Giriş

Yapay zeka (YZ) teknolojileri, günümüz dünyasında giderek daha fazla yer almakta ve toplumsal, ekonomik ve kültürel alanlarda derin etkiler yaratmaktadır. Bu teknolojilerin gelişimi ve yaygınlaşması, beraberinde önemli etik soruları da getirmektedir. Yapay zeka etiği, bu teknolojilerin tasarımı, geliştirilmesi ve kullanımı sırasında ortaya çıkan etik sorunları ele alan disiplinlerarası bir alandır.

Yapay Zeka Etiğinin Temel Konuları

1. İstatistiksel Adalet ve Eşitlik Sorunları

Yapay zeka sistemlerinde adalet kavramı, genellikle istatistiksel adalet çerçevesinde ele alınmaktadır. Ancak bu yaklaşım, adaletin eşitlikle karıştırılması gibi temel hatalar içerebilmektedir. İstatistiksel adalet yaklaşımı, benzer durumların benzer şekilde ele alınması prensibine dayanır, ancak bu durum adaletin daha geniş ve karmaşık doğasını yeterince yansıtmayabilir.

Ayrıca, önceden tanımlanmış sosyal gruplar içinde çalışan adalet anlayışı, adaletin kendisinin grup kompozisyonunu belirlemesine izin vermek yerine, mevcut grupları veri olarak kabul eder. Bu durum, adalet kavramının Aristo'dan bu yana gelen geleneksel anlamından uzaklaşmasına neden olabilir.

2. Sorumlu Yapay Zeka Çerçeveleri

Sorumlu Yapay Zeka (Responsible AI - RAI), yapay zeka kullanımıyla ilişkili etik ilkelerin standart çerçevelerle uyumlu hale getirilmesini amaçlar. Günümüzde, etik standartlar ve RAI uygulamaları genellikle birbirinden ayrılmış durumdadır, bu da her endüstrinin yapay zekayı etik olarak kullanmak için kendi standartlarını izlemesine yol açmaktadır.

Küresel firmalar ve devlet kuruluşları, ortak ve standart bir çerçeve tasarlamak için gerekli girişimlerde bulunmaktadır. Sosyal baskı ve yapay zekanın etik olmayan kullanımı, RAI'nin uygulanmasından ziyade tasarımını zorlamaktadır.

3. Büyük Dil Modelleri ve Eğitim Etiği

ChatGPT gibi büyük dil modellerinin (LLM) ortaya çıkışı, akademik ve eğitim topluluklarında endişe ve hareketlilik yaratmıştır. Bazıları, bu araçların insan benzeri metin üretme yeteneğini bilgi erişimi ve bilgisayar destekli öğrenmenin altın çağı olarak görürken, diğerleri bu durumun eşi görülmemiş düzeyde akademik dürüstlüzlük ve kopya çekmeye yol açabileceğinden endişe duymaktadır.

Yapılan araştırmalar, ChatGPT'nin piyasaya sürülmesinin, öğrenci denemelerinin hem uzunluğunda hem de stilinde önemli değişikliklerle aynı zamana denk geldiğini göstermektedir. Bu durum, akademik yayıncılık gibi diğer bağlamlarda da gözlemlenen değişiklikleri yansıtmaktadır.

4. Yapay Zeka Halüsinasyonları ve Yaratıcılık

Büyük dil modellerindeki halüsinasyonlar (gerçek olmayan bilgiler üretme) genellikle hatalar olarak kabul edilir. Ancak yaratıcı veya keşif bağlamlarında, bu "hatalar" inovasyon için beklenmedik yollar sunabilir.

"Purposefully Induced Psychosis" (PIP) gibi yaklaşımlar, spekülatif kurgu, interaktif hikaye anlatımı ve karma gerçeklik simülasyonları gibi yaratıcı görevler için LLM halüsinasyonlarını artırmayı amaçlar. Bu yaklaşım, halüsinasyonları bir kusur değil, bir hesaplamalı hayal gücü kaynağı olarak yeniden çerçevelendirir.

5. Model Güncellemeleri ve Şeffaflık Sorunları

Makine öğrenimi (ML) sistemleri, zaman içinde veri setindeki değişimler nedeniyle performans düşüşüne karşı savunmasızdır. Bu sorunu çözmek için, ML sistemlerinin düzenli olarak güncellenmesi önerilir.

