

Atıf (Cite as): Yenigün Altın, Ş., & Taşdelen, H. (2025). Yapay Zekâ Modellerinde Cinsiyet Önyargıları: ChatGPT ve Copilot Üzerine Nicel Bir İnceleme. *Akdeniz İletişim*, (49), 310-331. <https://doi.org/10.31123/akil.1629516>

Yapay Zekâ Modellerinde Cinsiyet Ön Yargıları: ChatGPT ve Copilot Üzerine Nicel Bir İnceleme

Gender Biases in Artificial Intelligence Models: A Quantitative Study on ChatGPT and Copilot

Şule YENİGÜN ALTIN¹²

Hacer TAŞDELEN³

Öz

Bu çalışma, günümüzde sağlıktan istihdama kadar çok çeşitli alanlarda kullanılan üretken yapay zekâ (YZ) araçlarının, bilgi üretirken cinsiyet konusunda ön yargılı bir tutum sergileyip sergilemediğini ortaya çıkarmayı amaçlamaktadır. Çalışmanın örneklemini, amaçlı örnekleme yöntemi kullanılarak görsel üretme özellikleriyle de öne çıkan YZ araçlarından ChatGPT ve Copilot olarak belirlenmiştir. Bu YZ araçlarından on dokuz meslek ve yirmi olumlu kişilik özelliğinin görselinin istemi yapılarak, üretilen meslek ve kişilik özelliklerinin görsellerinin en çok hangi cinsiyet ile ilişkilendirildiği analiz edilmiştir. Verilerin analizi içerik analizi tekniği ile gerçekleştirilmiş ve görsellerin cinsiyetlere göre dağılımları incelenmiştir. Çalışma sonucunda elde edilen bulgularda; erkek cinsiyetinin sorun çözen, teknik beceri gerektiren ve yüksek statü içeren mesleklerle ilişkilendirildiği, kadın cinsiyetinin ise sanat ve estetik yönü ağır basan az sayıda meslekle özdeşleştirildiği görülmüştür. Olumlu kişilik özellikleriyle ilgili bulgularda da benzer şekilde birçok özelliğin temsili erkek cinsiyetiyle ilişkilendirilmiştir. Sonuç olarak örnekleme yer alan YZ araçlarının erkek yanlı bir yaklaşımla çalıştığı ve ortaya koyduğu sonuçlarda erkek yanlı kabullerin yeniden üretimine katkı sunduğu anlaşılmıştır.

Anahtar Kelimeler: Yapay Zekâ, Cinsiyet Ön Yargısı, İçerik Analizi, ChatGPT, Copilot

Abstract

This study aims to reveal whether generative artificial intelligence (AI) tools used in a wide range of fields from health to employment today exhibit a gender-biased attitude while producing information. The sample of the study was determined as ChatGPT and Copilot, which are AI tools that also stand out with their visual production features, using the purposeful sampling method. Nineteen professions and twenty positive personality traits were requested from these AI tools, and the gender with which the generated profession and personality traits were most associated was analyzed. The data was analyzed with the content analysis technique and the distribution of the visuals by gender was examined. In the findings obtained as a result of the study; it was seen that the male gender was associated with problem-solving, technically skilled and high-status professions, while the female gender was identified with a small number of professions that were predominantly artistic and aesthetic. Similarly, in the findings regarding positive personality traits, the representation of many traits was associated with the male gender. As a result, it was understood that the AI tools in the sample worked with a male-biased approach and contributed to the reproduction of male-biased assumptions in the results they produced.

Keywords: Artificial Intelligence, Gender Bias, Content Analysis, ChatGPT, Copilot

¹ Dr. Öğr. Üyesi, Bingöl Üniversitesi, Sosyal Bilimler Meslek Yüksekokulu, Pazarlama ve Reklamcılık Bölümü, syenigun@bingol.edu.tr, ORCID: 0000-0002-5637-8737

² Sorumlu Yazar (Corresponding Author)

³ Dr., Bağımsız Araştırmacı, hacertasdelen@gmail.com, ORCID: 0000-0002-8122-0560

Giriş

Yapay zekâ (YZ) sistemlerinin eğitildikleri veri setlerindeki ön yargılı durumları yansıtmaları etik sorunlara yol açmaktadır. Özellikle son zamanlarda YZ kullanımının yaygınlaşmasıyla birlikte, bu sistemlerin karar süreçlerinde ön yargı barındırdığına dair tartışmalara akademik çalışmalarda sıkça yer verilmektedir. YZ teknolojileriyle ilgili araştırmalar, yalnızca teknik ilerlemelere odaklanmayıp bu sistemlerin etik ve toplumsal etkilerinin de derinlemesine araştırılmasını kapsamaktadır. Yaygın olarak kullanılan ve karmaşık bir yapıya sahip olan makine öğrenimi ve YZ destekli sistemlerin ayrıntılarını ortaya koymak ve bu sistemler tarafından alınan otomatik kararların eşitlikçi ve ön yargısız olup olmadığını göstermek genellikle güç bir durumdur. Toplumların sahip olduğu olumlu veya olumsuz deneyimler zaman içerisinde veriler aracılığıyla sistemlere aktarılmış ve aktarılmaya devam etmektedir. Bu veriler aracılığıyla eğitilen YZ sistemlerinin, kaynak materyallerinden aldığı negatif yargıları taşıdığı ve bunun sonucunda cinsiyet, ırk, yaş gibi niteliklere dayalı çeşitli ön yargıları yansıttığı görülebilmektedir (Shrestha ve Das, 2022).

YZ teknolojisi, geline son noktada, insanların günlük hayattaki çalışmalarından sağlık alanına, karar alma mekanizmalarından hukuk sistemine kadar birçok alanda kendini göstermektedir. Etki ve yaygınlık durumuna bakıldığında hızlı ilerleyiş gösteren bu teknoloji, beraberinde -adil erişim ve ön yargı başta olmak üzere- birçok endişeyi de taşımaktadır. YZ teknolojilerinin özellikle sağlık, istihdam, ceza adaleti, kredi puanlaması gibi çok çeşitli alanlarda kullanımının yaygınlaşması bu teknolojilerin önemini arttırmaktadır. Bu modeller, haksız sonuçlara yol açabilmekte, üretilmiş verilerdeki bireylerin temsilini etkileyen üretken ön yargılar dâhil olmak üzere mevcut eşitsizlikleri sürdürmekte ve eşitsizlikleri teşvik eden mevcut sosyo-kültürel ayrımcılıkları pekiştirebilmektedir (Angwin vd., 2022; Cirillo vd., 2020; Kaplan vd., 2024; Silberg ve Manyika, 2019; West vd., 2019). YZ'nin ön yargılı davranarak toplumun bazı kesimlerine karşı ayrımcılık yapma riski bu konuyla ilgili en büyük tehditlerden biri olarak gösterilmektedir (Ferrara, 2024). Makine zekâsı öncelikle kendisine sunulan verileri gözlemleyerek öğrenmektedir. Bir YZ'nin büyük miktarda veriyi işleme yeteneği bu sorunu bir ölçüde hafifletebilmekte; ancak, kullanılan veriler cinsiyetle ilgili basmakalıp kavramlar içeriyorsa YZ bu ön yargıları sürdürebilmektedir. Örneğin; YZ teknolojilerinin tasarımında erkeklerin aşırı temsil edilmesi cinsiyet eşitliğiyle ilgili geçmişten bugüne kadar atılmış adımların boşa gitmesine neden olabilmektedir (Leavy vd., 2020). YZ'nin kurucu ve araştırmacılarının cinsiyetinin genellikle erkek olması başlangıcından bu yana YZ'nin gelişim yolunu şekillendirmede önemli bir rol oynamaktadır. Araştırmalar, YZ modellerinin eğitildiği materyallerdeki ön yargılarla dolu verilere olan bağımlılığının, toplumda var olan ön yargıları pekiştirdiğini hatta daha kötü sonuçlara yol açabileceğini göstermektedir (Parsheera, 2018).

Hagerty ve Rubinov (2019), YZ ve etik ile ilgili çalışmalarında ilk olarak YZ'nin kültürel ortama bağlı olarak belirgin şekilde farklı sosyal etkilere sahip olma olasılığının yüksek olduğunu öne sürmüşlerdir. YZ'ye ilişkin algılar ve anlayışlar büyük olasılıkla yerel, kültürel ve sosyal bağlamlar tarafından derinden şekillendirilmektedir. İkinci olarak düşük ve orta gelirli ülkelerin YZ'nin olumsuz sosyal etkilerine karşı daha savunmasız olabileceğini ve olumlu sonuçlardan yararlanma olasılıklarının daha düşük olabileceğini savunmuşlardır. Bu durum sosyal eşitsizliği ve sosyal istikrarsızlığı artırabilecek tüm toplumları riske atabilecek ve potansiyel olarak çok uzak jeopolitik sonuçlara yol açabilecektir. Bu nedenle YZ sistemlerinin sosyal eşitsizliği nerede artırabileceğini ve zararların nasıl azaltılabileceğini anlamak için titiz, bağımsız araştırmalara ihtiyaç olduğu belirtilmiştir. Adalet, eşitlik, insan hakları gibi ilkeler, genellikle bilinçli ve

bilinçsiz ön yargılar nedeniyle pratiğe dönüştürülememektedir. Bu nedenle makinelerin insan ön yargılarını güçlendirmek yerine ortadan kaldırması ve YZ'nin toplumda daha fazla adalet sağlama hedefleriyle ilerlemesi için çaba göstermek gerekmektedir. Bunun için hükümetler, şirketler, üniversiteler; bilgisayar, sosyal bilim ve veri bilimi alanlarında çalışan araştırmacılar dâhil olmak üzere çeşitli paydaşlar arasında iş birliği gerekmektedir (Parsheera, 2018).

Dijital ayrımcılıkla ilgili yeterli terminolojinin olmaması ve gizli ön yargının olası varlığı, toplumda var olan ön yargıları normalleştirerek maskeleyebilmektedir. Dolayısıyla veri analitiğinde ayrımcılık sorunlarını çözmek için algoritmik şeffaflığı iyileştirmek için çalışmalar yapılması gerekmektedir (Favaretto vd., 2019). Prates vd. (2020), YZ modellerinin eğitildiği veri setlerinin toplumsal normlar ve kültürel değerler tarafından şekillendirilmiş bilgiler içermesinden dolayı bu veri setlerinin analiz edilerek toplumdaki basmakalıp cinsiyet rollerine dair sonuçlar elde etmenin mümkün olabileceğini ifade etmiştir. Buradan hareketle bu çalışmada, YZ araçlarının bilgi üretirken ön yargılı davranma riski cinsiyet bağlamından incelenmeye çalışılmıştır. Üretken YZ araçlarının günlük kullanım pratiklerinde ortaya koyduğu çıktılarda cinsiyete dair yaklaşımını anlamak ise çalışmanın temel amacını oluşturmuştur.

Araştırmanın amacına bağlı olarak aşağıdaki sorulara yanıt aranmaktadır:

S1: Üretken YZ araçları cinsiyet noktasında ön yargılı bir tutum sergilemekte midir?

S1a: Üretken YZ araçları, meslekler ile cinsiyet arasında ilişkilendirme yaparken ön yargılı davranmakta mıdır?

S1b: Üretken YZ araçları, olumlu kişilik özellikleri ile cinsiyet arasında ilişkilendirme yaparken ön yargılı davranmakta mıdır?

