

Antik Efsanelerden Üretken Yapay Zekâya: Gerçeklik, Dezenformasyon ve İnsanlık Mücadelesi

→ Şebnem ÖZDEMİR*

İnsanlığın en eski rüyası yapay zekâ, antik Yunan efsanelerinde/kaynaklarında, eski Çin ve eski Türk yazıtlarından, insanlığın temel mirası masallara kadar her alanda kendini göstermiştir. İlk geliştirildiği zamanlarda nasıl öğreneceği bilinmeyen makine (yapay zekâ), düşünce değişikliği ile veriden öğrenme yolculuğuna başlamıştır. Ancak “veri temelli makineler” geliştiricilere yeterli gelmemiş, veri olsun ya da olmasın insan gibi düşünebilen, insan gibi karar verebilen makineler yaratma hayali yükseliş göstermiştir. Çünkü yapay zekâ bir mühendislik disiplini olmaktan ötedir. Yapay zekâ, insanlığı taklit etme ve onu aşma sanatıdır (Özdemir, 2021). Kasım 2022’den itibaren başladığı düşünülen, önce üretici sonra üretken yapay zekâlar olarak adlandırılan süreç, insanlığın yapay zekâ yolculuğunda yeni bir dönemi başlatmıştır. Ancak bu başlayan yeni dönem, yapay zekânın tekil görev bazlı, verinin gösterdiği yolda ilerlediği değil, bir zekâ olarak var olduğu adımlar silsilesini de kapsamaktadır.

İnsanlık, geliştiricilerin en temel hayali yapay genel zekâ (artificial general intelligence – AGI) ile 2025 yılında buluşacak, 3. tür olarak, inorganik varlıkların, var olma ve yer edinme çabasına şahit olacaktır. Bu durum Hollywood filmlerinin de etkisi ile korkunç, yok edici olarak nitelense de daha korkunç ve yok edici bir süreç ile insanlık çok daha uzun süredir mücadele hâlinindedir. Bu mücadele, gerçeğe ulaşmanın, gerçek bilgiyi elde etmenin mücadelesidir.

Gerçeğin net duruşu, bilinen tarihin başlangıcından beri o ya da bu şekilde darbe almıştır. Antik çağlarda uygulanmış olan Damnatio Memoriae, kişilerin, bazen de o kişilerin ailelerinin, miraslarının yani buldukları dönem-

* Doç. Dr., İstinye Üniversitesi İktisadi, İdari ve Sosyal Bilimler Fakültesi Yönetim Bilişim Sistemleri Bölümü, sebnem.ozdemir@istinye.edu.tr

deki varlıklarının tarih sahnesinden silinmesini ifade etmektedir (Sijpesteijn, 1974; Flower, 1998; Huët, 2004). Bu şekilde gerçeğin bütünsel olarak yok edilmesi dışında, gerçeğin kirletilmiş, farklı formlarının oluşturularak yanlış bilgi hâline dönüştürülmüş, bir propaganda aracı olarak kullanılmış ve dezenformasyon hâli ile tarih sahnesinde gözlenmiştir. Örneğin Roma'da Antonius ile Kleopatra'nın ilişkisi karşısında, Octavian, Antonius'a karşı onun itibarını zedelemeyi amaçlı bilgiler paylaşmıştır. Dönemin madeni paraları üzerinde kazanmış kısa ve keskin bilgilerle Antonius, Kleopatra tarafından yozlaştırılmış ve onun kuklası hâline gelmiş biri olarak aktarılmıştır (Kaminska, 2017). Milattan önce 44. yıla ait bu olaydan sonra bir başka ürkütücü olay 1835 yılında gerçekleşmiştir. The New York Sun gazetesinin The Great Moon Hoax haberi ile en temel haber alma kanalının, bir dezenformasyon, bir sahte haber yayma aracı hâline gelişi resmî kayıtlara geçmiştir. Gazete, astronot Sir John Herschel'in ayda yaşayan bir ırkı keşfetmesine dair gerçeklikle alakası olmayan altı makale yayınlamıştır (Andrews, 2015).

