zekâ ile ilgili tartışmalarda, beyin ve bilgisayar arasındaki benzeşim ve ayrılıklar üzerinde durulmaktadır. Doğan (2002), gelecekte duyguları olan (insan duygularını simüle eden) sibernetik makineler olabileceğini iddia etmektedir. Bu düşünce her ne kadar tartışmaya açık olsa dahi, yapay zekânın önünde bulunan "işlem gücü ihtiyacı" (mevcut mikro işlemcilerin yetersizliği) ve enerji problemi düşünüldüğünde, sürecin oldukça uzayabileceği söylenebilir. Ayrıca sibernetik, sadece yapay zekâ ile ilgili olarak değerlendirilmemelidir. Malzeme tedarikinden, işlemci teminine kadar dünya ekonomisinin potansiyelleri ve kısıtlı hammadde kaynakları düşünüldüğünde; robotik sistemlerin yaygınlaşmasının önündeki zorluklar görülmektedir.

Günümüzde yapay zekâ çalışmaları tek merkezden değil, farklı amaç ve motivasyonları olan şirket ve araştırma kurumlarınca yürütülmektedir. Öne çıkan oluşumlardan en popüler olanı "OpenAI" adlı şirket olarak ifade edilebilir. 2015 yılında kurulmuş olan bu şirket özellikle Microsoft gibi köklü şirketlerin de desteğiyle insanlardan daha akıllı olan yapay zekâ (AI - Artificial Intelligence) algoritmaları üretmeyi hedeflemektedir (OpenAI, 2015). OpenAI, bu misyon çerçevesinde 30 Kasım 2022 tarihinde ChatGPT'yi genel kullanıma açmıştır (Kurian vd., 2023).

Yapay zekâ ile birlikte tartışılan konuların başında, bu algoritmaların toplum üzerinde ne tür etkilerinin olacağı gelmektedir. Sheikh (2020) yapay zekânın farklı alanlarda (iş, insan ilişkileri vb.) toplumsal hayata etkilerine dikkat çekmekte; insanlara inovasyon konusunda yardımcı olabileceğini iddia etmektedir. Ayrıca Sheikh, insanların yaşaması için gerekli olan yapısal unsurların izlenmesinde yapay zekâ algoritmalarının kullanılabilirliğini; kendi kendini tamir eden sistemlerle, bireylere sunulan hizmetin kesintisiz sürdürülebilmesinin sağlanabileceğini ifade etmektedir.

Yapay zekâ ile ilgili çoğunlukla pozitif söylemlerin kullanıldığı görülmektedir. Özellikle insan hayatının kolaylaştırılması noktasında yapay zekâ kullanımı ön plana çıkmaktadır. Yapay zekânın gelişimi ile ilgili olumlu başlıkların yanında olumsuz başlıklara da değinmek anlamlı olacaktır. yapay zekâ algoritmalarının hayatın her noktasına girmesi ile birlikte; özellikle insansız hava araçları ve robotların askeri kullanımı gibi sorun alanı sayılabilecek gündemler oluşmaktadır. Yaşanan teknolojik değişimlerle (Lokhorst ve Van den Hoven, 2012; Sharkey, 2012), yapay zekâ tabanlı sistemlerin insanlara vereceği zararlarda hukuki anlamda kimin sorumlu tutulacağı (Asaro, 2012); etik ilkeler ve yapay zekâ algoritmalarının sosyal ilişkiler üzerindeki etkileri (Levy, 2012) gibi birçok konu gündeme gelmektedir. Bu durum sadece savaş makineleri ile ilgili

düşünülmemelidir. Beer, insanlara dijital dünyada neyi göreceğini, duyacağını, kimleri tanıyacağını dayatan, dolayısıyla neyi sevip sevmeyeceğinden, kimlerle ilişki kuracağına kadar birçok alanda toplumsal hayata etki eden ve gündelik hayatta sürekli karşılaşılan yapay zekâ algoritmalarının “değer yüklü” ve “politik” olduklarını (Beer, 2017: 6) ifade etmektedir. Beer’den hareketle ilerleyen dönemlerde ideolojilerin yapay zekâ yoluyla domine edilebilecekleri ve hatta bireye ait sosyal tutumun, yapay zekâ yoluyla belirlenebileceği söylenebilir. Bu olası kontrol mekanizması, insan geleceği için en önemli tehditlerden biri olarak düşünülebilir. Yaşanan tartışmaların ilerleyen yıllarda artarak devam edeceği, özellikle hukuki düzenlemelere yönelik gelişmeler olacağı belirgindir. Her ne kadar yapay zekâ hayatın her noktasında hissedilse de ilgili algoritmalar ile birlikte en fazla gündeme gelen ve insanların en fazla karşılaştıkları sistemler, büyük dil modelleri (large language models) olmuştur. Çalışmanın bu kısmında büyük dil modellerinden bahsedilecektir.

1.2. Büyük Dil Modelleri (LLMs)

Büyük dil modelleri, dilsel birimlerin ve bunlar arasındaki ilişkilerin ağırlıklarıyla ilgili bilgiye sahip, yapay zekâ ağlarının etiketlenmemiş veri kullanarak ve gözetimli ya da yarı gözetimli yöntemler ile eğitilmesi sonucunda ortaya çıkan yapay zekâ araçları (Birhane vd., 2023) olarak tanımlanmaktadır. BDM’ler, insan benzeri dilsel özelliklere sahip ve son derece karmaşık dil çıktıları üretebilmektedir. BDM’lerin 2018 yılında duyurulması ile birlikte özellikle sözel yeteneklerinde ciddi gelişme meydana gelmiştir. BDM’lerin en önemli örnekleri, 175 milyar parametre üzerinde eğitilmiş bir GPT-3 modeli olan ChatGPT de dahil olmak üzere OpenAI GPT (Üretken Ön Eğitilmiş Dönüştürücü) ailesidir (Hayawi, K., Shahriar, S., & Mathew, S. S., 2023). GPT-4’ün hem metinleri hem de görüntüleri işleyebildiği bilinmektedir. GPT-4 ile birlikte sunulan “Sora” adlı görüntü işlemcisi (OPENAI, 2024) her ne kadar birey kullanımına açık olmasa da görüntü işlemenin geleceğine dair işaretler vermektedir. Yapılan akademik çalışmalarda genellikle Sora adlı görüntü işlemcisinden bahsedilmektedir. Yapay zekâ algoritmalarının özellikle video ve ses alanındaki hızlı gelişimi ile birlikte 2024 yılı itibariyle Krea, Kling, Runway gibi görüntü işlemcileri yayınlanmış; kullanıcıların hizmetine sunulmuştur. Bu algoritmalar referans bir resimden, yazıdan veya videodan yeni video içerikleri üretebilmektedir.