Çalışmada YZ araçları tarafından üretilen görsellerin cinsiyetle ilişkisi, meslek ve olumlu kişilik özelliklerine ait görseller şeklinde araştırma değişkenleri oluşturularak ele alınmıştır. Belirlenen meslekler ve olumlu kişilik özelliklerine ait üretilen görsellerin hangi cinsiyetle ilişkilendirileceği merakıyla gelişen çalışma; örneklem olarak seçilen üretken YZ araçlarının eğitimsel düşünme mantığını, cinsiyet konusundaki tavrını ve ön yargıları ne şekilde yansıttığını anlama noktasında yol gösterebilmektedir. Teknoloji ve cinsiyet ekseninden yapılan birçok çalışma, farklı dönemlerde kullanılan araçların erkek yanlı algoritmalara sahip olduğunu, yerleşik kalıp yargıları devam ettirdiğini ortaya koymuştur. Literatür bölümünde detaylı şekilde yer verilen bu çalışmalardan alınan ilhamla, bu araştırmada cinsiyet meselesi, üretken YZ araçlarının görsel oluşturma özellikleri aracılığıyla ürettiği görüntüler üzerinden irdelenmiştir. Üretken YZ araçlarının kullanıcı etkileşimiyle de gelişimlerini devam ettirdikleri göz önüne alındığında, çalışma sonucunda ortaya konacak tablonun durum saptama görevi taşıyacağı ve teknolojiyi üretenler ile kullanıcılar arasında yeni iletişim kurulumları bu noktaya dair farkındalık sağlayabileceği düşünülmektedir. Türkçe literatürde YZ ile ilgili yapılan çalışmalar içerisinde cinsiyet odaklı bir çalışma bulunmaması sebebiyle, çalışmanın literatüre özgün bir katkı sunması beklenmektedir.

1. Yapay Zekâ ve Ön Yargı

Teknolojik her kırılma anı beraberinde iyimser ve kötümser başta olmak üzere farklılaşan görüş ayrılıklarını getirmektedir. Kitle iletişim araçlarının gelişmesiyle birlikte teknolojinin toplumsal boyutunun da düşünülmesi gerektiği fikri, her yeni teknolojik gelişmede yeniden kendini göstermektedir (Güngör, 2018, s. 181). 21. yüzyıl özelinden bakıldığında bu kırılma anını yaşatan gelişme YZ teknolojisi olmuştur. Günlük yaşam pratiklerinde sağladığı kolaylık ile birlikte her geçen gün kullanım alanları artan YZ araçları, bazı çevreler tarafından kutsanırken

bazı grupların da endişe duymasına neden olmuştur. YZ teknolojilerini elektrik ile benzer tutan bilgisayar bilimci Andrew Ng, iyimser bakış açısı sergileyenlerden biri olarak YZ'nin tıpkı elektriğin zamanında her şeyi dönüştürdüğü gibi dönüştürücü etkiye sahip olduğunu dile getirmiştir (Lynch, 2017). YZ teknolojilerine endişe ile yaklaşanlar ise YZ araçlarının veri gizliliğini ihlal etme potansiyeline, ön yargıları sürdürme ve yeni ön yargılar oluşturma durumlarına odaklanmaktadır (Bliuc vd., 2018; Chan vd., 2023; Cyfirma, 2023; Rozado, 2023; Wu, 2022).

1950'li yıllardan bu yana düşüncesi beliren ve adını literatürde ilk olarak 1956 yılında gerçekleştirilen Dartmouth Konferansı'nda (McCarthy vd., 2006) duyuran YZ teknolojisi; insanlar gibi kararlar alabilen, problemleri çözebilen, öğrenebilen ve öğrendiklerine uyum sağlayabilen bilgisayar sistemlerinin geliştirilmesini içermektedir (Russell ve Norvig, 2010, ss. 1-2). İnsan zekâsının birtakım yönlerini taklit eden YZ teknolojisi, günümüzde kat ettiği yol sonucunda büyük veri kümelerinden öğrenme, tahmin yapma ve bağlamsal adaptasyon gibi bilişsel yeteneklere sahip durumdadır. YZ söz konusu olduğunda tanımsal düzeyde bilinmese de birçok kullanıcının bu teknolojiye entegre olmasını sağlayan araçlar üretken YZ kapsamında yer almaktadır. Üretken YZ, öncül verilerden öğrenilen kalıplara ve yapılara dayalı bir şekilde metinler, görseller, modeller, videolar ve kodlar oluşturabilen YZ sınıfını ifade etmektedir (Naumova, 2023). Örnek vermek gerekirse belirli istemler yazarak iletişime geçilen ChatGPT, Gemini, Copilot ve Claude gibi bilinen birçok sohbet odaklı ara yüzler, üretken YZ modelleri kapsamında yer almaktadır.

Üretken YZ modellerinin günlük yaşama kolay şekilde entegre olabilmesi ve işleyiş bazında kullanıcı dostu olması her geçen gün bu modellere olan ilgiyi arttırmaktadır. Öte yandan teknolojinin güdümündeki bu modellere yönelik kaygılar da beraberinde artış göstermekte ve bu kaygılar transformer mimarisine dayanan birçok sohbet ara yüzünün; güvenilirlik, doğruluk, ön yargı, genelleme, açıklanabilirlik, hesap verebilirlik, fikri mülkiyet, kontrol gerekliliği, gerçek zamanlılık, enerji tüketimi, güvenlik ve gizlilik endişeleri şeklinde (Crawford vd., 2023; Liebrez vd., 2023; Ray, 2023, ss. 140-141) dezavantajlı yönlerinin açığa çıkmasını sağlamaktadır. YZ araçları tarafından ortaya konan çıktıların her zaman doğru olmaması yanlış bilgilerin yayılmasına neden olabilmektedir. Bu sebeple, üretken YZ araçları ile elde edilen bilgilerin kontrol edilmesi gerekmektedir. Günümüzde ChatGPT ve Gemini gibi çok kullanılan sohbet ara yüzlerinin "hata yapabileceklerine" dair uyarılara yer vermesi, kullanıcıların bu noktada eleştirel düşünme kabiliyetlerine sahip olmalarını gerektirmektedir. Ön yargı, gizlilik ve güvenlik şeklindeki dezavantajlarıyla ciddi endişelere neden olan YZ araçları, manipülasyona açık durumdadır. Bu duruma, Microsoft tarafından 2016 yılında geliştirilen sohbet botu TAY örnek gösterilebilmektedir. Sosyal paylaşım ağı X –eski adıyla Twitter- platformunda insan davranışlarını analiz ederek paylaşımlarda bulunan bu sohbet botu, insanlardan fazla etkilenmesi sonucunda olumsuz ve tahrik edici paylaşımlarda bulunmaya başlamıştır (Wolf vd., 2017). Soykırımı desteklediği şeklinde bir yanıt veren TAY, kullanıma sunulduktan on altı saat sonra geri çekilmiştir (Wakefield, 2016).

Ön yargı esaslı endişelere bakıldığında ise aslında bu durumun doğrudan YZ teknolojisine özgü olduğunu söylemek yanlış olacaktır. Friedman ve arkadaşları tarafından 1996 yılında gerçekleştirilen çalışma, bilişim teknolojilerinin getirdiği ön yargılara odaklanmıştır. Bilişim teknolojilerinin ortaya çıkardığı ön yargıları sınıflandıran yazarlar; 1) önceden var olan, 2) teknik nedenlere dayanan ve 3) yeni ortaya çıkan şeklinde üçlü bir tipoloji üzerinde durmuştur. Bu bağlamda, YZ ve ön yargı çerçevesinde önem taşıyan önceden var olan ön yargının kökleri toplumsal kurumlara, uygulamalara ve tutumlara kadar uzanmaktadır. Önceden var olan ön

yargının oluşmasında bilişim sistemini tasarlayan bireylerden kaynaklanan ön yargılar ve toplumun genelinden kaynaklanan ön yargılar etkili olabilmektedir (Friedman ve Nissenbaum, 1996, ss. 333-334). YZ modelleri de, kaynak materyaller ile eğitilen ve yapay öğrenme metodları ile geliştirilen araçlar oldukları için bireysel ve toplumsal ön yargıları devam ettirme, yayma ve yeniden üretme potansiyeline sahip durumdadır (Bowman, 2023; Liang vd., 2021; Lippens, 2024). Toplumsal düzeyde teknolojik ayrıma tabi tutulacak kitleler üretebilecek bu teknoloji; kültürel, dilsel, cinsiyet, ırk, algoritmik ve ideolojik başta olmak üzere birçok ön yargı biçimini barındırmaktadır (Crawford ve Paglen, 2021; Ray, 2023, ss. 147-148). Irksal ön yargının örneğini oluşturan bir çalışmada, kişilerin suç işleme potansiyellerini gösteren bir YZ aracının siyah tenli insanları, beyaz tenli insanlara oranla daha çok suça meyilli hâlde gösterdiği görülmüştür (McKenna, 2019). YZ araçlarından GPT-3 modeli kullanılarak yapılan bir kelime ilişkilendirme çalışmasında, YZ modelinin şiddet ve terörizm ile ilgili olumsuz çağrışım yaratan kelimeler ile Müslümanları ilişkilendirdiği sonucuna ulaşılmıştır (Abid vd., 2021). YZ araçları tarafından ortaya konan ön yargılı çıktıların etkisine dair araştırma yapan Glickman ve Shanot (2024), bu araçlar ile etkileşim içinde olan kişilerin ön yargıyı içselleştirebileceklerini, ön yargılı çıktıya dair kanaatlerini değiştirebileceklerini ortaya koymuştur. Ön yargılı sonuçların kartopu etkisi yaratacak potansiyelde olduğunu ortaya koyan araştırmacılar, bu sebeple YZ araçlarının tarafsız çıktılar üretebilmesinin önemi üzerinde durmuştur. Etkileşim söz konusu olduğunda insan etkisini de göz önünde bulundurmak gerekmektedir. Kullanıcılardan aldığı istemler ve dönütler ile gelişim sürecini devam ettiren üretken YZ modelleri; ön yargılı, toksik ve manipüle edici istemler sonucunda yeni ön yargılar edinebilmektedir (Sharma, 2024).

Bir sonraki başlıkta örneklendirilecek cinsiyet ön yargısı ile birlikte toplumda “görünmez azınlıklar” (Pasquale, 2015, s. 38) yaratma ihtimali bulunan YZ araçlarının, ön yargı sorunları ile baş edebilmenin önemli yolunun, bu sorunları daha çok dile getirmekten geçtiği düşünülmektedir. Kaynak materyallerden, yazılımı geliştiren kişilerden önceden edinilmiş ön yargılar her ne kadar sistematik şekilde devam ediyor olsa da yeni ön yargıların oluşmaması ve teknik temsiliyet sorunlarının düzeltilmesi adına farkındalığı arttırmaya katkı sunacak türden çalışmaların daha çok gerçekleştirilmesi gerekmektedir.

2. Yapay Zekâ ve Cinsiyet Ön Yargısıyla İlgili Tartışmalar

Son yıllarda YZ teknolojisindeki olumlu gelişmelere rağmen bu sistemlerin hem eğitildikleri verilerde hem de çıktılarına uygulanan denetleme politikalarında örtük ön yargılara sahip oldukları, genellikle toplumsal eşitsizlikleri yansıttıkları ve bu eşitsizlikleri sürdürdükleri görülmektedir (Balestri, 2024, s. 65). YZ bağlamında ön yargı; veri toplama, algoritma tasarımı ve insan yorumlaması dahil olmak üzere çeşitli kaynaklardan ortaya çıkabilmektedir. Makine öğrenimi modelleri kendilerini eğitmek için kullanılan verilerde bulunan önyargıyı öğrenerek çoğaltabilmekte ve bu şekilde adil olmayan sonuçlara yol açabilmektedir. YZ’deki ön yargıyı belirlemek ve ele almak bu sistemlerin tüm kullanıcılar için adil ve eşit olmasını sağlamak için önem taşımaktadır (Ferrara, 2024, s. 2). YZ geliştiricileri, modellerinin dışında var olan gerçekliğe "temel gerçek" adını vermektedir ve YZ ile ilişkili olan ön yargı, bu gerçeklikten sapmalar, yanlış temsiller ve tahminler olarak tanımlanmaktadır (Zajko, 2022).