90'lı yıllardan itibaren internetin yaygınlaşması, dezenformasyon tarafında bir patlama etkisine neden olmuş olsa da bu etkinin büyümesine asıl katkı önce sosyal medya platformlarının ortaya çıkışı, ardından da akıllı telefonlarla her yerde ve her zaman bağlantıda kalma hâli ile oluşmuştur. Bireyler, haber alma kanalı olarak sosyal medyayı görmeye başladıklarında, bu durum olaylara anlık erişebilme gücü verse de dezenformasyon tarafında da büyük bir oyun alanı oluşturmuştur. Bu sayede etkileşimi yüksek dezenformasyon, gerçek bilgi karşısında dünyayı 9 kat daha hızlı dolaşmıştır. Ancak insanlık, bu kaos ortamında bile gerçeği bulabilmek için herhangi ek bir araç kullanmadan, farklı kaynakları okuyup teyit ederek, eleştirel düşünerek ve temel zihinsel süreçlerini devreye sokarak yolunu bulabiliyordu.

Kasım 2022'nin son gecesinde OpenAI firmasının ChatGPT'yi çıkarması ile üretici-üretken yapay zekâ dönemi herkesin erişebileceği ve deneyim kazanabileceği yapay zekâ ile etkileşim dönemini başlatmıştır (Park ve Gelles-Watnick, 2023). Başta çok eğlenceli gibi duran bu aracın ortaya çıkışı dezenformasyon tarafını da beslemektedir. Özellikle Midjourney, Dall-E, Stable Difusion, Flux, Runway, Sora, Kling ve Genimate gibi teknolojilerin görsel üretme, görseli canlandırma kabiliyeti farklı haber ve içeriklerin oluşturulmasında çeşitli olanaklar sağlamaktadır.

Ocak 2023 tarihinde görsel üreten makinelerde insan oluşturmak oldukça popüler hâle gelmiştir. İlk denemelerde, elde edilen insan görüntülerinde, benzer insan yüzlerinin çiziliyor olması, gözlerdeki odaklanma sorunu, diş sayısının fazlaca oluşturulması, parmak sayısının yüksek oluşu, ellerin duruşundaki çizim problemi heves kırıcı olsa da tespit etme tarafını kolaylaştırmaktadır.



Görsel 1. Ocak 2023 Midjourney ile Üretilen Görseller

1 Şubat 2023 tarihinde Midjourney uygulamasında oluşturulan görsellerde ilk önce farklı yüz üretmemet sorunun artık yavaşta olsa ortadan kalkmaya başlamıştır. Nitekim şubat ayında görüntü üreten makine girilen “iki güzelle bir mekânda çılgınca eğlenen Trump” promptunun çıktısında kadınların simalarında farklılaşmanın sağlandığı görülmektedir. Ancak bu çıktı, ilk bakışta, kadınların ellerinin doğru modellenememesi, yüz orantısındaki problemler ile yapay zekâ ile yapıldığı yine rahatlıkla tespit edilebilir durumdadır.



Görsel 2. 1 Şubat 2023 Midjourney Çıktısı

Mart 2023'e geldiğinde yapay zekânın Trump görselini üretmedeki başarısı da değişmiştir.



Görsel 3. Üretken Yapay Zekâ Trump Tutuklandı Çıktısı

Yaklaşık bir ayda makinenin gerçeğine yakın görsel üretme başarısındaki gelişme, dezenformasyon üretimi tarafını beslemiştir. Ancak Mart 2023'teki görseller dikkatli incelendiğinde kaval kemiklerinin modellenmesi, bacak uzunluğundaki oran, kavuşmuş ellerdeki hatalar, dikkatli gözler tarafından hâlâ ayırt edilebilmektedir.

Görsel üreten makinelere 2023 yılının ilk çeyreğinde istenilen görselin yüklenerek üzerinde değişiklik yapabilme özelliği gelmiştir. Bu özelliğin en can alıcı örneklerinden biri, 6 Şubat 2023 yılında ülkemizde gerçekleşen deprem sonrasında paylaşılan bazı görsellerle kendini göstermiştir.