BDM’ler ticaret, yazılım, eğitim ve benzer birçok alanda hizmet vermektedir. Bilimde ve eğitimde kullanımının yaygınlaşmasının ardından, BDM’lerin kötüye kullanıma ilişkin tartışmaların ortaya çıktığı görülmektedir (See vd., 2019). BDM’lerin insan diline çok benzer çıktılar elde edebilmesi, bu modellerin kullanımına ilişkin bazı kaygıların oluşmasına neden olmuştur. Liu vd.’ye göre, (2023) ChatGPT gibi dil modellerinin yanıtlar üretebilmesi, insan iletişimini taklit edebilmesi; akademik metinler oluşturabilmesi, etik sorununu gündeme getirmektedir. BDM’ler tarafından üretilen akademik içeriklerin hatasızlığı dikkat çekmekte; insan ve herhangi bir BDM tarafından üretilen metin arasında ayırım yapılmasını imkansız hale getirmektedir (Liu vd., 2023). BDM’ler arasında, daha önce de ifade edildiği üzere OpenAI ailesinin GPT modelleri önemli yer tutmaktadır. GPT modellerinin dışında Anthropic adlı şirket tarafından sunulan Claude ve Google tarafından üretilen Gemini (Bard) adlı dil modelleri de günümüzde dikkat çeken BDM’ler arasında yer almaktadır.

ChatGPT, OpenAI tarafından geliştirilen ve GPT dil modeline dayalı, metin tabanlı istemleri gerçekleştirme yeteneğine sahip gelişmiş bir robot (Liu vd.,2021; Kirmani, 2022) olarak tanımlanmaktadır. ChatGPT, “chatbot” olarak bilinen öncü yazılımların gelişmiş versiyonu olarak nitelendirilebilir. Chatbot’lar, yapay zekâ tabanlı öğrenme yöntemleri kullanılarak geliştirilen akıllı sistemlerdir (Taecharungroj, 2023). Chatbotlar, kullanıcıların sorgularını girdi olarak almakta; ses veya yazılı olarak yanıtlar oluşturmaktadır (Nithuna & Laseena, 2020; Thorat & Jadhav, 2020).Bu süreçte yapay zekâ desteğini kullanan chatbotlar, birçok alanda (bankacılık, müşteri hizmetleri vb.) endüstriyel olarak kullanılmaktadır.

Doğal dil işleme algoritmaları (NLP- Natural Language Processing) denildiğinde sadece GPT ailesinden bahsetmek yeterli olmayacaktır. Önceden eğitilmiş dil modelleri arasında BERT, GPT-

3 ve Alexa (Devlin vd., 2019; Floridi and Chiriatti, 2020) adlı algoritmalarından da bahsetmekte fayda vardır. BERT, çok çeşitli NLP görevlerinde yüksek performans elde etmiş bir modeldir ve birçok NLP görevi için kullanılmıştır (Xu vd., 2015; Miyato vd., 2016). OpenAI'nin GPT dil modelleri üzerine inşa edilen ChatGPT, bilinen arama motorlarının aksine bilgi derlemek için eğitilmiş veri setlerini kullanmaktadır (Rudolph vd., 2023: 4; Shah, 2022). Bu veri setleri daha önceden elde edilmiş ve işlenmiştir. Ayrıca kullanıcılardan elde edilen dönütlere göre algoritma kendini güncellemektedir. Susnjak (2022) tarafından yapılan çalışmada ChatGPT'nin üst düzey bilişsel görevleri gerçekleştirme ve metin üretme yeteneği değerlendirilmiştir. Araştırmaya göre, ChatGPT'nin eleştirel düşünme özelliği ortaya koyabildiği ve gerçekçi metinler üretebildiği sonucuna ulaşılmıştır.

BDM'lerin günlük hayatta kullanımıyla birlikte, bu algoritmaların performansı ile ilgili akademik araştırmaların arttığı görülmektedir. Aytekin ve Karabina tarafından yapılan bir çalışmada, GPT ailesinin Türkçe eş adlı kelimeler ve cümle içindeki anlamlarının doğruluğu tespit edilmeye çalışılmıştır. 200 eş adlı kelime üzerinden yapılan çalışmada, GPT ailesinin dilin anlamsallığı bağlamında gelişme gösterdiği, cümle içinde yer alan eş adlı kelimeleri ayırt edebildiği ve anlamlandırdığı görülmüştür (Aytekin & Karabina, 2024). Çalışmada elde edilen sonuçlara göre, büyük dil ailelerinin Türkçeyi doğru algılama kapasitesi her versiyonla birlikte artmaktadır.

Yapay zekâ tabanlı büyük dil algoritmalarının hızlı gelişimi, insanları bu algoritmaları hayatın her aşamasında kullanmaya yönlendirmektedir. Bu aşamaların bazıları olumlu olarak değerlendirilse dahi olumsuz olarak nitelenebilecek ve bireyin eğitim hayatına zarar verecek yaklaşımlar da ortaya çıkmaktadır. Bu çerçevede büyük dil modellerinin soru cevaplama potansiyeli, bir dolandırıcılık yöntemi veya haksız çıkar sağlama amaçlı olarak kullanılabilir. Deepfake gibi yapay zekâ algoritmalarının kişi hak ve hürriyetlerini aşındırıcı nitelikte kullanımına benzer bir durum, büyük dil modellerinde de ortaya çıkmış; özellikle eğitim sürecinde dil modelleri kullanılarak, yapay sahtecilik yolu açılmıştır. Çalışmanın ilerleyen kısmında “yapay zekâ sahteciliği” (AI cheating) kavramı detaylandırılacaktır.