Yapay zekâ sistemlerinin ön yargısız ve adil bir şekilde çalışması etik ve toplumsal fayda açısından önemli bir role sahiptir. Bu önemi açıklayan örnek çalışma, Amerika’da sağlık hizmetleri veren bir şirket tarafından kullanılan algoritmalar üzerinden gerçekleştirilmiştir. Çalışmada, sağlık şirketinin kullandığı algoritmanın siyahi hastaların sağlık ihtiyaçlarını sistematik bir şekilde hafife aldığı ve bu durumun tıpta uzun süredir devam eden irksal

eşitsizlikleri derinleştirdiği ortaya konulmuştur. Sağlık sistemi maliyetlerini bir parametre olarak kullanan algoritmalar bu şekilde bir eşitsizliğin ortaya çıkmasına yol açmıştır. Şirket tarafından her ne kadar bu ayrımcılığın kasıtlı bir şekilde yapılmadığı açıklansa da algoritmaların sağlık hizmetlerinde olduğu gibi her alanda sistemik tarihsel ön yargılardan beslendiği ve dolaylı olarak eşitsizlikleri yeniden ürettiği görülmüştür. Sağlık sistemindeki bu ön yargıların azaltılması adına yapılan araştırmalar, sağlık sisteminde kullanılan algoritmaların kişilerin gelecekteki sağlık harcamalarına yönelik tahminler yapması yerine kişilerin potansiyel hastalıklarını öngöreceği şekilde yapılandırılmalarının önemli olduğunu göstermektedir (Johnson vd., 2019).

Zhou ve Sanfilippo (2023), ABD merkezli bir YZ uygulaması olan ChatGPT ve Çin merkezli bir YZ uygulaması olan Ernie'nin cinsiyet ön yargısıyla ilgili kamuoyu algısını ölçmek için sosyal medya uygulaması X üzerinden "gender bias in chatgpt" ve "gender bias in Ernie" şeklinde sorgulama yaparak bir çalışma yapmışlardır. Çalışmada, ChatGPT'nin erkekleri doktorlarla, bilim ve teknoloji ile ilgili alanlarla; kadınları ise öğretmenler ve sanat ile uygun alanlarla özdeşleştirdiğine dair sonuçlara ulaşılmıştır. Yazarlar, ChatGPT'nin meslek ünvanları ve cinsiyet rolleri konusundaki ön yargıları hakkında şikâyetlerin de yaygın olduğunu ve ChatGPT'nin cümle tamamlama istemlerinde ön yargı açısından zengin yanıtlar ürettiğini belirtmiştir. YZ sistemleri tarafından oluşturulan içeriklerde yer alan cinsiyet ön yargılarını ortaya çıkarmak amacıyla yapılan başka bir çalışmada, belirlenen istemler sonucu kadın, erkek ve tarafsız bakış açıları almayı talep eden beş farklı istemle ChatGPT tarafından üretilen metinler incelenmiştir. Çalışma sonucunda, ChatGPT'nin metin taslağı oluşturmak için faydalı bir YZ aracı olduğu ancak hatalar ve ön yargıları en aza indirmek için sonuçların dikkatle kontrol edilmesi gerektiği belirtilmiştir (Urchs vd., 2024, s. 14). YZ ve cinsiyet ön yargısı üzerine çalışan Ferrara (2024), belirlediği YZ araçlarından CEO görselleri üretmelerini istemiştir. Ağırlıklı olarak erkek görselleri üreten bu modeller, cinsiyet ayrımcılığını yansıttığı gibi iş dünyasında CEO pozisyonunda kadınların az temsil edildiğini de ortaya koymuştur.

Hall ve Ellis (2023), YZ ve cinsiyet ön yargısı ile ilgili yapılmış makaleleri inceledikleri çalışmalarında, YZ algoritmalarındaki cinsiyet ön yargısının algoritmik tasarıma ve eğitim veri kümelerine eşit şekilde atfedilebileceğini belirtmişlerdir. Ortaya konan bu sistemik ön yargının azaltılmasına dair önerilere de yer verilen çalışmada, YZ geliştirme ekiplerindeki çeşitliliğe, farkındalığı artırmaya, insan denetiminin süreçlere dahil edilmesine ve tasarım sürecine etik ilkelerin entegre edilmesine vurgu yapılmıştır. New York Üniversitesi AI Now Enstitüsü tarafından yapılan bir çalışma da YZ teknolojilerini geliştiren kişilere odaklanmıştır. Çalışmada, YZ geliştiricilerinin ezici bir çoğunlukla beyaz ve erkek olmalarının geriye kalan grupların daha az temsil edilmesine ve eşitlikten uzak sonuçların çıkmasına neden olduğu savunulmuştur. Çözüm noktasında tavsiyelerin yapıldığı ilgili çalışmada, YZ'nin ön yargıyı ele alıp daha eşitlikçi sistemler yaratmasının yolunun, geliştirici ekipler içerisinde daha fazla kadın, renkli insan ve diğer yeterince temsil edilmeyen kişilerin eklenmesinden geçtiği belirtilmiştir (Guynn, 2019). Guijarro (2023), dijital asistanlıktaki cinsiyet ön yargılarına odaklandığı çalışmasında dijital asistanların büyük çoğunluğunun kadın kimliğiyle tasarlandığını gözlemlemiştir. Ona göre, kadın kimliğinin orantısız şekilde temsil edilmesi, toplumsal cinsiyete dayalı geleneksel iş bölümü normlarının dijital alanda da pekiştiğini göstermektedir. Konu ile ilgili yapılmış analizler, müşteri hizmetleri, yardım hatları ve operatörlük gibi asistanlıkla bağlantılı mesleklerin sıklıkla kadınsı rollerle ilişkilendirildiğini ortaya koymaktadır. Ortaya çıkan bu sonuç, kadınların lider konumda sunulmasından ziyade yardımcı rolü ile bağdaştırılması şeklinde de okunabilmektedir.

Çeviri araçları üzerinden cinsiyet ön yargısının araştırıldığı bir çalışmada; meslekler, işler ve tanımlardan oluşan bir liste Google Translate aracılığıyla Türkçe'den İngilizce'ye çevirtilmiştir. Mesleklerin çoğunluğunu erkek olarak çeviren araç, özellikle kalıplaşmış olan hemşire ve dadı gibi meslekleri kadın olarak vermiştir. Ayrıca, kişileri niteleyen sıfatlarla ilgili yapılan denemelerde "güzel" ve "çalışkan" gibi olumlu nitelikler erkek olarak karşılık bulmuş, "tembel" gibi olumsuz sıfatın çevirinde ise kadın zamiri kullanılmıştır (Wellner, 2020, s. 129). Xavier (2024), internet ara yüzlerinin de kadın kullanıcılara çoğunlukla erkek cinsiyeti gibi yaklaştığını belirtmiştir. Bu durumun özellikle isim, sıfat, fiil gibi dilbilgisi bileşenlerinin öznenin cinsiyetine göre değiştiği dillerde daha sık görüldüğünü vurgulamıştır. Örneğin İbranice'de "ben öğretmenim" cümlesinde "öğretmen" kelimesi erkek ve kadın özneler için farklı olduğundan ve Google Translate üzerinden İngilizce'den İbranice'ye çeviri yapılırken yalnızca erkek kalıbının kullanıldığını belirtmiştir. Çok dilli araçların örtük ön yargılarını araştırarak Prates ve arkadaşları (2020) da, Google Translate gibi istatistiksel çeviri araçlarının cinsiyete dayalı ön yargı konusunda güçlü bir eğilim sergilediğini ve bu ön yargıların muhtemelen onları eğitmek için kullanılan gerçek dünya verilerinden kaynaklandığını belirtmiştir. ChatGPT aracılığıyla yapılan mesleklerle ve eylemlere dayalı çeviriler incelendiğinde de düşük kaynaklı dillerden İngilizce'ye yapılan çevirilerin meslekleri -örneğin doktoru erkek; hemşire ve aşçıyı kadın- örtük olarak ön yargılı davranarak çevirdiği gözlemlenmiştir (Ghosh ve Çalışkan, 2023, s. 908). Türkçe literatürde ise Çifci ve Başfıncı (2020) çalışmalarında, mesleklerle yönelik geliştirilen meslek robotları aracılığıyla, toplumsal cinsiyet algılarının katılımcı düzeyinde nasıl şekillendiğini ortaya koymayı amaçlamışlardır. Çalışmanın bulgularına göre, katılımcılar meslek robotlarını hem cinsiyet hem de kişilik özellikleri açısından değerlendirdiklerinde, cinsiyetçi algılama eğiliminde oldukları sonucuna ulaşmışlardır.

İncelenen bütün çalışma sonuçlarına bakıldığında, YZ teknolojisi ile kuşanmış birçok aracın görevlerini yerine getirirken cinsiyet eşitliğine dikkat etmediği söylenebilmektedir. İş yapış biçimlerinde ataerkil kodlardan izler barındıran bu araçların, toplumsal düzende cinsiyet başta olmak üzere birçok noktada ayrımcılığı artırma potansiyelinde olması kaygı yaratmaktadır.

3. Yöntem

Bu çalışmada nicel araştırma yöntemlerinden içerik analizi kullanılarak görsel üreten YZ araçlarının, belirlenen istemler (prompt) sonucunda ürettiği çıktılarda cinsiyete dair yaklaşımı saptanmaya çalışılmıştır. Bu araçların cinsiyete dair ön yargılı davranma durumlarını saptama amacı öncelikli olduğu için çalışma durum belirleme çalışması olarak değerlendirilebilmektedir. Betimleyici çalışmalar içerisinde yer alan durum belirleme-saptama araştırmalarında, araştırılmak istenen karakteristiklerin sayımı ve ortaya dökümü ile durum yansıtılmaya çalışılmaktadır (Aziz, 2015, s. 26). Verilerin analiz sürecinde başvuru alan içerik analizi, önceden belirlenmiş kategoriler esas alınarak sistematik ve tekrarlanabilir bir şekilde içeriği ölçmeyi amaçlayan bir analiz tekniğidir. İçerik analizinde önemli olan kriterler arasında yer alan nesnellik ve sistematiklik, analiz edilecek verinin geçerliliğini arttırmak adına önem taşımaktadır. Nesnellik kriterinin sağlanması, içerik analizi yapılacak verinin her araştırmacı tarafından aynı kurallarla değerlendirilebilmesini sağlamakta, sistematiklik ise belirlenen kuralların tüm veri setinde tekrarlanabilirliğini yansıtmaktadır (Bryman ve Bell, 2011, ss. 280-290). Sistematiklik ve tekrarlanabilirlik kriterleri baz alındığında, üretken YZ araçları ile cinsiyete dair ön yargılı yaklaşımın anlaşılmasının üretilen görseller aracılığıyla daha uygun olacağı düşünülmüştür. Metin formatında kapsamlı bilgiler verebilen YZ araçlarının, kullanıcıya geniş perspektif sunması sebebiyle doğrudan bir yaklaşımını tespit etmek güçleşmektedir. Bu endişe sebebiyle sadece görsel istemlerin yapıldığı çalışmada, örneklem olarak belirlenen YZ

araçlarından Tablo 1’de gösterilen 19 meslek ve 20 olumlu kişilik özelliğine dair görseller istenmiştir. İstemi yapılacak meslekler belirlenirken Türkiye’de ve dünyada Google üzerinden en çok sorgusu yapılan yirmi meslek baz alınmıştır (Remitly, 2022). Referans alınan bu listede “hukukçu” şeklinde belirlenen mesleğin, listede yer alan “avukat” ve “hâkim” mesleklerini kapsamı sebebiyle bu istem yapılmamıştır. Olumlu kişilik özelliklerinin, üretken YZ araçlarında hangi cinsiyete karşılık geldiğini anlamak adına meslekler dışında bu özellikler üzerinden de istemler yapılmıştır.

