VERİ BİLİMİ DERGİSİ
www.dergipark.gov.tr/veri

Gerbner'in Yetiştirme Teorisi Perspektifinden Yapay Zekâ'da Yanlılık Problemi

Duygu ERGÜN[1*], Ural Altan BOZKURT[2], Baran BİNGÖL[2], Seyhmus BASKIN[2], Şuheda TEMUR[3], Savaş TAKAN[2]

[1]*Atılım Üniversitesi, Mimarlık ve Güzel Sanatlar Fakültesi, Ankara,*
[2]*Ankara Üniversitesi, Yapay Zeka ve Veri Mühendisliği, Ankara,*
[3]*Uludağ Üniversitesi, Bilgisayar Mühendisliği, Bursa*

Makale Bilgisi

Başvuru:
18/04/2023
Kabul:
17/09/2023

Özet

Günümüzde, hayati önem taşıyan kararların alındığı durumlarda ve güvenlik açısından kritik birçok ortamda yaygın biçimde kullanılmaları dolayısıyla yapay zekâ uygulamalarının adilliği kritik bir konu haline gelmiştir. Çünkü bu sistemlerin kullanımı zaman içerisinde adillik, yanlılık ve mahremiyet gibi çeşitli yönlerden hatalı sonuçlar ortaya koymaya başlamıştır. Bunun üzerine, yapay zekâ kaynaklı beklenmedik sonuçlarla başa çıkmak için teknolojiler geliştirilmeye başlanmış ve sorunların çözümü için şirketler genellikle algoritma odaklı hatalara odaklanmıştır. Ancak kullanılan çözümler genellikle birçok yapay zekâ algoritmasında işe yaramamaktadır. Çünkü sorunun nedeni sadece algoritma değildir; aynı zamanda, örneğin derin öğrenmede, neden-sonuç ilişkisinin kolayca kurulamadığı verinin kendisidir. Ayrıca istatistiksel veya sezgisel algoritmalarda sınırlar belirsiz olmakta ve bu da istatistiksel veya sezgisel analiz için belirli bir standardın oluşturulamamasına neden olabilmektedir. Öte yandan, adalet sadece algoritmaya göre değil, aynı zamanda bağlam ile ilgili verilere bağlı olarak da değişebilmektedir. Bu açıdan bakıldığında, makalede odaklandığımız nokta verilerin nasıl olması gerektiğidir ki bu bir istatistik meselesi değildir. Hatta bağlamdan dolayı ülkeden ülkeye ve kültürden kültüre değişebilmektedir. Bu nedenle sadece bilgisayar bilimlerinin değil iletişim, sosyoloji, sanat, hukuk vb. sosyal bilimlerin de konuya katkısı oldukça önemlidir. İnsan kaynaklı veriler, gündelik hayatın bir yansıması olarak değerlendirilebileceğinden, toplumdaki çeşitli grupların hangi bağlamlarda var olduklarına dair önemli ipuçları içermektedir. Bu noktadan hareketle, bilgisayar bilimleri ve sosyal bilimlerin ortak katkısıyla, bu çalışmada elde edilen ipuçlarını kullanarak yapay zekâ algoritmalarından kaynaklanabilecek olası toplumsal tehlikeleri öngörmek amaçlanmaktadır. Bu doğrultuda günümüzün en temel sorunlarından biri olan "savunmasız ve dezavantajlı" gruplara özel bir senaryo üzerinden söz konusu tablo ortaya konmuştur. Ardından, gelmiş geçmiş en yaygın kitle iletişim teorisi olan Gerbner'in "yetiştirme teorisi", makine öğrenmesi perspektifinde yeniden yorumlanarak, makine öğrenmesinin veriye bağlı olarak yol açabileceği olası toplumsal ve kitlesel sorunlara dikkat çekilmiştir. Makalenin, bütüncül bir yaklaşımın (yani algoritma ve veri kombinasyonunun) ve disiplinler arası bir değerlendirmenin önemine katkı sunması beklenmektedir.