Yapay zekâ algoritmalarının barındırdığı riskler sadece dolandırıcılık (deepfake) ve sahtecilik amaçlı kullanım ile sınırlı değildir. Kabak ve Kırbaş'ın yapmış olduğu “ChatGPT with Risks and Opportunities” başlıklı çalışmada, GPT ailesi özelinde yapay zekâ kullanımı ile ortaya çıkabilecek risklerin sınıflandırması yapılmıştır. Çalışmaya göre, yapay zekâ kullanımının yaygınlaşması ile birlikte karşılaşma ihtimalimiz olan riskler şunlardır: Bilgi eksikliği, hatalı varsayım, gizlilik sorunları, siber güvenlik tehdidi, önyargılı içerik, işten çıkarılma (iş imkanlarının azalması), intihal, aşırı bağımlılık, kısıtlı içerik, empati eksikliği, yaratıcılık eksikliği, ve bağlamsal sınırlamalar (Kabak & Kırbaş, 2023). Belirtilen riskler arasında çalışma açısından en önemli başlıklar intihal, yaratıcılık eksikliği, ve bağlamsal sınırlamalar olarak düşünülebilir. Bahsi geçen bu risklere Massaro'da değinmekte ve yapay zekâ kullanımı ile birlikte klasik eğitimin anlamsızlaşabileceğini (Massaro, 2023) iddia etmektedir.

1.3. Yapay Zekâ Sahteciliği ve Yapay Zekâ Kopyası

“AI cheating” ifadesi (Akbari, 2024; Massaro, 2023; Smith, 2024) son yıllarda literatüre girmiş bir tanımlama olarak karşımıza çıkmaktadır. Kavram Türkçeye “yapay zekâ kopyası” olarak çevrilebilir. Kavramın bu şekilde kullanılması ilerleyen çalışmalar açısından bir yol gösterici olarak değerlendirilebilir. “AI cheating” ifadesi, akademik kaynaklarda kullanılmamakla birlikte, daha çok konuyla ilgili haber veya gazete makalelerinde yer almaktadır. Yine Türkçe literatürde “yapay zekâ kopyası” şeklinde bir ifadeye rastlanmamıştır. Bu çerçevede kavramı tanımlamanın gerekliliği ortaya çıkmaktadır. Smith'e göre AI cheating (yapay zekâ kopyası), öğrencilerin gelişmiş bilgisayar programlarını kullanarak okul ödevlerini veya sınavlarını hileli bir şekilde tamamlarken, bu çalışmalarını kendilerine aitmiş gibi göstermeleridir (Smith, 2024). Sınav esnasında veya ödev yapılırken öğrenciler, soruları çözmek veya ilgili ödevi hazırlamak için yapay zekâ problem çözme araçlarını kullanabilmekte; bu durum, haksız bir avantaj sağlamaktadır. Yapay zekâ kopyasının eğitim sürecine etkisi, sadece haksız rekabete yol açması olarak değerlendirilmemelidir. Massaro, yapay zekâ kopyacılığının ABD üniversitelerinde

umutsuz ve onarılamaz bir bozulmaya yol açtığını ifade etmektedir. Massaro'ya göre haber medyasının dizginsiz eğlenceye dönüşmesi, internetin yanlış bilgilendirilmiş kişi ve dar görüşlerle dolu olması gibi nedenlerle, çevrimiçi beslenen öğrenciler bu dar görüşlülük sarmalına mahkum olmakta; ortaya çıkan eserler, bu darlık ve yanlışlık atmosferinden beslenmektedir. Bu durum, eğitimin asli amaçlarından olan eleştirel ve mantıksal düşünmeyi yok eden, körleştiren bir süreçtir (Massaro, 2023). Massaro, sadece hak ihlallerine değinmenin yanlış olacağını, yapay zekâ kopyacılığının bireylerin özgün düşünme kabiliyetini yok ettiğini ifade etmekte; bu durumun daha ciddi sonuçlar doğuracağını belirtmektedir. Yapay zekâ kopyacılığının eğitimi manipüle ettiği iddiası sadece Massaro ile sınırlı değildir. Bir yapay zekâ gözetmenlik teknolojisi şirketi olan Rosalyn'nin kurucusu ve CEO'su Nur Akbari'ye göre, yapay zekânın kontrolsüz kullanımı, ilerleyen süreçte eğitimi anlamsız hale getirebilir. Spor branşlarında yapılan doping gibi, yapay zekânın eğitimde kontrolsüz kullanımı da haksız rekabete yol açabilmektedir. Akbari, “yapay zekânın her şeyi yapacağı” şeklinde ortaya çıkan birey ve öğrenci yönelimlerinin, insanların eğitimden uzaklaşmasına neden olabileceğini ifade etmektedir (Akbari, 2024). Akbari ve Massaro'dan hareketle yapay zekâ kopyasının eğitimde haksız rekabete yol açan, diğer yandan eğitimin asli fonksiyonlarını yok ederek, bireylerin düşünme kapasitelerini sınırlayan etkilerinin olduğu söylenebilir.

Yapay zekâ algoritmaları eğitim öğretim süreçlerinde haksız rekabete yol açtığı gibi, diğer çevrimiçi alanlarda da bu haksız rekabetin varlığından bahsetmek mümkündür. Eğitimde yapay zekânın kullanımı yapay zekâ kopyası olarak adlandırılırken; diğer çevrimiçi alanlarda ortaya çıkan haksız rekabet, “yapay zekâ hilesi” ifadesi ile tanımlanmaktadır. Yapay zekâ hilesi, çevrimiçi oyunlar, yarışmalar gibi faaliyetlerde haksız bir avantaj elde etmek için yapay zekânın kullanılması şeklinde tanımlanabilir. Bu hile yönteminde yapay zekâ sistemlerinin hızlı öğrenme ve uyum sağlama yeteneği kullanılmakta ve bu durum, yapay zekâ algoritmalarını kullananları, rakiplerinden daha iyi performans gösterebilecek hale getirmektedir.