Tablo 1. İstemleri Yapılan Meslekler ve Olumlu Kişilik Özellikleri

Kategori	İstem Değişkenleri
Meslek	Avukat, Bilgisayar programcısı, Blogger, Dansçı, DJ, Doktor, Girişimci, Hakim, Influencer, İtfaiyeci, Oyuncu, Öğretmen, Pilot, Profesör, Psikolog, Şarkıcı, Uçuş görevlisi/ Kabin memuru, Yazar, Youtuber
Özellik/Sıfat	Başarılı, Cesur, Cömert, Çalışkan, Dikkat çekici, Dikkatli, Dürüst, Eğlenceli, Enerjik, Fedakâr, Girişken, Hoşgörülü, Kibar, Merhametli, Özgüvenli, Sorumluluk sahibi, Titiz, Yardımsever, Zeki, Zengin

Bu çalışmanın evrenini görsel üretebilen bütün üretken YZ araçları oluşturmaktadır. YZ teknolojisi ve transformer mimarisine dayanan birçok üretken YZ aracının görsel üretebilmesi göz önüne alındığında geniş bir evrenin varlığından söz edilebilmektedir. Bu sebeple, çalışmanın daha kolay yürütülebilmesi sebebiyle örneklem seçimine gidilmiştir. Örneklem seçimi yapılırken mevcut görsel üreten YZ araçları bilinirlik, kullanım potansiyeli ve ücretsiz erişim kriterleri üzerinden değerlendirilmiştir. Bilinirlik kriterinde Adobe Firefly, ChatGPT, Copilot, Gemini Advanced, Flux, Midjourney ve Leonardo.ai üretken araçları öne çıkmıştır. Adobe Firefly, Midjourney ve Leonardo.ai araçları, her istem sonucunda dört opsiyon sunarak kullanıcıya seçim hakkı sunan tasarımsal varyasyonlar ortaya koymaktadır. Kullanıcıya opsiyon sunularak üretilen görsellerde, en az bir görüntüde farklı cinsiyetin verilmesi, cinsiyet ve ön yargı düzleminde sağlıklı bir değerlendirme yapılmasının önüne geçmiştir. Bu sebeple, tek promptta birden fazla görsel üreten araçlar kapsam dışında tutulmuştur. Flux ve Gemini Advanced araçları tek görsel üzerinden çıktılar ortaya koymalarına rağmen ücretsiz şekilde erişim noktasında -çalışmanın yapıldığı tarih aralığında- yetersiz görülmüştür. Amaca yönelik örnekleme tekniğinden yararlanılan bu süreçte, araştırma amacına ve sorularına katkı sağlayacak ve zengin veriler sunacak araçların (Patton, 2002, s. 46) ChatGPT ve Copilot olduğuna karar verilmiştir. Bunun başlıca nedenleri; ChatGPT ve Copilot araçlarının yüksek bilinirliği, kullanım kolaylığı ve ücretsiz erişim seçeneğinin geniş kapsamı olmuştur. Üretken YZ modellerinden biri olan ve OpenAI şirketi tarafından 2022 yılının Kasım ayının sonlarına doğru tanıtılan ChatGPT, şirketin GPT 3,5 ve GPT-4 modellerine entegre şekilde çalışan bir sohbet arayüzü hizmeti olarak geliştirilmiş ve şirketin bilinirliğine geniş ölçekte katkı sağlamıştır. Ücretli ve ücretsiz kullanım seçenekleri olan ChatGPT; bağlamsal anlayışı, dil oluşturma yetenekleri, görev uyarlanabilirliği, çok dilli yeterliliği, ölçeklenebilirliği, sıfır atış ve az atış öğrenimi ve ince ayar potansiyeli gibi insan-makine etkileşiminde gelişmiş özelliklere sahip durumdadır (Du vd., 2023; King ve chatGPT, 2023). Gelişmiş bağlam anlayışı ve ince ayar yetenekleri gibi özelliklere sahip olan ChatGPT, günlük yaşam pratiklerinden bilimsel çalışmalara kadar birçok alanda kendine yer bulmuştur (Mbakwe vd., 2023). Çok dilli olması sebebiyle küresel düzeyde kullanımı yaygınlaşan ChatGPT (Zhuo vd., 2023) başlangıçta sadece metin tabanlı bir sohbet ara yüzü iken daha sonra aynı şirketin görsel üretim aracı Dall-E ile entegre şekilde hareket ederek görsel üretmeye başlamıştır. Ücretsiz kullanım seçiminde tekil

kullanıcıya günlük sınırlı görsel hakkı sunan ChatGPT, “create an image/ görsel oluştur” komutundan sonra yapılan isteme, tek bir görsel şeklinde çıktı sunmaktadır. Çalışmanın diğer örnekleme olan Copilot, Microsoft - ChatGPT’nin sahibi OpenAI şirketi ile iş birliği içerisinde olan- markasının 2023 yılına kadar Bing Chat, adı altında başlattığı YZ teknolojisine sahip aracının kapsamının genişletilmesiyle ortaya çıkmıştır (Muchmore, 2024). Bağlamsal anlayış, çeviri, etkileşimli sohbet ve görsel üretme özelliklerine sahip bu araç, ChatGPT’den farklı olarak ücretsiz şekilde sınırsız görsel üretebilmektedir.

Çalışmanın veri toplama sürecine geçilmeden önce çalışmayı verimli sürdürebilmek adına birtakım varsayımlar üzerinden hareket edilmiştir. Bu bağlamda, örnekleme oluşturan üretken YZ araçlarından istemler yapılırken dil, kullanıcı etkileşimi ve cinsiyet noktalarında farklı sonuçlar çıkabileceği varsayılarak araştırma kurgusu yapılmıştır. Görsel üreten araçların kişisel verilerden yola çıkarak cinsiyete göre farklı sonuçlar ortaya koyabileceği düşünülerek çalışmanın araştırmacısı olan iki kadın kullanıcı hesabı dışında iki erkek kullanıcı hesabından da aynı görsellerin istemi yapılmıştır. YZ araçlarının eğitim dilinin İngilizce olması ve bu dil ile yapılan aramaların evrensel –konumdan bağımsız- bir şekilde değerlendirilebileceği varsayımıyla iki kullanıcının Türkçe, iki kullanıcının da İngilizce istemler yapması belirlenmiştir. Dile dair gruplandırma yapılırken cinsiyet faktörü göz önünde tutulmuş ve İngilizce ya da Türkçe istem yapan kullanıcıların farklı cinsiyetlerde olmasına özen gösterilmiştir. 17 Aralık- 31 Aralık tarihleri arasında toplanan veriler, ortak çalışma klasöründe arşivlenmiştir (Ek 1).

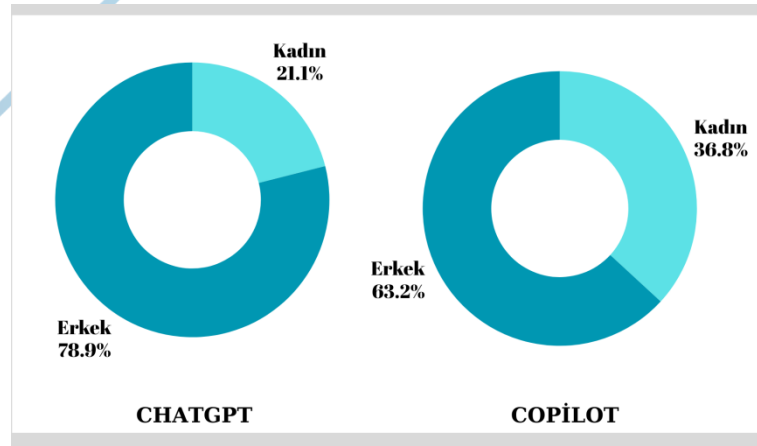
Varsayımlar ve çalışmanın zaman aralığı göz önüne alındığında ilgili araştırmanın örnekleme dayalı, zamansal, dilsel ve niceliksel sınırlılıklarının bulunduğunu söylemek mümkündür. İstemlerde kullanılan dillerin sadece Türkçe ve İngilizce olması, diğer dillerin kapsam dışında tutulduğunu göstermektedir. Cinsiyet değerlendirmesi yapılırken biyolojik cinsiyet olarak kabul edilen kadın ve erkek cinsiyetleri üzerinden değerlendirme yapılmıştır. Bu noktada heteronormatif bir yaklaşımdan ziyade üretilen görsellerin kolay değerlendirilebilmesi adına bu şekilde bir sınırlılığa başvurulmuştur. Çalışma sonucunda ortaya konan bulguların, sadece belirlenen kategoriler, örneklem ve kullanıcılarla çerçevelenmesi, çalışmanın niceliksel sınırlılığını göstermektedir.

4. Bulgular ve Yorum

ChatGPT ve Copilot üretken YZ araçlarından on dokuz meslek ve yirmi olumlu kişilik özelliği için her kullanıcı ayrı şekilde görsel istemiştir. Meslek ve olumlu kişilik özellikleriyle ilişkilendirilen meslek kategorileri birbirinden bağımsız şekilde değerlendirildiği için bunlara dair bulgular iki ayrı başlıkta verilmiştir.

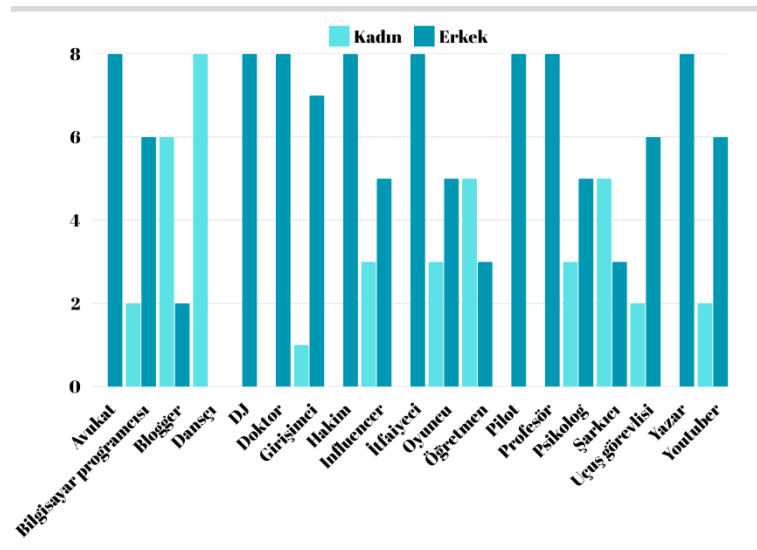
4.1. Mesleklere İlişkin Bulgular

Veri setinde yer alan on dokuz mesleğe dair istem yapılırken Türkçe ve İngilizce şekilde “*içinde X (belirlenen meslek) olan bir görsel oluştur*” komutu ile görseller üretilmiştir. Dört kullanıcının her meslek için ayrı görseller istemesiyle beraber iki üretken YZ aracından toplam 152 görsel elde edilmiştir. Bu görsellerin 108’i erkek cinsiyeti ile temsil edilirken sadece 44 görselde istenen mesleğin karşılığı kadın olarak verilmiştir. Bu araçların kendi aralarında kıyaslama yapıldığında Copilot tarafından üretilen görsellerde kadın cinsiyetinin ChatGPT’ye göre biraz daha fazla olduğu görülmüştür (Şekil 1).