Anahtar Kelimeler: *Yapay zekâ, Yetiştirme teorisi, Makine öğrenmesi, Veri yanlılığı, Dezavantajlı gruplar*

* İletişim e-posta: duygu.ergun@atilim.edu.tr

The Problem of Bias in Artificial Intelligence from the Perspective of Gerbner's Cultivation Theory

Abstract

Today, the fairness of artificial intelligence applications has become a critical issue due to their widespread use in situations where vital decisions are made and in many security-critical environments. Because the use of these systems has started to produce erroneous results in various aspects such as fairness, bias, and privacy over time. In response, technologies have started to be developed to deal with unexpected results from artificial intelligence, and companies have generally focused on algorithm-oriented errors to solve problems. However, the solutions used generally do not work in many artificial intelligence algorithms. This is because it is not only the algorithm that causes the problem but also the data itself, where, for example, in deep learning, the cause-effect relationship cannot be easily established. From this perspective, our focus in this paper is on how the data should be, which is not a matter of statistics. It may even vary from country to country and culture to culture due to context. For this reason, the contribution of not only computer science but also social sciences such as communication, sociology, art, law, etc. to the subject is very important. Since human-derived data can be considered as a reflection of daily life, it contains important clues about the contexts in which various groups in society exist. From this point of view, with the joint contribution of computer sciences and social sciences, it is aimed to predict possible social dangers that may arise from artificial intelligence algorithms by using the clues obtained in this study. In this direction, the picture in question is presented through a scenario specific to "vulnerable and disadvantaged" groups, which is one of the most fundamental problems of today. Then, Gerbner's "nurture theory", which is the most widespread mass communication theory of all time, is reinterpreted in the perspective of machine learning, and attention is drawn to the possible social and mass problems that machine learning may cause depending on the data. The article is expected to contribute to the importance of a holistic approach (i.e. the combination of algorithms and data) and an interdisciplinary evaluation.

Keywords: *Artificial intelligence, Cultivation theory, Machine learning, Data bias, Disadvantaged groups*

1 Giriş

Günümüz algoritmalarının doğruluğu çok önemli bir noktadadır. Çünkü siyahi vatandaşlar, Müslümanlar, engelliler, yaşlılar ve diğer azınlıklarla ilgili algoritmaların neden olduğu bazı talihsiz durumlar olmuştur. Bu talihsizliklerin nedenlerine baktığımızda birkaç nokta dikkat çekmektedir. Bunlardan ilki her geçen gün popülaritesi artan derin öğrenme algoritmalarıdır. Ancak bu tür teknikler geriye dönük neden- sonuç ilişkisi kurma konusunda sorunludur. Algoritmaların neden olabileceği talihsiz durumların bir diğer nedeni ise istatistiksel ya da sezgisel analiz için belirli bir standardın olmamasıdır [1]. Başka bir deyişle, istatistiksel ya da sezgisel algoritmalarda sınırlar belirsizdir. Ancak bu belirsizliğin veri ile ilişkisi ortaya koyulmamıştır. Algoritmaların büyük çoğunluğunun bu sınıfa girdiği düşünüldüğünde bu durum ciddi bir sorun olarak karşımıza çıkmaktadır.

Öte yandan, algoritmanın doğruluğunun bağlamıyla çok ilgisi vardır. Doğru bir algoritma yanlış veriler üzerinde doğru bir sonuç gösterebilir mi? Hata toleransı nedir? Aynı zamanda doğru algoritma doğru veriyi ne kadar hatalı bir şekilde bilgiye dönüştürür? Ya da yanlış algoritma doğru veri ile ne tür sonuçlar üretebilir? Yanlış bir algoritma, yanlış verilerle ne kadar doğru sonuç verebilir? Algoritma bu kadar çok veriyle ilişkiliyken, algoritmayı veriden ayrı değerlendirmek ne kadar uygun olur?

Belki de bu iki yapı geçmişten gelen alışkanlıklara göre birbirinden ayrılmış olabilir [2]. Ancak algoritmaların da veri olduğunu unutmamak gerekmektedir. Dolayısıyla veri temelli yaklaşım zaten algoritmaları da içerecektir. Bu açıdan bakıldığında algoritma ve veri arasındaki sınır ortadan kalkmaktadır. Var olan bu gerçeklik çerçevesinde bilgisayar bilimlerinde kullanılan test mekanizmalarının veri adaletine uyarlanması faydalı sonuçlar doğurabilmektedir.