Yapay zekâ kopyası/hilesi, genel bir yaklaşımla “yapay zekâ sahteciliği” olarak sınıflandırılabilir ve eğitimden çevrimiçi oyunlara birçok alanda kendini gösterebilmektedir. Çevrimiçi oyunlarda kullanılan programlanmış botlar veya ödev hazırlama algoritmaları ile hazırlanan eğitim materyalleri bunlara örnek olarak verilebilir. Diğer yandan çalışmanın çerçevesini oluşturan çevrimiçi sınavlarda yöneltilen soruların cevaplanmasında, yapay zekâ algoritmaları ve özellikle dil modellerinin kullanılabilmesi düşünülmektedir. Bu kullanım, bireysel yanıtları ile sistemde bulunan diğer öğrencilerin haksızlığa maruz kalmasına yol açabilmektedir. Her ne kadar BDM ve yapay zekâ algoritmaları kısmen kapalı sistemler olsa da CoPilot ve Gemini gibi asistanların çevrimiçi arama yaparak geri dönüş sağlama, internet kaynaklarını kullanarak doğru yanıtı erişebilmeleri, büyük dil modelleri ve chatbotların çevrimiçi sınavlarda kullanımı ile ilgili soru işaretlerini beraberinde getirmektedir.

Özellikle çevrimiçi sınavlarda yapay zekâ kullanımının engellenmesi için bazı yenilikçi sistemlerden bahsetmek doğru olacaktır. Yapay zekâ kopyasının, çevrimiçi sınavlarda daha karmaşık sistemlerle ve yine çoğunlukla yapay zekâ kullanılarak önlenebileceği düşünülmektedir. Bu noktada “online proctoring” (çevrimiçi gözetmen) uygulamaları ön plana çıkmaktadır. Günümüzde yapay zekâ kopyacılığını çevrimiçi sınavlarda engellemek amacıyla üretilmiş Digixam, Ai-proctor, Mercer-Mettl, ProctorEdu, Respondus Monitor gibi yazılımlar dikkat çekmektedir. Bu yazılımlar, bir katılımcının çevrimiçi sınavda kopya çekmek için kullanabileceği tüm potansiyel yöntemleri (ekran paylaşımı, arkadaşların veya akrabaların yardımını almak, akıllı telefonlar, büyük dil modelleri, chatbotlar vb.) çevrimiçi gözetmenlik yaparak durdurmayı hedeflemektedir. Burada kişinin göz ve beden hareketlerinin takibi, sekmelerin kontrolü, ortamın dinlenmesi, ortamın gözetilmesi gibi birçok parametre devreye girmektedir.

2. Metodoloji

Çalışma iki aşamadan oluşmaktadır. Öncelikle yapay zekâ kopyası olarak tanımlanan kavramı açıklamak amacıyla literatüre başvurulmuştur. Alanın yeni olması ve kavramın daha yeni tartışılır olması nedeniyle geniş bir akademik çerçevenin varlığından söz etmek mümkün değildir. Mevcut

akademik çalışmaların (Abdelaal vd., 2019; Oravec, 2023) çoğunlukla yapay zekânın akademik amaçlı kullanımına dair sorunlara yoğunlaştığı söylenebilir. BDM’lerin çevrimiçi sınavlarda kopya amaçlı kullanımına ilişkin literatürde yeterli çalışma bulunmamaktadır. Bu nedenle BDM’lerin çevrimiçi sınavlarda sahtecilik amaçlı kullanım potansiyellerinin (doğru yanıt verme performansları) belirlenmesi ikincil amaç olarak seçilmiştir. Bu aşamada yapay zekâ tabanlı büyük dil modellerinin ön lisans ve lisans düzeyinde çevrimiçi sınavlardaki başarısının ölçülmesi hedeflenmiştir. Amaçlı örneklem tekniği çerçevesinde iletişim fakülteleri ve sosyal bilimler-teknik bilimler meslek yüksek okullarında eğitim veren akademisyenlerden çevrimiçi ölçme değerlendirme sürecinde kullandıkları sınav soruları talep edilmiştir. Veri paylaşmayı kabul eden akademisyenlerden 17 derse ait çevrimiçi sınav içeriği dijital olarak temin edilmiştir. Bu sınavlarda yer alan 173 adet soru seçilmiştir. Açık uçlu sorular çalışmada kullanılmamış, sadece çoktan seçmeli sorular değerlendirilmiştir. Seçilen sorular tek tek incelenmiş ve iki kategoriye ayrılmıştır. İlk kategori, salt bilgiye dayalı ve basit bir web taraması ile cevapları bulunabilecek sorulardır. İkinci kategori ise yorum gerektiren, birden fazla konu veya bilginin kıyaslanması gereken, yoruma dayalı sorulardır. Kategorize edilen sorular, 2024 yılı Ocak-Şubat aylarında yaygın kullanımı olan ChatGPT-3.5, Microsoft Bing (GPT-4), Google Bard, Opera Aria, LLaMA-2-7b, LLaMA-2-13b, LLaMA-2-70b, GPT-3.5-Turbo adlı modellerin chat botlarına yöneltilerek, yanıt vermeleri istenmiştir. Yapılan internet taraması sonucunda 2024 yılı itibariyle GPT-3, GPT-4, Cohere, Falcon, LLaMA, Claude, BLOOM, LaMDA, PaLM, BERT, DeepMind, Vicuña, Alpaca, Databricks, LLaMA-2-7b-32k (TogetherAI), Meta AI, Mistral AI, Stability AI, Gemini, StableLM, LongLLaMA, Xgen-7b adlı BDM’lerin aktif olduğu görülmüştür. Bu modellerin tamamı yerine, amaçlı örneklem tekniği çerçevesinde sekiz modele ilgili sorular yöneltilmiştir. İlerleyen çalışmalarda daha fazla dil modelinin kullanılması tavsiye edilebilir. Araştırma devam ederken Google Bard, “Gemini” (9 Şubat 2024); Microsoft Bing ise “Copilot” adını alarak varlığına devam etmiştir. Burada Microsoft Bing ve Opera Aria adlı botların yine OpenAI tarafından geliştirilen GPT-3.5 veya GPT-4 dil modellerini kullandığını hatırlatmakta fayda vardır. Chat botlar yoluyla dil modellerinin yanıtlaması istenilen sorulara verilen yanıtlar, öğretmenlerden gelen cevaplarla karşılaştırılmış, dil modellerinin doğru ve yanlış cevapları kaydedilmiştir. Kaydedilen bu cevaplar belirtilen iki kategoride sınıflandırılmış ve bir durum analizi yapılmıştır.