Şekil 1: YZ Araçlarının Meslek Görsellerine Dayalı Cinsiyet Dağılımları (ChatGPT-Copilot)

ChatGPT tarafından üretilen meslek görsellerinden 60'ı (%78,9) erkekleri işaret ederken 16 (%21,1) görselde, istemi yapılan meslek kadın cinsiyeti ile ilişkilendirilmiştir. Mesleklere dayalı cinsiyet ön yargısını devam ettiren Copilot, ChatGPT'ye oranla biraz daha fazla kadın tasvirine yer vermiş olsa da ürettiği 76 meslek görselinin 28'i kadın cinsiyetini, 48 görselin ise erkek cinsiyetini yansıttığı sonuçlarına ulaşılmıştır. Literatürde yer verilen çalışma sonuçlarıyla benzer bulguların ortaya çıktığını gösteren bu tablo, örnekleme yer alan iki üretken YZ aracının, erkek yanlı algoritmik sistemi desteklediği şeklinde yorumlanabilmektedir. Üretilen görsellerin önemli çoğunluğunun erkek cinsiyeti ile ilişkilendirilmesi; YZ geliştiricilerinin erkek ağırlıklı olmasına, üretilen araçların eğitim kaynaklarındaki erkek yanlı örneklemelere ve kadınların ekonomik katılım seviyesinin düşük olmasına dayandırılabilir. Dünya Ekonomik Forum'u (WEF) tarafından düzenli şekilde gerçekleştirilen Küresel Cinsiyet Farkı Raporu'nun (2024) sonuçlarına göre kadınların iş gücüne katılım oranı erkeklere göre düşüktür. Bunun ardında yatan nedenler ise kadınların daha zor koşullar altında çalıştırılması, alt pozisyonların kadınlar için uygun görülmesi ve pandemi gibi herhangi bir acil durumda ilk olarak gözden çıkarılacak elemanların kadınlardan oluşmasıdır. Bu açıdan bakıldığında, iş gücüne katılım noktasında erkeklerden geride olan kadınların, mesleki temsil noktasında daha az gösterilmesi üretken YZ araçlarının örneklem yanlılığı yapmasına sebebiyet vermektedir.



Şekil 2: Cinsiyetlerle İlişkilendirilen Mesleklerin Dağılımı

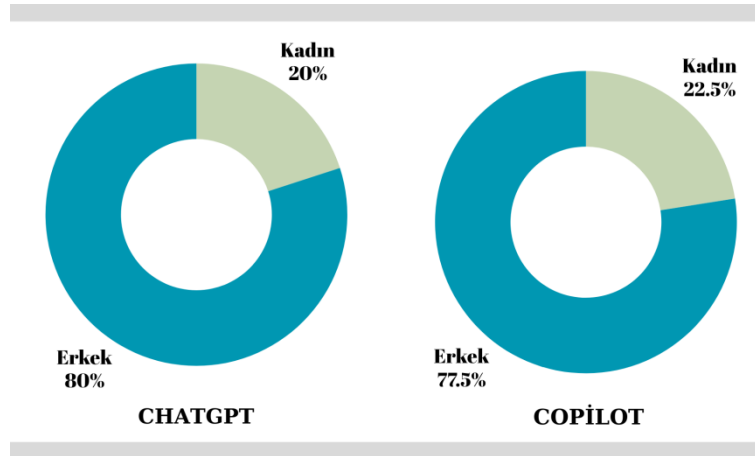
Bu araçlar tarafından üretilen meslek görsellerinin dağılımına bakıldığında (Şekil 2); avukat, DJ, doktor, hâkim, itfaiyeci, pilot, profesör ve yazar mesleklerinin tamamının erkeklerle ilişkilendirildiği görülmüştür. Dansçı mesleği ile ilgili görsel üreten ChatGPT ve Copilot araçlarının, bu mesleğe dair ürettiği bütün görseller kadın cinsiyetini yansıtmıştır. Blogger, öğretmen ve oyuncu meslekleri ile ilgili üretilen görsellerde kadın temsili erkek temsiline fazla olsa da kalan bütün meslekler çoğunlukla erkek üzerinden gösterilmiştir. Belirlenen mesleklerin cinsiyetlerle ilişkilendirilme durumunda örneklem yanlılığı ve referans değer kıstaslarının etkili olduğu düşünülmektedir. Bu bağlamda, örneklem yanlılığı erkeklerin sayıca fazla olduğu mesleklerde kendini gösterirken referans değer kıstası da çok bilinen kişilerin baz alınarak ilişkilendirme yapılmasını içermektedir. Örnek vermek gerekirse tamamen erkekler ile bağdaştırılan mesleklerden biri olan DJ'lik mesleğinde dünyaca ünlü kişilerin çoğunun erkek olması, üretken YZ araçlarının bu şekilde bir eşleştirme yapması şeklinde değerlendirilebilmektedir. Burada ünlü olanı öne çıkaran mekanizmaların inanırılığı ve güvenilirliği her ne kadar sorun teşkil etse de mevcut tabloyu yorumlama noktasında göz önünde bulundurulabilecek bir etmen olarak görülmektedir. Erkekler ile ilişkilendirilen mesleklerin ortak özelliklerine bakıldığında, bu kategori içerisinde yer alan her bir mesleğin analitik karar verme ve sorun çözme süreçleriyle ilgili olduğu söylenebilmektedir. Tamamı erkekler ile ilişkilendirilen avukat, doktor, hâkim ve itfaiyeci gibi mesleklerin elzem bir sorun olduğunda başvurulacak kişileri işaret etmesi bu mesleklerin bilişsel boyutunu göstermekte ve erkekleri daha analitik olarak algılayabilmektedir. Bunun yanı sıra bütün görselleri kadınlar ile eşleştirilen dansçılık mesleğinin ise estetik ve duygusal boyutunun ağır basması cinsiyetler arasında yapılan ayrımcı bakış açısının bir başka göstergesidir. Her iki cinsiyet için de sorun yaratan bu durum kalıp yargıların devam etmesine neden olmaktadır. Cinsiyetler bazında meslek seçiminde etkisini gösteren bu yargılar, itfaiyeci olmak isteyen bir kadının kendini engellemesine ya da dansçı olmak isteyen bir erkeğin bu mesleği feminen olarak değerlendirmesine yol açabilmektedir.

Benzer bulgular Nikolic ve Jovicic'in (2023) YZ araçlarının ürettiği görsellerle ilgili yaptıkları deneysel çalışmalarında da elde edilmiştir. Çalışmada cinsiyet stereotiplerinin YZ araçları tarafından görsel olarak yeniden üretilmesinin erkek egemen yönündeki stereotipleri güçlendirdiği, bu sonuçların mevcut eşitsizliklerin bir örneği olduğu ve daha eşitlikçi bir sistem yaratmak için yalnızca yeni YZ araçlarına güvenilemeyeceği belirtilmiştir. Meslek görsellerini inceleyen Garcia ve Lazaro'nun (2023) çalışmasında da bu çalışmanın sonuçlarıyla paralellik gösteren bulgulara ulaşılmıştır. Yazarlar tarafından yapılan çalışmada, hemşire, hizmetçi, ilköğretmeni, şarkıcı, terzi, sekreter gibi mesleklerin görselleri kadınları temsil etmiş; marangoz, taksi şoförü, kamyon şoförü, uçak pilotu, tamirci, inşaat işçisi, asker, mühendis, polis, bankacı, bilgisayar uzmanı, politikacı gibi mesleklerin görselleri ise erkekleri göstermiştir. Benzer şekilde Avrupa Uygulamalı Bilimler Üniversitesi (2024), YZ'nin cinsiyet ile ilgili ön yargısını araştırmak amacıyla Leonardo AI ve DALL-E tarafından görselleştirilen 50 mesleği inceleyerek YZ tarafından oluşturulan görsellerde, kökleşmiş kalıp yargıların yansıtıldığı sonucuna ulaşmıştır. Araştırmaya göre incelenen mesleklerin neredeyse yalnızca erkeklerle ilişkilendirildiği kadınların ise öncelikli olarak sosyal ve bakım rollerinde tasvir edildiği ortaya çıkmıştır. Görüldüğü gibi benzer çalışmalardan elde edilen bulgular, bu çalışmadan elde edilen bulguları desteklemektedir. YZ araçları meslekleri cinsiyetlerle ilişkilendirirken algoritmik olarak yanlı davranmakta ve ilgili görselleri yansıtırken mesleklerin büyük çoğunluğunu erkek cinsiyeti ile ilişkilendirmektedir. Bu veriler toplumda var olan mesleklerle cinsiyet arasındaki ilişki konusunda bazı mesleklerde erkeklerin daha öncelikli olabileceğine dair algıyı güçlendirmeye katkı sağlayabilmektedir. Aynı şekilde kadınların da meslek seçimi yaparken

bazı mesleklerle ilgili tercihleri olumsuz yönde etkilenebilmekte ve bu mesleklerin erkekler için uygun olduğuna dair toplumdaki kanı pekiştirilebilmektedir.

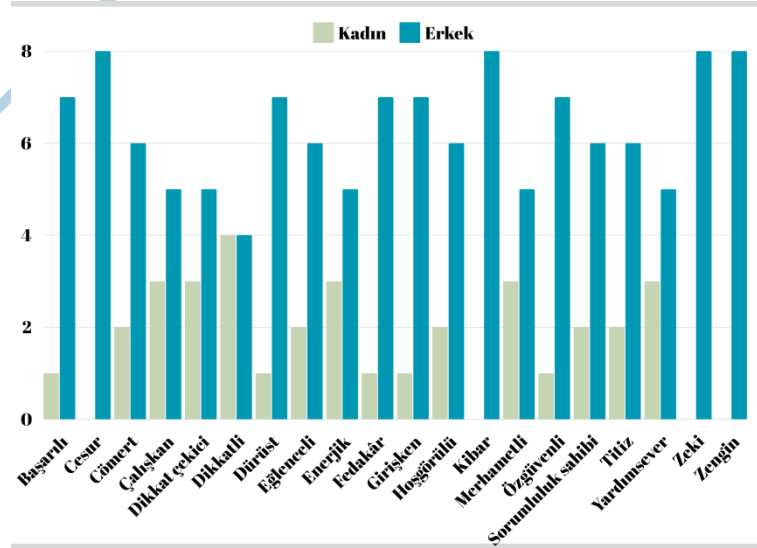
4.1. Olumlu Kişilik Özelliklerine İlişkin Bulgular

Mesleklerle ilgili istemler yapıldıktan sonra olumlu kişilik özelliklerinin karşılığının, üretken YZ araçlarında hangi cinsiyeti temsil ettiğini anlamak adına belirlenen yirmi olumlu kişilik özelliği üzerinden istemler yapılmıştır. Meslek görsellerinde olduğu gibi “içinde X (belirlenen özellik) olan bir görsel oluşturun” komutu üzerinden istenen görseller, her kullanıcı tarafından ayrı şekilde alınmıştır. Bunun sonucunda ortaya çıkan 160 görselin 126’sında erkek cinsiyeti yer alırken 34 görselde kadın cinsiyeti temsil edilmiştir. ChatGPT ve Copilot arasında karşılaştırma yapıldığında mesleklerde ortaya çıkan sonuca benzer durum görülmüştür (Şekil 3).