Ancak model güncellemeleri, ML destekli karar verme sürecine "güncelleme opasitesi" adı verilen yeni bir opasite (şeffaflık eksikliği) türü getirir. Bu durum, kullanıcıların bir güncellemenin bir ML sisteminin mantığını veya davranışını nasıl veya neden değiştirdiğini anlayamaması durumunda ortaya çıkar. Bu tür bir opasite, ML'deki kara kutu sorununa yönelik mevcut çözümlerin büyük ölçüde ele almakta yetersiz kaldığı çeşitli ayırt edici epistemik ve güvenlik endişeleri sunar.

Yapay Zeka Etiğinde Güncel Tartışmalar

Veri Gizliliği ve Güvenliği

Yapay zeka sistemleri, büyük miktarda veri üzerinde eğitilir ve çalışır. Bu verilerin toplanması, işlenmesi ve saklanması sırasında gizlilik ve güvenlik sorunları ortaya çıkabilir. Kişisel verilerin korunması, veri sahibinin rızası ve veri güvenliği, yapay zeka etiğinin önemli konularıdır.

Algoritmik Önyargı ve Ayrımcılık

Yapay zeka sistemleri, eğitildikleri veriler üzerinden öğrenirler. Eğer bu veriler önyargılı ise, yapay zeka sistemleri de bu önyargıları öğrenebilir ve kararlarında yansıtabilir. Bu durum, cinsiyet, ırk, etnik köken veya diğer korunan özelliklere dayalı ayrımcılığa yol açabilir.

Şeffaflık ve Açıklanabilirlik

Yapay zeka sistemlerinin nasıl karar verdiğini anlamak, özellikle derin öğrenme gibi karmaşık modeller söz konusu olduğunda zor olabilir. Şeffaflık ve açıklanabilirlik, yapay zeka sistemlerinin kararlarının anlaşılabilir ve denetlenebilir olmasını sağlamak için önemlidir.

Sorumluluk ve Hesap Verebilirlik

Yapay zeka sistemlerinin kararları sonucunda ortaya çıkan zararlardan kim sorumludur? Bu soru, yapay zeka etiğinin en zorlu sorularından biridir. Sorumluluk ve hesap verebilirlik, yapay zeka sistemlerinin tasarımcıları, geliştiricileri, kullanıcıları ve hatta sistemlerin kendisi arasında nasıl dağıtılmalıdır?

Sonuç

Yapay zeka etiği, teknolojinin hızla geliştiği ve toplumsal etkilerinin derinleştiği bir dönemde giderek daha önemli hale gelmektedir. Etik ilkelerin ve standartların geliştirilmesi, yapay zeka teknolojilerinin insanlığın yararına kullanılmasını sağlamak için kritik öneme sahiptir.

Yapay zeka etiği alanındaki araştırmalar ve tartışmalar, teknolojinin gelişimiyle birlikte devam edecektir. Bu alanda disiplinlerarası işbirliği, farklı perspektiflerin ve uzmanlıkların bir araya getirilmesi, etik sorunların kapsamlı bir şekilde ele alınması için önemlidir.

Kaynakça

1. Brusseau, J. (2025). Four Bottomless Errors and the Collapse of Statistical Fairness. arXiv:2504.13790v1.
2. Gadekallu, T. R., Dev, K., Khowaja, S. A., Wang, W., Feng, H., Fang, K., Pandya, S., & Wang, W. (2025). Framework, Standards, Applications and Best practices of Responsible AI: A Comprehensive Survey. arXiv:2504.13979v1.
3. Leppänen, L., Aunimo, L., Hellas, A., Nurminen, J. K., & Mannila, L. (2025). How Large Language Models Are Changing MOOC Essay Answers: A Comparison of Pre- and Post-LLM Responses. arXiv:2504.13038v1.
4. Pilcher, K., & Tütüncü, E. K. (2025). Purposefully Induced Psychosis (PIP): Embracing Hallucination as Imagination in Large Language Models. arXiv:2504.12012v1.
5. Hatherley, J. (2025). A moving target in AI-assisted decision-making: Dataset shift, model updating, and the problem of update opacity. arXiv:2504.05210v1.