3. Bulgular ve Yorum

Çalışmada metodoloji kısmında belirtildiği üzere 173 adet soru üzerinden hareket edilmiştir. 8 farklı Chatbot (CB) uygulamasına sorular yöneltilmiştir. Bu CB’lerin bazıları (Aria, Bing, GPT-3.5 Turbo) GPT tabanlı uygulamalardır. Versiyon ve özellikler değişmekle birlikte çoğunlukla aynı büyük dil modelinin farklı versiyonlarını kullanmaktadır. CB/BDM’lere cevapları basit internet taraması ile bulunabilecek 158 soru yönlendirilmiştir. Sorulara verilen cevapların doğrulukları ve frekansları Tablo 1’ de yer almaktadır.

Tablo 1: Bilgiye dayalı sorulara CB/BDM’ler tarafından verilen yanıtların doğru yanlış dağılımı

	ChatGPT 3.5		Microsoft Bing		Google Bard		Opera Aria		LLaMA-2- 7b		LLaMA-2- 13b		LLaMA-2- 70b		GPT-3.5- Turbo	
D/f	90	%56, 9	12	%80, 7	10	%67, 7	86	%54, 4	52	%32, 9	55	%34, 8	46	%29, 1	98	%62
Y/f	68	%43, 1	31	%19, 7	51	%32, 3	72	%45, 6	10	%67, 6	10	%65, 3	11	%70, 2	60	%38, 9
Top	15	%10	15	%10	15	%10	15	%10	15	%10	15	%10	15	%10	15	%10
.	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0

Tablo 1’de görüldüğü üzere CB/BDM’ye yöneltilen ve internet taraması sonucu cevaplarına ulaşılabilen (yoruma dayanmayan, bilgiye dayalı) sorulara en fazla doğru yanıtı Microsoft Bing CB’si vermiştir. Bing veya yeni adıyla “Copilot” soruların %80,3 (n:127)’ünü doğru, %19,7 (n:31)’sini yanlış cevaplamıştır. Bing CB’den sonra en fazla doğru cevap Google Bard tarafından verilmiştir. Bard chatbotu soruların %67,7 (n:107) sini doğru yanıtlamıştır. En az doğru yanıt veren BDM/CB ise LLaMA ve türevleri olmuştur.

Tablo 2: Bilgi ve yoruma dayalı sorulara CB/BDM’ler tarafından verilen yanıtların doğru / yanlış dağılımı

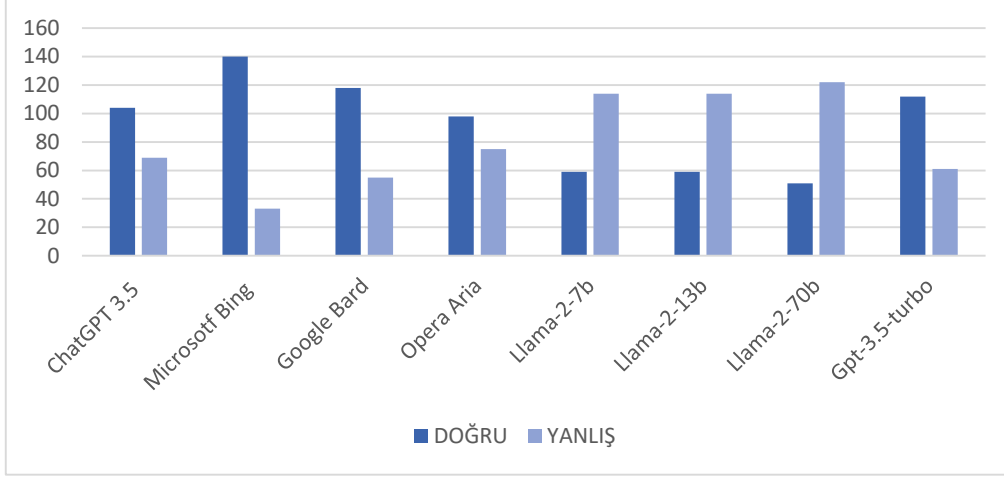
	ChatGPT 3.5		Microsoft Bing		Google Bard		Opera Aria		LLaMA-2- 7b		LLaMA-2- 13b		LLaMA-2- 70b		GPT-3.5- Turbo	
D/f	14	%93,3	13	%86,6	11	%73,3	12	%80	7	%46,6	4	%26,6	5	%33,3	14	%93,3
Y/f	1	%6,7	2	%13,4	4	%26,7	3	%20	8	%53,4	11	%73,4	10	%66,7	1	%6,7
Top.	15	%100	15	%100	15	%100	15	%100	15	%100	15	%100	15	%100	15	%100

Tablo 2’de CB/BDM’ne yöneltilen ve internet taraması sonucu cevaplarına ulaşılmakta zorlanılan (yoruma dayanan, bilgi ve yoruma dayalı) sorulara en fazla doğru yanıtı %93,3 (n:14) oranla GPT-3.5 Turbo ve GPT-3.5 vermiştir. GPT-3.5 Turbo’dan sonra Microsoft Bing chatbotu gelmektedir. Bing veya yeni adıyla “Copilot” soruların %86,6’sını (n:13) doğru cevaplamıştır. En az doğru yanıt veren BDM/CB ise LLaMA ve türevleri olmuştur. Bu noktada yoruma dayalı soruların (örneklem) kısıtlı olması nedeniyle, sonuçların genel bir çıkarım yapmaktan uzak olduğu düşünülebilir.