Şekil 3: YZ Araçlarının Olumlu Kişilik Özelliği Görsellerine İlişkin Cinsiyet Dağılımları (ChatGPT-Copilot)

Copilot YZ aracı, ChatGPT’ye göre düşük bir farkla daha çok kadın cinsiyeti resmetmiştir. Copilot’un olumlu kişilik özelliklerine dair ürettiği 80 görselin 18’i (%22,5) kadın, kalan görsellerde ise temsil erkek cinsiyetine yönelik olmuştur. ChatGPT tarafından üretilen görsellerin ise 16’sında (%20) olumlu kişilik özelliğinin karşılığında kadın cinsiyeti yer almıştır. ChatGPT (%80) ve Copilot (%77,5) YZ araçlarının, erkekler ile olumlu kişilik özelliklerini ilişkilendirme oranlarına bakıldığında erkek yanlı bir yaklaşımın varlığından söz etmek mümkündür.



Şekil 4: Olumlu Kişilik Özellikleri ile İlişkilendirilen Görsellerin Dağılımı

Şekil 4’te verilen olumlu kişilik özellikler ile cinsiyetlerin ilişkilendirilme sonuçlarına bakıldığında ChatGPT ve Copilot tarafından üretilen görsellerde; cesur, kibar, zeki ve zengin özellikleri için üretilen bütün görsellerin erkekleri tasvir ettiği görülmüştür. Başarılı, dürüst, fedakâr, girişken ve özgüvenli sıfatlarının her iki üretken YZ aracındaki karşılığında her özellik için ortaya konan 8 görselin 7’sini erkekler oluşturmuştur. Kadınların hiçbir olumlu özellikte çoğunlukta olmadığı bu sonuçlarda, kadınlar ile en çok ilişkilendirilen olumlu özellik dikkatli özelliği olmuştur. Bu özelliği devam ettiren ve kadınlar ile ilişkilendirilen –erkeklerden daha az şekilde- diğer özellikler de çalışkan, dikkat çekici, enerjik, merhametli ve yardımsever özellikleri olmuştur. Sözcüklerin ayrımcılık düzeyindeki etkisi Menegatti ve Rubini tarafından yapılan araştırmada da ortaya koyulmuştur. Bu çalışmada, kadınlar ile özdeşleştirilen sözcüklerin şefkat ve iyilik ile ilişkili olması, erkeklerin ise “iddialı”, “etkin” (Menegatti ve Rubini, 2017, s. 1-2) gibi baskın roller yüklenen özelliklerle temsil edilmesi cinsiyet eşitsizliğinin sözcükler aracılığıyla ne şekilde yapıldığına bir örnek teşkil etmektedir.

Ortaya çıkan bu sonuçlar tıpkı mesleklerde olduğu gibi olumlu kişilik özellikleri görsellerinde de üretken YZ araçlarının örneklem yanlılığına başvurduğunu, kalıplaşmış yargıları devam ettirdiğini göstermektedir. Üretken YZ araçlarının örneklem yanlılığı yaptığı düşünülen özelliklerin başında “zengin” gelmektedir. Dünyanın en zengin insanların paylaşıldığı listelerde yıllardır ilk on sırada erkekler bulunmaktadır ve listenin çoğunluğu da erkeklerden oluşmaktadır (Forbes, 2024). Sadece bu veri bile gerçeğe yakın ve doğruluk oranı fazla sonuçlar üretmek amacıyla görevlendirilen YZ araçlarının tarafı sonuçları göstermesinde etkili olan etmenlerden biridir. Bunun dışında, geçmişten günümüze dek devam eden eril zihniyetin tezahürü erkekler ile ilişkilendirilen özelliklerin daha belirleyici sevide olmasına ve liderlik ile ilişkilendirilmesine yol açmaktadır. Eve bakma ve para kazanma sorumluluğu yüklenen erkekler ile ev işleri ile ilgilenme, çocuklara bakma görevi verilen kadınların (Olsen, 1999) varlığının kabul edildiği ataeril düzenin, günümüzde bir seviyede yol kat ettiği bilirse de geçmiş algıların her yeni teknolojiye yeniden örtük şekilde kendine yer bulduğunu söylemek mümkündür. Bu çerçeveden bakıldığında; “cesur, kibar ve zeki” olan erkek dışarı ile bağlar kurabilmek ve para kazanmak için gerekli özelliklere sahip görünmekte, nitekim parayı kazanan olarak da –zengin- erkek sunulmaktadır. Kadın ise ailesine ve evine zaman ayırırken

gereken dikkate sahip olmakta, anaç karakter ile ilişkilendirilen merhamet ve yardımseverliği sergilemekte aynı zamanda etrafına iyi görünmek için dikkat çekici ve enerjik olmaktadır.

Sonuç ve Tartışma

“Teknolojinin ikiliği” kavramını ortaya atan Wanda J. Orlikowski (1991, s. 11) burada teknolojiyi nesnel bir güç olarak gören ve toplumsal olarak inşa edilmiş bir ürün olarak karşıt şekilde değerlendiren görüşlere alternatif sunmuştur. Ona göre teknoloji, belirli bir toplumsal bağlamda çalışan aktörler tarafından fiziksel olarak inşa edilmektedir ve bu özelliğiyle nesnel bir güç olarak görülebilmektedir. Pasif teknoloji bakış açısının ötesinde duran bu yaklaşımda, teknolojinin ikiliğini yaratan diğer taraf ise toplumsal inşada üstlendiği roldür. Teknolojiyi, onu üreten aktörlerin kendisine yükledikleri farklı anlamlar, vurguladıkları ve kullandıkları çeşitli özellikler aracılığıyla toplumsal inşa gerçekleşme boyutuyla ele alan yazar, kapsamlı bir teknoloji değerlendirmesinin iki yönlülüğüne dikkat çekmiştir. Eleştirel bakış açısıyla teknolojiye yaklaşan Langdon Winner da teknolojinin sosyo-ekonomik boyutuna ve toplumsal inşadaki rolüne odaklanmıştır. Yazara göre, teknoloji yalnızca insan faaliyetlerine yardımcı değildir, aynı zamanda bu faaliyetleri ve anlamları yeniden şekillendiren bir güç durumundadır (1989, s. 6). Teknolojiye genel çerçeveden yaklaşan bu erken dönem çalışmalarıyla benzer bir yaklaşımla hareket eden bu çalışma da teknolojinin toplumsal inşadaki görevini göz önüne almaktadır. Çalışmada, bu sorun cinsiyet ön yargısı bağlamında ve 21. yüzyılın teknolojisi olarak nitelendirebileceğimiz YZ teknolojisi özelinde incelenmiştir.

Üretken YZ araçlarından ChatGPT ve Copilot’un örneklem olarak seçildiği çalışmada, her iki aracın görsel üretme özelliği incelenmiştir. Belirlenen meslekler ve olumlu kişilik özellikleri ile bu araçların cinsiyetleri ilişkilendirmesi üzerinden kurgulanan çalışma sonuçlarında, üretilen görsellerin büyük bir çoğunluğunun öznesi erkek cinsiyeti olmuştur. Üretilen meslek görsellerinde sorunların çözümünde doğrudan muhatap alınan (avukat, doktor, hakim) teknik beceriler gerektiren (pilot, itfaiyeci) ve yüksek statü gösteren (profesör, yazar) birçok meslek erkekler ile eşleştirilmiştir. Kadınlar ile eşleştirilen mesleklerin başında ise estetik yönü ağır basan dansçılık mesleği gelmiştir. Eğitim ve eğlence alanlarıyla kadın cinsiyetini özdeşleştiren bu araçlar, öğretmen ve şarkıcı gibi mesleklerin temsilde çoğunlukla kadına yer vermiştir. Dijital teknolojiler ile birlikte kendini gösteren blogger, Influencer ve Youtuber gibi mesleklerin temsillerinde ise geleneksel mesleklere oranla daha az ayrımcılık yapılmıştır. Bu meslekler arasında yer alan blogger çoğunlukla kadınlarla ilişkilendirilmiş, Youtuber mesleği çoğunlukla erkeklerle bağlantılı görülmüş, Influencer mesleği ise birbirine yakın oranlarda temsil edilmiştir. Olumlu kişilik özelliklerine dair görsellerin istemlerinin de yapıldığı çalışmada cesur, kibar, zeki ve zengin özelliklerinin tamamı erkek cinsiyetiyle resmedilmiş; hiçbir özellikte kadınlar çoğunlukta yer almamıştır. Erkeklerle eşit oranda kadın temsili sağlayan dikkatli özelliği, kadınlarla ilişkilendirilen özelliklerden olmuştur. Bu özellik dışında; çalışan, dikkat çekici, enerjik, merhametli ve yardımsever özellikleri de kadınlarla eşleştirilen –erkeklerden az şekilde- özelliklerden olmuştur. Meslekler ve olumlu kişilik özelliklerine ait görsel sonuçlarına dair genel çıkarım yapıldığında, ChatGPT ve Copilot araçlarının algoritmik olarak erkek yanlı davrandığı açıkça görülen bir durum olmaktadır. Bu durumun altında yatan sebeplerin başında ise örneklem etkisi, yerleşik kalıp yargılar, teknoloji geliştiriciler ve cinsiyetler arasındaki teknolojiye erişim farklılıklarının geldiği düşünülmektedir. Gross’un (2023) belirttiği gibi, YZ araçları arasında en yaygın kullanılan ChatGPT gibi sistemler cinsiyete dayalı ön yargılar içerebilmekte ve bu durum toplumsal eşitsizlikleri güçlendirebilecek potansiyele sahip olabilmektedir. YZ teknolojisinin toplumsal yaşamı dönüştürme ve iyileştirme potansiyeli son

derece büyük olsa da bu potansiyel etik prensiplere bağlılık ve etkin düzenleyici politikalar olmadan sürdürülebilir bir şekilde gerçekleşmemektedir. YZ tasarımının tarafsız ve adil bir şekilde yapılması toplumsal adaletin sağlanmasında önemli bir rol oynamaktadır. Literatürde yapılan araştırmalar, üretken YZ araçlarında belirlenen cinsiyet ayrımcılığının, toplumdaki mevcut eşitsizlik durumlarını pekiştirme potansiyeline işaret etmekte ve bu teknolojilerin etik sorumluluğuna vurgu yapmaktadır. Üretken YZ araçları bireylerin hayatına birçok fayda sağlasa da eşitsizlikler noktasında belirli düzenlemelerin yapılmasına gereksinim duyulmaktadır. YZ yanlılığıyla ilgili gerekli tedbirler alınmadığı takdirde, bu araçlar toplumsal eşitsizlikleri daha fazla arttırarak yeniden üretebilmekte ve bu eşitsizlikleri daha çok derinleştirebilmektedir.