Bir X gazetesi tarafından soldaki resim “Enkazda Kur’an-ı Kerim bulan Uzak Doğulu kişi Müslüman oldu” haberiyle servis edilmiştir. Bir başka Y gazetesi, haberin görsellerinin yapay zekâ ile üretilerek halkın kandırıldığını iddia etmiş ve sağdaki resmi paylaşmıştır.



Görsel 4. X ve Y Gazeteleri Tarafından Paylaşılan Deprem ile İlgili Haber Görselleri

Haberin doğruluğundan ziyade görseller incelendiğinde, X gazetesi tarafından verilen görselin (sol) orijinal fakat görüntü kalitesi düşük olduğu anlaşılmaktadır. Y gazetesinin verdiği görsel (sağ), ilk resim yapay zekâyâ yüklenerek elde edilmiştir. Soldaki resimdeki görevlinin atkısı, sağdaki resimde makine tarafından sakal ile atkı arasında bir şeye dönüştürülmüştür. Sağdaki resimde Kur’an-ı Kerim’in kavrandığı elde temsiliyet hatası bulunmaktadır. Yine sağdaki resimde arka planda yer alan tabeladaki yazılar okunaksızdır, kelimeler oluşmamaktadır. Bu durum, o tarihlerdeki görsel üreten makinelere ortak bir sorundur.

Görsel üreten makinelerdeki gelişme devam ederken, ses üretiminde de üretken yapay zekâlar aynı yıllarda kendini göstermeye başlamıştır. BBC News, Sudan iç savaşına yönelik olarak üst düzey bir yöneticinin sesinin modellen-diği olay (Goodman ve Hashim, 2023), Alman haber kuruluşu Tagesschau tarafından, yayınlarının çarpıtılmasına neden olan modellenmiş sesler (Reveland ve Siggelkow, 2023) ilk kritik örneklerdir. 2024 yılına gelindiğinde, özellikle yılın ilk çeyreğinde ses üretimine yönelik üretken yapay zekâlarda yükseliş başlamış, modellemede başarımları artmıştır. Bu durum, özellikle konuşmanın sentezlenmesinin daha da kolay hale getirmiş, kötüye kullanım tarafında sıradan vatandaşları bile etkileyecek kritik vakaların yaşanmasına neden olmuştur. Nitekim 2024 yılı biterken, FBI yayınladığı uyarı raporunda, aile bireylerinin kendi aralarında fiziksel olarak toplandıklarında bir parolanın belirlenmesi, parolayı bir kâğıda yazarak saklamaları, online görüşmeler ve telefon görüşmelerinde önce bu parolayı vermeleri/istemeleri gerektiği ifade edilmiştir.

2024 yılı biterken görüntü üreten yapay zekâlarda aşağıdaki sorunlar düzelmiş, dezenformasyon yaymada başarısı yüksek içerikler üretilebilir hâle gelmiştir.

Ellerin oluşturulması, bir nesneyi kavrama durumu yılın başındaki problemlerden arındırılmıştır (ellere dikkat ederken artık, parmakların açıklığına, tırnakların tüm parmaklarda aynı şekilde temsiliyetine dikkat edilmelidir. Bir insanın bir başka insanın belinden kavramasında da orantı problemi gözlenebilir).

Yazıların taklidi, harflerin oluşturulması, kelime yazımındaki karmaşa ve anlamsızlık giderilmiştir. Bir uyarı tablosu, açıklama tabelası, mağaza tabelası artık kolaylıkla oluşturulmaktadır. Bu nedenle yazılar üzerinden varsa anlamsal ve mantıksal sorunlarına bakılmalı, lokasyon ile uyarı/mağaza tabelasının uyuma durumu kontrol edilmelidir.

Dişlerdeki sayısal problem (32 dişten fazla diş çizimi) giderilmiştir, bu nedenle sayısal olarak kontrolü işe yaramamaktadır. Ancak açık şekilde ağız ve diş görselinin olduğu çıktılarda, diş yüzeylerine detaylı bakılmalıdır. Diş yüzeyi bir estetik müdahale söz konusu değilse kısmı ölçüde pürüzlüdür. Diş minesini de incelenmeli, keskin ve düz bir çizgi şeklinde modellenenlerden şüphe edilmelidir.