Tablo 3: Tüm sorulara Chat bot /BDM’ler tarafından verilen yanıtların doğru yanlış dağılımı

	ChatGPT 3.5		Microsoft Bing		Google Bard		Opera Aria		LLaMA-2- 7b		LLaMA-2- 13b		LLaMA-2- 70b		GPT-3.5- Turbo	
D/f	10	%60,	14	%80,	11	%68,	98	%56,	59	%34,	59	%34,	51	%29,	11	%64,
	4	1	0	9	8	2		6		1		1		5	2	7
Y/f	69	%39,	33	%19,	55	%31,	75	%43,	11	%65,	11	%65,	12	%70,	61	%35,
		9		1		8		4	4	9	4	9	2	5		3
Top	17	%10	17	%10	17	%10	17	%10	17	%10	17	%10	17	%10	17	%10
	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0

Tablo 3’te CB/BDM’lere yöneltilen tüm soruların sonuçları yer almaktadır. Bu verilere göre en başarılı CB/BDM, %80,9 doğru yanıt vermesi nedeniyle Microsoft Bing chatbotudur. Yeni ve güncel adıyla Copilot, toplam 173 sorunun 140’ına doğru yanıt vermiştir. Bilgiye dayalı ve yoruma dayalı sorulara verdiği cevapların oranı birbirine yakındır. Diğer yandan en başarılı dil modeli, “LLaMA” ve türevleri (LLaMA-2-7b, LLaMA-2-13b, LLaMA-2-70) olarak belirlenmiştir. Microsoft Bing içinde yer alan chatbot, Microsoft kaynaklarına göre GPT-4 tabanlıdır. GPT-3’e nazaran güncel internet kaynaklarına ulaşabilmesi nedeniyle daha isabetli sonuçlar verdiği düşünülmektedir.



Grafik 1: Chatbot /BDM'lerin yöneltilen soruları yanıtlama performansı

Grafik 1'de tüm CB/BDM karşılaştırılarak bir performans tablosu oluşturulmuştur. Grafikte görüldüğü üzere doğru yanlış oranlarında en iyi performans Microsoft Bing'e aittir. Bu durum bazı verilerle açıklanabilir. Microsoft Bing içinde yer alan chatbot (Copilot) Microsoft kaynaklarına göre GPT-4 tabanlıdır. Bing botunun GPT-3 veya 3.5'e nazaran güncel internet kaynaklarına ulaşabilmesi nedeniyle daha isabetli sonuçlar verdiği düşünülmektedir. GPT-3 ve 3.5'in gerçek zamanlı internete erişimi olmadığı bilinmektedir. ChatGPT-3 ve 3.5'i yayınlayan OpenAI'ye göre bu iki dil ailesi, 2022'nin başlarına kadar olan verilere dayanarak çalışmaktadır.

Sonuç

Yapay zekâ kopyası (AI cheating) kavramı, yapay zekâ algoritmalarının günlük hayatın bir parçası haline gelmesi ve özellikle eğitim alanında sahtecilik amaçlı kullanılması nedeniyle ortaya çıkmıştır. Kavram, bütüncül bir yaklaşımla eğitim ve ölçme değerlendirme sürecinin tamamı ile ilintilidir. Ortak yaklaşımlara göre, makine öğrenmesi temelli yapay zekâ algoritmaları olan büyük dil modelleri, çoğunlukla ödev, proje, makale yazımı ve benzer akademik içeriklerin oluşturulmasında kullanılmaktadır. Yapay zekâ algoritmaları yoluyla ortaya çıkan sahtecilik eğilimleri, eğitim sürecinde haksız rekabete yol açtığı gibi, eğitimin temel amaçları çerçevesinde değerlendirilen, analitik düşünme yetisinin öğrenciler tarafından kazanılmasının önünde önemli bir engel olarak düşünülmektedir. Ödev, proje, makale yazımı gibi süreçlerin dışında, yapay zekâ ve büyük dil modellerinin en fazla etkilediği alanın, ölçme ve değerlendirme süreci olduğu söylenebilir. Ölçme ve değerlendirme sürecinde yapay zekâ kopyacılığı için çoğunlukla BDM'ler kullanılmaktadır.

Çalışma kapsamında yapay zekâ, yapay zekâ kopyası ve BDM'lerin tanımı yapılmış, seçilen BDM'ler üzerinden, dil modellerinin uzaktan eğitim sürecinde kullanılan çevrimiçi soruları cevaplama performansları sorgulanmıştır. BDM'lerin çevrimiçi sınav performansları farklı üniversitelerde verilen 17 farklı derse ait sınav soruları üzerinden değerlendirilmiştir. Toplamda 212 soru içinden 173 soru belirlenmiş, 8 farklı CB'ye bu sorular yöneltilmiştir. Toplamda 1384 sorgulama gerçekleştirilmiştir. İlerleyen süreçte daha büyük veri setleri ve daha fazla insan kaynağı ile kapsamlı çalışmalar yapılması önerilmektedir. Elde edilen sonuçlar şu şekilde sıralanabilir:

CB ve BDM'ler her ne kadar yeni bir teknoloji olsalar ve gelişmeye devam etseler de mevcut altyapıları ile (lisans ve ön lisans seviyesinde) uzaktan eğitim sürecinde ölçme ve değerlendirme

amaçlı yöneltilen soruların büyük kısmına doğru yanıt verecek potansiyeli barındırmaktadır. Mevcut çalışmada Microsoft Bing (Copilot) adlı asistan GPT-4 dil modelinden aldığı destekle, sorulan soruların %80'inden fazlasına doğru yanıt vermiştir. Veriler çerçevesinde BDM'lerin çevrimiçi sınavlarda başarılı oldukları görülmüştür.

İnternette arama yapma özellikleri aktif olan BDM ve botlar, daha doğru sonuçlar sağlamaktadır. Yanlış verilen cevapların, farklı kaynaklardan elde edilen bilgilerin çatışmasından kaynaklandığı düşünülmektedir. Burada internette yer alan bilgilerin doğruluğu ile ilgili tartışmalar akla gelmektedir.

ChatGPT, Copilot ve Bard (Gemini) gibi BDM ve CB'lerin başarısının arkasında, büyük verilere ulaşım sağlayabilmeleri ve bunları işleyebilecek işlem gücüne sahip olmaları gibi etkenlerin olduğu düşünülmektedir. GPT-3, 3.5, 4 dil modelleri OpenAI ve Microsoft; Bard (Gemini) ise Google tarafından desteklenmektedir. LLaMA ailesi BDM'lerin ise hem işlem gücü hem de big dataya ulaşmada yaşadıkları sorunlarla birlikte gelişiminin yavaş devam ettiği görülmektedir.