Toplumsal yansıması olan ve kökleşmiş ön yargıları yansıtan basmakalıpların, YZ araçlarında kendini daha az göstermesinin yolunun, veri setlerindeki çeşitlilikten geçeceği düşünülmektedir. YZ sistemleri için kullanılan eğitim verilerinde çeşitliliği ve eşitliği teşvik etmek ve teknolojiyi etkileyecek yerleşik yargılar hakkında insanlar arasında daha fazla farkındalık oluşturmak bu noktada önemli bir çözümü oluşturmaktadır (García-Ull ve Melero-Lázaro, 2023). Liu (2021)'nin de belirttiği gibi YZ'nin yaygınlaşması, insanların düşünce ve davranış biçimlerini bilgisayar mantığına yaklaştırmakta ve bu etkileşimin yayılması YZ kültürünü popüler hâle getirmektedir. Toplum için bir yandan önemli faydalar ve kolaylıklar getiren bu kültür, ciddi zorluklara da yol açabilmekte ve bu zorlukların çözümü pek kolay olmamaktadır. Bu nedenle, teknolojinin toplumlara daha yararlı ve sürdürülebilir bir hâle gelmesi için YZ alanındaki uzmanların, sosyal bilimler alanında çalışan kişilerle iş birliği yapması büyük önem taşımaktadır. YZ geliştiricileri kadar kullanıcılara da sorumluluğun düştüğü bu süreçte, kullanıcıların YZ araçlarını kullanırken ve sonuçlarını yorumlarken daha dikkatli ve eleştirel olmaları gerekmektedir (Carter vd., 2024). Kullanıcı etkileşimine açık bir şekilde gelişimlerini sürdüren bu araçlardan istemler yapılırken eşitlikçi davranmak, kalıp yargıları desteklememek, ön yargılı şekilde verilen bir sonuca eleştiri yapmak ve bunu YZ aracına detaylı şekilde anlatmak sorunun çözümünde gerekli olan bilinci geliştirebilecek ve farkındalığı arttırabilecektir.

Bu çalışmanın kurgusu yapılırken üretken YZ araçlarının İngilizce ve Türkçe istemler ile farklı cinsiyetlere sahip kullanıcı isteklerine ayrı sonuçlar verebileceği varsayılmıştır. Elde edilen bulgular neticesinde bu varsayımları güçlendirecek görsellere rastlanmamış, ChatGPT ve Copilot her iki dilde ve farklı kullanıcılarda da cinsiyet temsili benzer şekilde yapmıştır. Çalışmada cinsiyet ön yargısı öncelikli olduğu için sonuçlar bu bağlamda değerlendirilmiştir. Üretilen görsellerde cinsiyet dışında ırk ve bedensel görünüm noktasında da yanlı sonuçların ortaya koyulduğu saptanmıştır. Bu açıdan bakıldığında ileride yapılacak çalışmaların diğer ön yargı türlerini de kapsayacak şekilde genişletilmesi, farklı örneklem gruplarıyla hareket edilmesi önerilmektedir.

Açıklamalar

* *Etik Kurul Onayı:* Araştırma etik kurul onayı gerektirmemektedir.

* *Yayın Etiği:* Bu çalışma, "Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi" kapsamında belirtilen kurallara uygun olarak hazırlanmıştır. Ayrıca, makale intihal tespit yazılımlarıyla (Turnitin / iThenticate) taranmış ve herhangi bir intihal tespit edilmemiştir.

* *Yazar Katkı Oranı:* Her iki yazarın katkı oranı %50'dir.



* **Çıkar Çatışması:** Çalışmanın yazarlarının herhangi bir kurum veya bu kurumun çalışanlarıyla araştırmayı etkileyebilecek düzeyde doğrudan veya dolaylı olarak herhangi bir finansal, ticari, hukuki veya profesyonel ilişkisi/çıkartı söz konusu değildir.

* **Akademik Finansal Destek:** Çalışma; herhangi bir akademik finansal destek kuruluşu (TÜBİTAK, BAP, Avrupa Birliği, Birleşmiş Milletler, vs.) tarafından desteklenmemiştir.

* **Yazar Beyanı:** Çalışma bir tezden üretilmemiş ve herhangi bir kongre/ sempozyum/ konferansta sunulmamıştır.

Structured Extended Abstract

AI models can lead to unfair results in the outputs they produce, maintain existing inequalities, including productive biases affecting the representation of individuals in the generated data, and reinforce existing socio-cultural discriminations that promote inequalities (Angwin et al., 2022; Cirillo et al., 2020; Kaplan et al., 2024; Silberg & Manyika, 2019; West et al., 2019). Prates et al. (2020) stated that since the data sets on which AI models are trained contain information shaped by social norms and cultural values, it may be possible to obtain results about stereotypical gender roles in society by analyzing these data sets. From this point of view, in order to contribute to the studies in literature, this study examined the potential for generative AI tools to behave biasedly while producing information in the context of gender. Many studies conducted on the axis of technology and gender have revealed that the tools used in different periods have male-biased algorithms and continue established stereotypes. The fact that no study has yet been conducted in the Turkish literature addressing this issue within the framework of AI models gives original value to the relevant study.

Research Methodology

In this study, the issue of gender was examined through the images produced by generative AI tools through their visual creation features. Considering that generative AI tools continue their development with user interaction, it is thought that the picture to be presented as a result of the study will serve as a status determination and can provide awareness about this point while establishing new communications between technology producers and users. The following questions are asked depending on the purpose of the research:

RQ1: Do generative AI tools exhibit a prejudiced attitude in terms of gender?

RQ1a: Do generative AI tools act prejudiced when making associations between professions and gender?

RQ1b: Do generative AI tools act prejudiced when making associations between gender and positive adjectives that describe people?

In this status determination study, where content analysis was used, ChatGPT and Copilot generative AI tools were selected as samples. These tools were asked to produce images related to nineteen professions and twenty features that describe people positively. The main problem of the study was the curiosity about which genders ChatGPT and Copilot would associate the professions or positive features within the images they produced.

Research Results

After the analysis of the data, it was seen that both AI tools in the sample visualized male-biased results. Of the 152 profession images produced, 108 (71.1%) were represented by the male gender, while forty-four (28.9%) were represented by the female gender. When these



tools were compared with each other, it was seen that the female gender was slightly higher in the images produced by Copilot than in ChatGPT. The professions of lawyer, DJ, doctor, judge, firefighter, pilot, professor and writer were all associated with men. The profession most associated with the female gender was presented as dancing, while the professions of blogger, teacher and actor were mostly represented with the female gender. Similar results were also found for the features that positively describe people other than professions. Out of 160 images produced regarding the features, 126 images were represented by the male gender, and only thirty-four of them were represented by the female gender. In the images produced by ChatGPT and Copilot, all the images produced for the features of brave, polite, intelligent and rich reflected the male gender. In both productive AI tools, seven out of eight visuals presented for each feature were men, in return for the adjectives successful, honest, altruistic, enterprising and self-confident. In these results, where women were not the majority in any positive feature, the careful feature was most associated with women. Other features associated with women less than men were hardworking, remarkable, energetic, compassionate and helpful.

Conclusion & Discussion

When the general conclusion is made about the results of adjectives that describe professions and people positively, it is clearly seen that ChatGPT and Copilot tools are male-biased. The underlying reasons for this situation are thought to be mainly due to the sample effect, established stereotypes, technology developers and differences in access to technology between genders. At this point, it is of great importance for experts in the field of AI to cooperate with people working in the field of social sciences in order to make technology more useful and sustainable for societies. In this process, where responsibility falls to users as much as AI developers, users need to be more careful and critical when using AI tools and interpreting their results (Carter et al., 2024). When making requests from these tools that continue their development in a way that is open to user interaction, acting egalitarian, not supporting stereotypes, criticizing a result given in a biased way and explaining this to the AI tool in detail can develop the consciousness necessary to solve the problem and increase awareness.

Kaynakça

- Abid, A., Farooqi, M., ve Zou, J. (2021). Persistent anti-Muslim bias in large language models. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 298-306. <https://doi.org/10.1145/3461702.3462624>
- Angwin, J., Larson, J., Mattu, S., ve Kirchner, L. (2022). Machine bias. İçinde *Ethics of data and analytics*. Auerbach Publications.
- Aziz, A. (2015). *Sosyal bilimlerde araştırma yöntemleri ve teknikleri*. Nobel Yayıncılık.
- Balestri, R. (2024). *Examining multimodal gender and content bias in ChatGPT-4o* (arXiv:2411.19140). arXiv. <https://doi.org/10.48550/arXiv.2411.19140>
- Bliuc, A.-M., Faulkner, N., Jakubowicz, A. ve McGarty, C. (2018). Online networks of racial hate: A systematic review of 10 years of research on cyber-racism. *Computers in Human Behavior*, 87, 75-86. <https://doi.org/10.1016/j.chb.2018.05.026>



- Bowman, S. R. (2023). *Eight things to know about large language models* (arXiv:2304.00612). arXiv. <https://doi.org/10.48550/arXiv.2304.00612>
- Bryman, A. ve Bell, E. (2011). *Business Research Methods: Bell, Emma: 9780199668649: Amazon.com: Books*. Oxford University Press. <https://www.amazon.com/Business-Research-Methods-Alan-Bryman/dp/0199668647>
- Carter, C., Kurkinen, V., König, L., Ruohonen, A., ve van Vloten, E. C. (2024). *Exploring gender bias in ChatGPT* [Graduate]. University of Helsinki.
- Chan, A., Salganik, R., Markelius, A., Pang, C., Rajkumar, N., Krasheninnikov, D., Langosco, L., He, Z., Duan, Y., Carroll, M., Lin, M., Mayhew, A., Collins, K., Molamohammadi, M., Burden, J., Zhao, W., Rismani, S., Voudouris, K., Bhatt, U., Maharaj, T. (2023). Harms from increasingly agentic algorithmic systems. *2023 ACM Conference on Fairness, Accountability, and Transparency*, 651-666. <https://doi.org/10.1145/3593013.3594033>
- Çifci, B. S., ve Basfirinci, C. (2020). Yapay zekâ konusunun toplumsal cinsiyet kapsamında incelenmesi: Mesleklere yönelik bir araştırma. *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 29(4), Article 4. <https://doi.org/10.35379/cusosbil.819510>
- Cirillo, D., Catuara-Solarz, S., Morey, C., Guney, E., Subirats, L., Mellino, S., Gigante, A., Valencia, A., Rementeria, M. J., Chadha, A. S., ve Mavridis, N. (2020). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. *Npj Digital Medicine*, 3(1), 1-11. <https://doi.org/10.1038/s41746-020-0288-5>
- Crawford, J., Cowling, M., ve Allen, K.-A. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *Journal of University Teaching and Learning Practice*, 20(3), Article 3. <https://doi.org/10.53761/1.20.3.02>
- Crawford, K., ve Paglen, T. (2021). Excavating AI: The politics of images in machine learning training sets. *AI & SOCIETY*, 36(4), 1105-1116. <https://doi.org/10.1007/s00146-021-01162-8>
- Cyfirma. (2023, Ocak 21). *ChatGPT AI in security testing: Opportunities and challenges*. CYFIRMA. <https://www.cyfirma.com/outofband/chatgpt-ai-in-security-testing-opportunities-and-challenges/>
- Du, H., Teng, S., Chen, H., Ma, J., Wang, X., Gou, C., Li, B., Ma, S., Miao, Q., Na, X., Ye, P., Zhang, H., Luo, G., ve Wang, F.-Y. (2023). Chat with ChatGPT on intelligent vehicles: An IEEE TIV Perspective. *IEEE Transactions on Intelligent Vehicles*, 8(3), 2020-2026. IEEE Transactions on Intelligent Vehicles. <https://doi.org/10.1109/TIV.2023.3253281>
- Favaretto, M., De Clercq, E., ve Elger, B. S. (2019). Big data and discrimination: Perils, promises and solutions. A systematic review. *Journal of Big Data*, 6(1), 12. <https://doi.org/10.1186/s40537-019-0177-4>
- Ferrara, E. (2024). Fairness and bias in artificial intelligence: A brief survey of sources, impacts and mitigation strategies. *Sci*, 6(1), Article 1. <https://doi.org/10.3390/sci6010003>
- Forbes. (2024). *Billionaires list—the richest people in the world ranked*. Forbes. <https://www.forbes.com/billionaires/>