Bir verinin kullanılabilmesi için sahip olması gereken nitelikler nelerdir ya da hangi durumlarda kullanılabilir? Bir veri tıp alanında kullanılacaksa hangi standartlara sahip olmalıdır? Doğru verileri hangi ölçütlerle oluşturulabilir? Algoritmaya hangi özellikler dahil edilmelidir? Peki var olması gereken veriler ve halihazırda var olan verilerle arasındaki ilişki nedir? Aralarında bir dönüşüm gerekli midir? Tüm bu soruların, veriye dayalı test mekanizmalarını şekillendirmek için yol gösterici olduğu düşünülmektedir. Söz konusu soruların cevaplanabilmesi için kullanılacak her verinin bir üst verisinin olması ve verinin nerede ve hangi bağlamda kullanılacağının belirtilmesi gerekmektedir. Bunları anlayabilmek ve doğru sonuçlar üretebilmek için disiplinler arası çalışmalara ihtiyaç duyulmaktadır.

İstatistik, mevcut verilerin doğru örneklenmesi üzerine inşa edilmiştir [3]. Bu çalışmada odaklanılan sorun daha ziyade doğru bağlamda veri üzerinde çalışmaktır. İhtiyaç duyulan verinin ne olduğu, nasıl olduğu, nasıl oluşturulduğu, test metodolojilerinin neler olduğu gibi konularda literatürün zenginleştirilmesi gerektiğini düşünülmektedir. Bu açıdan makalede odaklanılan konu, verinin nasıl olması gerektiğidir ki bu bir istatistik konusu değildir. Kaldı ki verinin nasıl olması gerektiği konusu tek başına bilgisayar bilimlerinin araştırma konusu olamaz ve iletişim, sosyoloji, hukuk, tıp, psikoloji, makine vb. birçok farklı disiplinle iş birliği yapılması gerekmektedir. Verinin ait olduğu disiplinle iş birliği yaparak çözüm üretmek, veriden beslenen algoritmaların doğruluğu açısından önemlidir.

Veriler birçok farklı ve tartışmalı değer üzerine inşa edilebilmektedir. Örneğin, Amerika'daki düşünme biçimi farklı bir ülkede farklı şeyler ifade edebilmektedir. Bu noktada bağlam çok önemli hale gelmektedir. Örneğin herkesi etkileyebilecek bir yapay zekâ teknolojisinde herhangi bir metin kullanılıyorsa, bu metnin en azından insan hakları evrensel beyannamesine uygun olması ve mümkünse kullanıldığı kültüre yönelik değerlere sahip olması beklenmektedir. Otonom araçlarda kullanılan bir sensör verisinin de aynı titizlikle kültürel ve evrensel değerler çerçevesinde test edilmesi gerekmektedir [4]. Bu, bir sayının evrensel bir test mekanizmasından nasıl geçirileceği konusunda üzerinde düşünülmesi gereken önemli bir konudur. Nasıl ki ülkeden ülkeye ya da küresel ölçekte değişiklik gösterebilen birçok sayısal kalite standardı varsa, bu konuda da temel kalite standartlarının geliştirilmesi ve yayınlanması gerekmektedir.

Günümüz algoritmalarının büyük çoğunluğunda aynı işi yapan az sayıda test ve çok fazla program bulunmaktadır. Github ve Bitbucket gibi bulut tabanlı iş birliğine dayalı sürüm kontrol sistemleri bu sorunu daha da kötüleştirmektedir [5], [6]. Çünkü bu programlar çok önemli programlara eklenebilmekte ve programın işlevselliği açısından ciddi zafiyetlere neden olabilmektedir. Bu kadar çok programın kolayca var olması özgürlük açısından güzel görünse de veri tabanlı ciddi bir güvenlik zafiyetini de beraberinde getirebilmektedir. Tüm bu nedenlerle algoritmanın yanında ve öncesinde veri odaklı bir standardizasyon elzem görünmektedir.