Botların, BDM desteğiyle yüksek doğrulukta soru çözebilme potansiyeli ve bu algoritmaların çevrimiçi sınavlarda kullanılma olasılığı, özellikle online sınavların geçerliliği ve güvenilirliği ile ilgili soru işaretleri doğurmaktadır. Anlamli bir gözetim ve kontrol mekanizması olmaksızın yapılacak online sınavlarda yapay zekânın kullanımı haksız rekabete yol açabilecektir. Bu rekabetin önlenmesi için bazı önlemlerin alınması düşünülebilir. Burada öncelikli olarak eğitim sürecinde sorulan soruların “salt bilgiye dayalı olmaması” gerektiği sonucuna ulaşılmaktadır. İnternet yoluyla yapılacak temel aramalarla sonucu ortaya çıkan soruların kullanımının kısıtlanması, bunun yerine öğrencinin sözel veya yazılı olarak cevap verebileceği soru-cevap tiplerinin tavsiye edilmesi düşünülebilir.

Ölçme ve değerlendirme sürecinde sınavların çevrimiçi yapılması durumunda, eğitim kurumları “kamera açtırma” veya “görsel ortam kontrolü” gibi yöntemlere başvurmaktadır. Bu yöntemlerin yetersiz ve etkisiz olduğu düşünülmektedir. Çevrimiçi sınavlar söz konusu olduğunda gelişmiş yapay zekâ araçları soruları çözerek veya doğrudan yanıtlar sağlayarak, gerçek zamanlı yardım sunabilmektedir. Ayrıca bu yapay zekâ araçlarından bazıları ekran paylaşımını veya uzaktan gözetimi yöneterek, yapılan sahteciliğin tespit edilmesini zorlaştırabilmektedir. Bu noktada yapay zekâ sahteciliğinin önlenmesi için yine yapay zekâ araçlarına başvurulması ve bu alanda özelleştirilmiş uygulamaların kullanılması gerekliliği ortaya çıkmaktadır. Çalışmanın alanyazın kısmında değinilen “online proctoring” (çevrimiçi gözetmen) adı verilen uygulamaların çevrimiçi sınav sürecinde kullanılması tavsiye edilebilir.

Yapay zekânın hayatımıza birçok yenilik getirdiği şüphesizdir. İlerleyen süreçte yapay zekânın gelişim ve büyük dil modellerinin yaygınlaşması ile birlikte bu yeniliklerin artacağı düşünülebilir. Diğer yandan bu gelişimin yeni açmazlara yol açtığı ve bu açmazlara karşı yeni perspektiflerin geliştirilmesinin gerektiği ortadadır. Klasik eğitim ve sınav sistemi ile devam etmenin yerine özellikle yapay zekânın kullanımının desteklenmesi ve daha akılcı eğitim ve ölçme değerlendirme sistemlerinin oluşturulması kaçınılmaz durmaktadır.

Kaynakça

Abdelaal, E., Mills, J., & Walpita Gamage, S. (2019). Artificial Intelligence Is a Tool for Cheating Academic Integrity.

- Akbari, N. (2024, Şubat 28). The AI Cheating Crisis: Education Needs Its Anti-Doping Movement. Education Week. <https://www.edweek.org/technology/opinion-the-ai-cheating-crisis-education-needs-its-anti-doping-movement/2024/02>
- Andrade, I. M. D., & Tumelero, C. (2022). Increasing customer service efficiency through artificial intelligence chatbot. *Revista de Gestão*, 29(3), 238-251. <https://doi.org/10.1108/REGE-07-2021-0120>
- Asaro, P. M. (2012). A body to kick, but still no soul to damn: Legal perspectives on robotics. P. Lin, K. Abney, G. A. Bekey (Ed), *Robot ethics: The ethical and social implication of robotics* (169-186 ss.). Massachusetts: MIT Press.
- Aytekin, Ç., & Karabina, T. B. (2024). Chatgpt'nin Farklı Büyük Dil Modelleri Performanslarının Türkçedeki Eş Adlı Kelimeler Üzerinden İncelenmesi. *İstanbul Aydın Üniversitesi Sosyal Bilimler Dergisi*, 16(3), Article 3.
- Beer, D. (2017). The social power of algorithms. *Information, Communication & Society*, 20(1), 1–13.
- Birhane, A., Kasirzadeh, A., Leslie, D. & Watcher, S. (2023). Science in the age of large language models. *Nature Reviews Physics* 5, 277–280
- Campa, R. (2020). Fourth Industrial Revolution and Emotional Intelligence: A Conceptual and Scientometric Analysis, *Changing Societies & Personalities*, 4(1), 8–30. DOI: 10.15826/csp.2020.4.1.087
- Cellan-Jones, R. (2014). Stephen Hawking Warns Artificial Intelligence Could End Mankind. BBC News. <https://www.bbc.com/news/technology-30290540>
- Chomsky, N., Roberts, I., & Watumull, J. (2023). The false promise of ChatGPT. *New York Times*. <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>
- Devlin J., Chang M.W., Lee K., Toutanova K. (2019) BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, Minnesota. Association for Computational Linguistics.
- Doğan, A. (2002). Yapay Zekâ. Ankara. Kariyer
- Floridi L., Chiriatti M. (2020). Gpt-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30:681–694.
- Goodfellow I.J, Shlens J., Szegedy C. (2014). Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572.
- Hayawi, K., Shahriar, S., & Mathew, S. S. (2023). The imitation game: Detecting human and ai-generated texts in the era of large language models. arXiv preprint arXiv:2307.12166.
- Johnson, P. (2017). 99 Facts On The Future Of Business in The Digital Economy, SAP: [https://www.slideshare.net/sap/99-facts-on-the-future-of-business-in-the-digital-economy- \[01.07.2020\]](https://www.slideshare.net/sap/99-facts-on-the-future-of-business-in-the-digital-economy- [01.07.2020]).
- Kabak, T., & Kirbaş, İ. (2023). Chatgpt With Risks And Opportunities. *Mehmet Akif Ersoy Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 14(2), 365-376. <https://doi.org/10.29048/makufebed.1271477>
- Kirmani, A.R. (2022). Artificial intelligence-enabled science poetry, *ACS Energy Letters*, Vol. 8, pp. 574-576. <https://doi.org/10.1021/acseenergylett.2c02758>