- Friedman, B., ve Nissenbaum, H. (1996). Bias in computer systems. *ACM Trans. Inf. Syst.*, 14(3), 330-347. <https://doi.org/10.1145/230538.230561>
- García-a-Ull, F.-J., ve Melero-Lázaro, M. (2023). Gender stereotypes in AI-generated images. *profesional de la información*, 32(5), Article 5. <https://doi.org/10.3145/epi.2023.sep.05>
- Ghosh, S., ve Caliskan, A. (2023). ChatGPT perpetuates gender bias in machine translation and Ignores Non-gendered Pronouns: Findings across Bengali and five other Low-resource languages. *Proceedings of the 2023 ACM Conference on International Computing Education Research V.1*, 397-415. <https://doi.org/10.1145/3568813.3600120>
- Glickman, M., ve Sharot, T. (2024). How human–AI feedback loops alter human perceptual, emotional and social judgements. *Nature Human Behaviour*, 1-15. <https://doi.org/10.1038/s41562-024-02077-2>
- Gross, N. (2023). What ChatGPT tells us about gender: A cautionary tale about performativity and gender biases in AI. *Social Sciences*, 12(8), Article 8. <https://doi.org/10.3390/socsci12080435>
- Guijarro, S. T. (2023). Sesgos de género de la asistencia digital en español. *Gender on Digital. Journal of Digital Feminism*, 1, 35-58. <https://doi.org/10.35869/god.v1i.5061>
- Guynn, J. (2019, Nisan 16). *AI, facial recognition too white, too male: More minorities needed*. USA Today. <https://www.usatoday.com/story/tech/2019/04/17/ai-too-white-male-more-women-minorities-needed-facial-recognition/3451932002/>
- Güngör, N. (2018). *İletişim: Kuramlar ve yaklaşımlar* (4. bs). Siyasal Kitapevi.
- Hagerty, A., ve Rubinov, I. (2019). *Global AI ethics: A review of the social impacts and ethical implications of artificial intelligence* (arXiv:1907.07892). arXiv. <https://doi.org/10.48550/arXiv.1907.07892>
- Hall, P., ve Ellis, D. (2023). A systematic review of socio-technical gender bias in AI algorithms. *Online Information Review*, 47(7), 1264-1279. <https://doi.org/10.1108/OIR-08-2021-0452>
- Johnson, C. Y., Boodman, S. G., Nirappil, F., Diamond, D., Sun, L. H., Johnson, M., ve Bisset, V. (2019, Ekim 24). Racial bias in a medical algorithm favors white patients over sicker black patients. *Washington Post*. <https://www.washingtonpost.com/health/2019/10/24/racial-bias-medical-algorithm-favors-white-patients-over-sicker-black-patients/>
- Kaplan, D. M., Palitsky, R., Arconada Alvarez, S. J., Pozzo, N. S., Greenleaf, M. N., Atkinson, C. A., ve Lam, W. A. (2024). What's in a name? Experimental evidence of gender bias in recommendation letters generated by ChatGPT. *Journal of Medical Internet Research*, 26, e51837. <https://doi.org/10.2196/51837>
- King, M. R. ve chatGPT. (2023). A conversation on artificial intelligence, chatbots, and plagiarism in higher education. *Cellular and Molecular Bioengineering*, 16(1), 1-2. <https://doi.org/10.1007/s12195-022-00754-8>
- Leavy, S., Meaney, G., Wade, K., ve Greene, D. (2020). Mitigating gender bias in machine learning data sets. İçinde L. Boratto, S. Faralli, M. Marras, ve G. Stilo (Ed.), *Bias and*

- social aspects in search and recommendation* (C. 1245, ss. 12-26). Springer International Publishing. https://doi.org/10.1007/978-3-030-52485-2_2
- Liang, P. P., Wu, C., Morency, L.-P., ve Salakhutdinov, R. (2021). Towards understanding and mitigating social biases in language models. *Proceedings of the 38th International Conference on Machine Learning*, 6565-6576. <https://proceedings.mlr.press/v139/liang21a.html>
- Liebrezn, M., Schleifer, R., Buadze, A., Bhugra, D., ve Smith, A. (2023). Generating scholarly content with ChatGPT: Ethical challenges for medical publishing. *The Lancet Digital Health*, 5(3), e105-e106. [https://doi.org/10.1016/S2589-7500\(23\)00019-5](https://doi.org/10.1016/S2589-7500(23)00019-5)
- Lippens, L. (2024). Computer says 'no': Exploring systemic bias in ChatGPT using an audit approach. *Computers in Human Behavior: Artificial Humans*, 2(1), 100054. <https://doi.org/10.1016/j.chbah.2024.100054>
- Liu, Z. (2021). Sociological perspectives on artificial intelligence: A typological reading. *Sociology Compass*, 15(3), e12851. <https://doi.org/10.1111/soc4.12851>
- Lynch, S. (2017, Mart 10). *Andrew Ng: Why AI Is the New Electricity*. Stanford Graduate School of Business. <https://www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity>
- Mbakwe, A. B., Lourentzou, I., Celi, L. A., Mechanic, O. J., ve Dagan, A. (2023). ChatGPT passing USMLE shines a spotlight on the flaws of medical education. *PLOS Digital Health*, 2(2), e0000205. <https://doi.org/10.1371/journal.pdig.0000205>
- McCarthy, J., Minsky, M. L., Rochester, N., ve Shannon, C. E. (2006). A proposal for the dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Magazine*, 27(4), Article 4. <https://doi.org/10.1609/aimag.v27i4.1904>
- McKenna, M. (2019, Ekim 5). *Three notable examples of AI bias*. <https://aibusiness.com/responsible-ai/three-notable-examples-of-ai-bias>
- Menegatti, M., ve Rubini, M. (2017). Gender bias and sexism in language. İçinde *Oxford Research Encyclopedia of Communication*. <https://oxfordre.com/communication/display/10.1093/acrefore/9780190228613.001.0001/acrefore-9780190228613-e-470>
- Muchmore, M. (2024, Haziran 18). *What Is Copilot? Microsoft's AI assistant explained*. PC Mag. <https://www.pcmag.com/explainers/what-is-microsoft-copilot>
- Naumova, E. N. (2023). A mistake-find exercise: A teacher's tool to engage with information innovations, ChatGPT, and their analogs. *Journal of Public Health Policy*. <https://doi.org/10.1057/s41271-023-00400-1>
- Nikolic, K., ve Jovicic, J. (2023, Nisan 3). *Reproducing inequality: How AI image generators show biases against women in STEM*. UNDP. <https://www.undp.org/serbia/blog/reproducing-inequality-how-ai-image-generators-show-biases-against-women-stem>
- Olsen, K. (1999). *Daily life in 18th-century England*. The Greenwood Press.
- Orlikowski, W. J. (1991). *The duality of technology: Rethinking the concept of technology in organizations*. Massachusetts Institute of Technology.

- Parsheera, S. (2018). A gendered perspective on artificial intelligence. *Machine Learning for a 5G Future*, 1-7. <https://doi.org/10.23919/ITU-WT.2018.8597618>
- Pasquale, F. (2015). *The black box society: The secret algorithms that control money and Information*. Harvard University Press. <https://www.jstor.org/stable/j.ctt13x0hch>
- Patton, M. Q. (2002). *Qualitative research ve evaluation methods* (3. bs). SAGE Publications Ltd. <https://us.sagepub.com/en-us/nam/qualitative-research-evaluation-methods/book232962>
- Prates, M. O. R., Avelar, P. H., ve Lamb, L. C. (2020). Assessing gender bias in machine translation: A case study with Google Translate. *Neural Computing and Applications*, 32(10), 6363-6381. <https://doi.org/10.1007/s00521-019-04144-6>
- Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, 3, 121-154. <https://doi.org/10.1016/j.iotcps.2023.04.003>
- Remitly. (2022). *Dream jobs around the world study*. Remitly. <https://www.remitly.com/gb/en/landing/dream-jobs-around-the-world>
- Rozado, D. (2023). The political biases of ChatGPT. *Social Sciences*, 12(3), Article 3. <https://doi.org/10.3390/socsci12030148>
- Russell, S. J., ve Norvig, P. (2010). *Artificial intelligence: A modern approach* (3rd ed). Prentice Hall.
- Sharma, M. (2024, Aralık 27). *What is AI bias? Almost everything you should know about bias in AI results*. TechRadar. <https://www.techradar.com/computing/artificial-intelligence/what-is-ai-bias-almost-everything-you-should-know-about-bias-in-ai-results>
- Shrestha, S., ve Das, S. (2022). Exploring gender biases in ML and AI academic research through systematic literature review. *Frontiers in Artificial Intelligence*, 5, 976838. <https://doi.org/10.3389/frai.2022.976838>
- Silberg, J., ve Manyika, J. (2019). Notes from the AI frontier: Tackling bias in AI (and in humans). *McKinsey Global Institute*, 18.
- UE. (2024, Eylül 25). *DALL-E and Leonardo AI: When artificial intelligence generates outdated gender roles*. UE. <https://www.ue-germany.com/news-centre/press/when-artificial-intelligence-generates-outdated-gender-roles>
- Urchs, S., Thurner, V., Aßenmacher, M., Heumann, C., ve Thiemichen, S. (2024). *How prevalent is gender bias in ChatGPT? -- Exploring German and English ChatGPT Responses* (arXiv:2310.03031). arXiv. <https://doi.org/10.48550/arXiv.2310.03031>
- Wakefield, J. (2016, Mart 24). Microsoft chatbot is taught to swear on Twitter. *BBC News*. <https://www.bbc.com/news/technology-35890188>
- WEF. (2024, Haziran 11). *Global gender gap report 2024*. World Economic Forum. <https://www.weforum.org/publications/global-gender-gap-report-2024/>
- Wellner, G. P. (2020). When Ai is gender-biased. *Humana Mente*, 13(37). <https://philpapers.org/rec/WELWAI>

- West, S. M., Whittaker, M., ve Crawford, K. (2019). *Discriminating systems: Gender, race, and power in AI*. <https://ainowinstitute.org/publication/discriminating-systems-gender-race-and-power-in-ai-2>
- Winner, L. (1989). *The whale and the reactor: A search for limits in an age of high technology*. University of Chicago Press.
- Wolf, M. J., Miller, K. W., ve Grodzinsky, F. S. (2017). Why we should have seen that coming: Comments on Microsoft's Tay "Experiment," and Wider Implications. *The ORBIT Journal*, 1(2), 1-12. <https://doi.org/10.29297/orbit.v1i2.49>
- Wu, G. (2022, Aralık 22). *8 Big problems with OpenAI's ChatGPT*. MUO. <https://www.makeuseof.com/openai-chatgpt-biggest-probelms/>
- Xavier, B. (2024). Biases within AI: Challenging the illusion of neutrality. *AI & SOCIETY*. <https://doi.org/10.1007/s00146-024-01985-1>
- Zajko, M. (2022). Artificial intelligence, algorithms, and social inequality: Sociological contributions to contemporary debates. *Sociology Compass*, 16(3), e12962. <https://doi.org/10.1111/soc4.12962>
- Zhou, K. Z., ve Sanfilippo, M. R. (2023). *Public perceptions of gender bias in large language models: Cases of ChatGPT and Ernie* (arXiv:2309.09120). arXiv. <https://doi.org/10.48550/arXiv.2309.09120>
- Zhuo, T. Y., Huang, Y., Chen, C., ve Xing, Z. (2023). *Red teaming ChatGPT via Jailbreaking: bias, robustness, reliability and toxicity* (arXiv:2301.12867). arXiv. <https://doi.org/10.48550/arXiv.2301.12867>

EK 1

Araştırma kapsamında elde edilen görsellerin erişim klasörü: <https://drive.google.com/drive/folders/1F9y7B5-xSgdZRNkIXCmci-XbLjx0Vz5V?usp=sharing>