Algoritmaların temel dayanağının veri olduğu bir ortamda bu gibi talihsizliklerin yaşanmaması için sadece algoritmalara odaklanmak yeterli olmayacaktır. Veriyi araştırmadan algoritmaların doğruluğuna odaklanmak, bugüne kadar konuyla ilgili yapılan çalışmaların en büyük eksikliği olabilir. Bu çalışmada Facebook'un algoritmalarında oluşturduğu veri seti, savunmasız ve dezavantajlı gruplarla ilişkilendirilebilecek bazı kelimeler açısından incelenmiştir. Savunmasız ve dezavantajlı gruplar ifadesinin kapsamı kültürden kültüre, coğrafyadan coğrafyaya ve tarihsel perspektife göre değişebilmektedir. Bu sorunu araştırırken, bu durumun medya mesajlarını alan kitleler üzerinde de benzer bir etkiye sahip olabileceği keşfedilmiştir. Bu sebeple veri seçiminin önemine dikkat çekmek amaçlanmıştır.

Daha önce de belirtildiği gibi, neden olabileceği sosyal sorunlara dikkat çekmek için gereklilik olgusal verilerle tanımlanmıştır. Buna ek olarak, Gerbner'in "yetiştirme teorisi"[7], makine öğreniminin insan verileri nedeniyle neden olabileceği olası sosyal ve kitlesel sorunlara dikkat çekmek için makine öğrenmesi perspektifinde yeniden yorumlanmıştır.

2 İlgili Çalışmalar

Makine öğrenimi sistemleriyle ilgili adillik, kullanılan algoritmalara ve kullanılan verilere göre ayrılabilir. Bu nedenle, ilgili çalışmaları algoritma merkezli adillik ve veri merkezli adillik olarak ikiye ayırdık.

2.1 Algoritma Merkezli Adillik

Algoritmik yanlılık esas olarak makine öğrenimi

sürecinde kullanılan yazılımlardan kaynaklanmaktadır. Algoritmik adaletsizlikle mücadele etmek için kullanılan çeşitli çalışmalar vardır.

Algoritmik ayrımcılığın adil olup olmadığını test etmek için yazılım test yöntemleri de kullanılmaktadır. Galhotra ve arkadaşları, herhangi bir Test Oracle'ı (programın bir testi geçip geçmediğini veya başarısız olup olmadığını belirleyen bir mekanizma) gerektirmeden otomatik olarak ayrımcılık testleri oluşturmak için kullanılan Themis adlı bir test senaryosu oluşturma aracı sunmaktadır [8]. Bu araç, algoritmaların ırk ve yaşa karşı adilliğini ölçmek için nedensel ayrımcılık puanı hesaplamaktadır.

Adilliği test etmek için, bir yazılım aracı olarak makine öğrenimi modelleri, sistemlerin yalnızca girdilerinin ve çıktılarının ayrımcılığa karşı kontrol edildiği kara kutu testi olarak kabul edilir. Aggarwal ve arkadaşları, bireysel ayrımcılığı tespit etmek ve test girdilerini otomatik olarak oluşturmak için bir yöntem sunmaktadır [9]. Test girdilerini otomatik olarak oluşturmak, yol kapsamını en üst düzeye çıkarmak ve farklı girdilerde yürütme yolu kısıtlamalarını toplamak için dinamik bir sembolik yürütme tekniği kullanmaktadırlar. Ayrıca deneysel bir çalışma da sunmaktadırlar. Ayrıca Galhotra ve arkadaşları Themis aracına kıyasla yaş, cinsiyet ve ırk özelliklerine göre 8 farklı kriter için değerlendirme sağlamaktadır. Themis %6,4 başarı sağlarken onlar %34,8 ortalama başarı puanı elde etmiştir.

Ayrıca, Udeshi ve arkadaşları makine öğrenimi modelindeki adillik ihlallerini tespit etmek için ölçeklenebilir test oluşturma yaklaşımını tanıtmışlardır [10]. Yerel arama için aday test girdileri haline gelen hassas parametreleri (örneğin ırk, din, cinsiyet) tespit etmek için kullanılan rastgele test girdilerinin üretildiği küresel arama algoritmasını kullanmışlardır. Yerel aramada benzer test girdilerini tespit etmek için aday girdilerin komşuluğunda arama yapmışlardır. Altı yaygın makine öğrenimi sınıflandırıcısı için deneysel bir değerlendirme sağlamışlardır.