- Kozma, R., Alippi, C., Choe, Y. ve Morabito, F. C. (Eds.). (2018). Artificial intelligence in the age of neural networks and brain computing. United States: Academic Press.
- Kurian, N., Cherian, J. M., Sudharson, N. A., Varghese, K. G., & Wadhwa, S. (2023). AI is now everywhere. *British Dental Journal*, 234(2), 72-72. <https://doi.org/10.1038/s41415-023-5461-1>
- Liu, X., Zheng, Y., Du, Z., Ding, M., Qian, Y., Yang, Z., & Tang, J. (2021). GPT understands, too. arXiv preprint arXiv:2103.10385.
- Lee, E. (2023). Is ChatGPT a false promise?. Berkeley Blog. <https://blogs.berkeley.edu/2023/03/19/is-chatgpt-a-false-promise/>
- Levy, D. (2012). The ethics of robot prostitutes. P. Lin, K. Abney, G. A. Bekey (Ed), *Robot ethics: The ethical and social implication of robotics* (223-232 ss.). Massachusetts: MIT Press.
- Liu, Z., Yao, Z., Li, F., & Luo, B. (2023). Check me if you can: Detecting ChatGPT-generated academic writing using CheckGPT. arXiv preprint arXiv:2306.05524.
- Lokhorst, G.A. & van den Hoven, J. (2012). Responsibility for military robots. P. Lin, K. Abney, G. A. Bekey (Ed), *Robot ethics: The ethical and social implication of robotics* (145-156 ss.). Massachusetts: MIT Press.
- Mark Massaro, O. C. (2023, Ağustos 23). AI cheating is hopelessly, irreparably corrupting US higher education [Text]. The Hill. <https://thehill.com/opinion/education/4162766-ai-cheating-has-hopelessly-irreparably-corrupted-us-higher-education/>
- McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. (1955). A proposal for Dartmouth Summer Research Project on Artificial Intelligence. <http://wwwformal.stanford.edu/jmc/history/dartmouth.pdf>. Accessed 13 January 2023.
- Miyato T., Dai A.M., Goodfellow I. (2016). Adversarial training methods for semi-supervised text classification. arXiv preprint arXiv:1605.07725.
- Nithuna, S. & Laseena, C. (2020). Review on implementation techniques of chatbot. In *Proceedings of the 2020 International Conference on Communication and Signal Processing (ICCSP)*, Chennai, India, 28–30 July 2020; pp. 157–161.
- OpenAI. (2015). OpenAI. <https://openai.com/about/> (Erişim Tarihi: 17.04.2024).
- Oravec, J. A. (2023). Artificial Intelligence Implications for Academic Cheating: Expanding the Dimensions of Responsible Human-AI Collaboration with ChatGPT and Bard. *Jl. of Interactive Learning Research*, 2(34), 213-237.
- Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Massachusetts: Harvard University Press.
- Pool, C. R. (1997). Brain-based learning and students. *The Education Digest*, 63(3), 10.
- Rudolph, J., Tan, S. & Tan, S. (2023), Chatgpt: Bullshit Spewer Or The End Of Traditional Assessments In Higher Education? *Journal Of Applied Learning & Teaching*, 6(1), 1-22.
- Russell, S., & Norvig, P. (2021). *Artificial intelligence: a modern approach* (3rd ed). Prentice Hall.
- See, A., Pappu, A., Saxena, R., Yerukola, A., & Manning, C. D. (2019). Do massively pretrained language models make better storytellers?. arXiv preprint arXiv:1909.10705.

- Shah, C. (2022), The Rise Of Aı Chat Agents And The Discourse With Dilettantes. *Information Matters*, 2 (12). <https://Informationmatters.Org/2022/12/The-Rise-Of-Aichat-Agents-And-The-Discourse-WithDilettantes/>
- Sharkey, N. (2012). Killing made easy. P. Lin, K. Abney, G. A. Bekey (Ed). *Robot ethics: The ethical and social implication of robotics* (111-128 ss.). Massachusetts: MIT Press.
- Sheikh, S. (2020). *Understanding the Role of Artificial Intelligence and Its Future Social Impact*, IGI Global
- Smith, A. (2024). AI Cheating: Tips to Avoid and Detect. <https://screenapp.io/blog/how-to-avoid-and-detect-ai-cheating-with-exam-assignments-and-essays>
- Susnjak, T. (2022). ChatGPT: The end of online exam integrity? Preprint. ArXiv:2212.09292v1. <https://doi.org/10.48550/arXiv.2212.09292>
- Taecharunroj, V. (2023). What can ChatGPT do? Analyzing early reactions to the innovative AI chatbot on Twitter. *Big Data and Cognitive Computing*, 7(1),35. <https://doi.org/10.3390/bdcc7010035>
- Thorat, S.A. & Jadhav, V. (2020). A review on implementation issues of rule-based chatbot systems. In *Proceedings of the International Conference on Innovative Computing & Communications (ICICC)*; SSRN: Rochester, NY, USA, 2020. <http://dx.doi.org/10.2139/ssrn.3567047>
- Xu Y., Chen L., Wei J., Ananiadou S., Fan Y., Qian Y, Chang E.I., Tsujii J. (2015). Bilingual term alignment from comparable corpora in english discharge summary and chinese discharge summary. *BMC Bioinform.*, 16:149:1–149:10.

Çatışma beyanı

Makalenin yazarı, bu çalışma ile ilgili taraf olabilecek herhangi bir kişi ya da finansal kuruluş ile ilişkisi bulunmadığını dolayısıyla herhangi bir çıkar çatışmasının olmadığını beyan eder.

Destek ve teşekkür

Çalışmada herhangi bir kurum ya da kuruluştan destek alınmamıştır.