Gözlerin odaklanmasındaki sorunlar da ortadan kaldırılmıştır. Bu nedenle artık göz uzayına bakılmalıdır. Bir nesneye/kişiye/duruma bakan kişinin her iki gözünün göz uzayında (gözünün içindeki yansıma) benzerlikler oluşmaktadır. Benzer şekilde gözün alt kirpiklerinin sıralanması, sıklığı, yerleşimi, göz pınarının modellenme biçiminin kontrolü de kritiktir.

İster metin ister ses ister görüntü olarak elde edilsin, yapay zekâ tarafından oluşturulan çıktıların dezenformasyon oluşturmadaki kolaylaştırıcı etkisi büyüktür. Bu etkinin tespitine yönelik olarak çeşitli çalışmalar bulunmaktadır (Jawahar ve diğerleri, 2020; Fröhling ve Zubiaga, 2021; Stiff ve Johansson, 2022; Uchendu ve diğerleri, 2023; Kumarage ve diğerleri, 2023). Çalışmaların ilk yıllarında görüntü ve ses tarafında tespiti daha kolay, metin tarafında ise bireyin entelektüel seviyesi ile alakalı bir tespit yeteneği gözlenirken, 2025 yılı itibarıyla çıktı ne olursa olsun salt insan duyuları ve zihinsel kabiliyetlerle ile tespit artık mümkün değildir.

Kaynakça

- Andrews, E. (2015). The Great Moon Hoax, The History Channel. Accessed 06/01/25: <http://www.history.com/news/the-great-moon-hoax-180-years-ago?-linkId=16545579>
- Fröhling L. & Zubiaga, A. (April 6, 2021). 'Feature-based detection of automated language models: tackling GPT-2, GPT-3 and Grover'. Peer J Computer Science. Vol. 7. E443.
- Goodman, J. & Hashim, M. (October 5, 2023). 'AI: Voice cloning tech emerges in Sudan çivil war'. BBC.
- Harriet I. Flower; Rethinking "Damnatio Memoriae": The Case of Cn. Calpurnius Piso Pater in AD 20. Classical Antiquity 1 October 1998; 17 (2): 155–187. doi: <https://doi.org/10.2307/25011081>
- Huet, V. (2004). IMAGES ET "DAMNATIO MEMORIAE." Cahiers Du Centre Gustave Glotz, 15, 237–253. <http://www.jstor.org/stable/24359195>
- Jawahar, G., Abdul-Mageed, M. & Lakshmanan, L. (2020). 'Automatic Detection of Machine Generated Text: A Critical Survey'. In Proceedings of the 28th International Conference on Computational Linguistics, pages 2296–2309, Barcelona, Spain (Online). International Committee on Computational Linguistics
- Kaminska, I. (2017). A module in fake news from the info-wars of ancient Rome. Financial Times. Accessed 12/01/25: <https://www.ft.com/content/aaf2bb08-dca2-11e6-86ac-f253db7791c6>
- Kumarage, T., Garland, J., Bhattacharjee, A., Trapeznikov, K., Ruston, S. & Liu, H. (2023). 'Stylometric Detection of AI-Generated Text in Twitter Timelines'. ArXiv

- Park, E. & Gelles-Watnick, R. (August 28, 2023). 'Most Americans haven't used ChatGPT; few think it will have a major impact on their job'. Pew Research Center.
- Reveland, C. & Siggelkow, P. (November 13, 2023). 'False Tagesschau audio files in circulation'. Tagesschau
- Sijpesteijn, P. J. (1974). Macrinus' Damnatio memoriae und die Papyri. *Zeitschrift Für Papyrologie Und Epigraphik*, 13, 219–227. <http://www.jstor.org/stable/20180617>
- Stiff, H. & Johansson, F. (2022). 'Detecting computer-generated disinformation'. *International Journal of Data Science Analytics*, vol 13, pp 363–383.
- Uchendu, A., Le, T. & Lee, D. (July 05, 2023). 'Attribution and Obfuscation of Neural Text Authorship: A Data Mining Perspective'. *ACM SIGKDD Explorations Newsletter*, Volume 25, Issue 1, pp 1–18.