Hertweck ve arkadaşları, algoritmanın özelliklerini kullanmak yerine, istatistiksel adillikle ilgili ahlaki yönlere odaklanmayı amaçlamışlardır [11]. İstatistiksel adilliği, saf matematiksel kullanımı genişleterek ahlaki perspektifiyle ilgili bağımsızlık ölçütü olarak kullanmaktadırlar. Önerileri Friedler ve arkadaşlarının bir uzantısıdır. Friedler ve arkadaşları, doğuştan yetenekli olan bireyin yaşamının neden olduğu önyargının gerçekleşmiş yeteneklere dönüşmesi ile ilgilidir. Uzantıya dayanarak, iki karşı örnek de sunmuşlar ve önerilerinin evrensel olarak doğru olmadığı sonucuna varmışlardır [12].

Konuyla ilgili son olarak IBM tarafından, makine öğrenimi modellerindeki önyargıyı tespit etmeye ve ortadan kaldırmaya yardımcı olabilecek açık kaynaklı bir yazılım araç seti olan "AI Fairness 360 (AIF360)" geliştirilmiştir[1].

2.2 Veri Merkezli Adillik

Yaşam boyunca genler, çevre ve yaşam tarzındaki farklılıkları dikkate alarak kişiselleştirilmiş önleyici ve tedavi edici stratejiler bulmaya çalışan hassas tıp gibi farklı alanlardaki veri kümelerinde adillik üzerine çeşitli ampirik çalışmalar mevcuttur. Cinsiyet ve toplumsal cinsiyet eşitsizlikleri, biyotıp ve sağlık hizmetleri alanına yönelik bir derlemede ele alınmaktadır [13]. Bu derlemede yazarlar, sağlık alanındaki cinsiyet ve toplumsal cinsiyet farklılıklarını anlamak için mevcut ana biyomedikal veri türlerini ve çeşitli yapay zekâ teknolojilerinin rolünü vurgulamayı amaçlamaktadır. Sonuçlara dayanarak, eşitsizlikleri azaltmak için küresel sağlık ve hastalık ortamının iyileştirilmesine yönelik öneriler sunmaktadırlar.

Bethhall ve arkadaşları tarafından yapılan ilginç bir çalışmada toplum ve tasarımcılar için bir ikilemden bahsetmektedir: ilki ırksal grup eşitsizliklerine karşı kör olmak ve böylece sistemik eşitsizliği artık ölçmeyerek ırksallaştırılmış sosyal eşitsizliğin farkına varmak veya ikincisi olarak ırkı reddeden bir şekilde ırksal kategorilerin bilincinde olmaktır [14]. Üçüncü bir seçenek olarak, grup adilliği müdahalelerinden önce, ayrımcılık kalıplarını dinamik olarak tespit etmek için denetimsiz öğrenme ile makine öğrenimi sistemlerinin, sosyal eşitsizliklerin, sosyal ayrışmanın ve tabakalaşmanın temel nedenini, dezavantajlı kategorileri daha fazla sabitlemeden azaltabileceğini öne sürmektedirler. Bu yaklaşımın dezavantajı uygulanabilirlik olabilmektedir.

Kearns ve arkadaşları zengin alt grup ayrımı üzerine bir algoritma önermektedir. Sunulan algoritmanın kapsamlı bir ampirik değerlendirmesini, adilliğin söz konusu olduğu dört gerçek veri kümesi üzerinde

[1] https://www.ibm.com/opensource/open/projects/ai-fairness-360/

uygulamışlardır ve aşağıdaki durumlarda algoritmanın temel yakınsamasını araştırmışlardır [15]. Programın bir testi geçip geçmediğini veya başarısız olup olmadığını belirleyen bir mekanizma olan Test Oracle'ları yerine hızlı sezgiseller ile örneklendirilmiş, adillik ve doğruluk arasındaki değiş tokuşları ölçmüş ve bu yaklaşımı son algoritmalarla karşılaştırmışlardır.

Yazılım testlerinde veri ve algoritmaların ayrılması yeni değildir. Bu konudaki çalışmaların neredeyse tamamı hata tespiti noktasında veri ve algoritma ayrımı yapmaktadır.

Verinin mi yoksa algoritmanın mı hatalı olduğunun belirlenmemesi ciddi sorunlara neden olabilmektedir çünkü veri göz ardı edildiğinde bir algoritma tamamen sorunsuz görünebilmektedir. Veri ve algoritmanın kavramsal olarak ayrı tutulması, mevcut sorunların göz ardı edilmesine neden olabilmektedir. Özellikle, bu sorunlardan kaçınmak için veri ve algoritmaların birlikte ele alınması gerektiği düşünülmektedir.

Öte yandan, daha önce bahsedilen çalışmaların hiçbiri sorunun tanımlanmasına ilişkin sistematik bir bakış açısı geliştirmemiştir. Dahası, mevcut çalışmalar bu sorunun ciddiyetini ve genişliğini ortaya koymamıştır. Makalenin önceki çalışmalardan temel farkı, sorunu sistematik bir yaklaşımla tanımlamayı amaçlamasıdır. Ardından, ele alınan sorunun nasıl çözülmesi gerektiği konusunda sosyal bilimlerle etkileşim içinde olan disiplinler arası bir yaklaşımın gerekliliğini vurgulamasıdır. Bu doğrultuda, ortaya koyulan problemin, Gerbner'in yetiştirme teorisine dayanan sosyal bilimsel bir kavramsal çerçevede tartışılmış olmasıdır.

Yukarıda bahsedilen kavramları doğrulamak için Facebook tarafından oluşturulan Türkçe ve İngilizce veri setleri kullanılmıştır. Bunlar arasından, özellikle savunmasız ve dezavantajlı grupları ifade edebilecek kelimeler belirlenmiştir.

Daha sonra kelimeler arasındaki benzerlikleri gözlemlenmiştir. İnceleme ve sonuçların detayları aşağıda verilmiştir.

3 Veri Analizi ve Sonuçlar

Çalışmamızda, Facebook tarafından CBOW ve Skip-Gram modelleri kullanılarak oluşturulan veri kümeleri metinleri sayısal vektörlere çevirmek için kullanılmıştır (Word2Vec). Facebook bu veri kümelerini birçok farklı dilde yayınlamıştır ve bu çalışmada, güvenilirlik açısından Joulin ve arkadaşlarının veri kümesi tercih edilmiştir [16]. Veri kümelerinde, Common Crawl ve Wikipedia üzerinde fastText kullanılarak eğitilmiş 157 dil için önceden eğitilmiş kelime vektörlerinden yararlanılmıştır. Bu modeller, 300 boyutunda, 5 uzunluğunda karakter n-gramları, 5 boyutunda bir pencere ve 10 negatif ile konum ağırlıkları ile CBOW kullanılarak eğitilmiş ve veri analizinde Python kullanılmıştır.

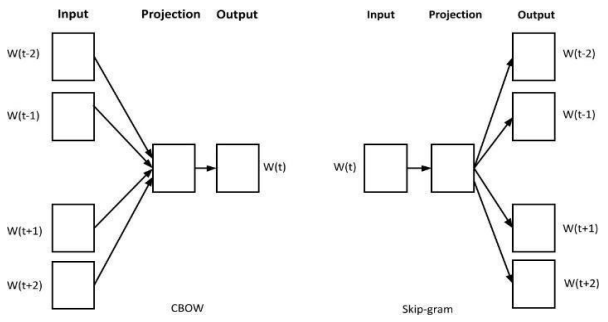
Sonrasında, kırılgan ve dezavantajlı gruplarla ilgili ifadeleri tespit edip, bu veri setinden kırılgan ve dezavantajlı gruplarla ilgili tespit edilen kelimelere en yakın diğer kelimeler belirlenmiştir.

Çalışmada kullanılacak veri insan olgusunun var olduğu her yerden seçilebilirdi. Ancak dil, insanlık için en sıkıştırılmış veri aktarım biçimlerinden biridir. Örneğin bir görselden çıkarılan bilgi, bilgisayarda kapladığı alanla kıyaslandığında eksik kalmaktadır. Bu nedenle metinlerin incelenmesi tercih edilmiştir. Ancak bu seçim, sorunun metinle ilgili olduğunu göstermez. Sorun daha ziyade, insan tarafından bırakılan izlerle ilgilidir.

Veri analizi yöntemi ve sonuçları aşağıdaki alt bölümlerde açıklanmıştır.

3.1 Yöntem

Word2Vec, Google araştırmacısı Thomas Mikolov ve arkadaşları, Grave ve arkadaşları tarafından geliştirilen, sözcükleri vektör uzayında ifade etmeye çalışan, dağılım hipotezine dayalı, denetimsiz (etiketsiz) ve tahmin tabanlı bir modeldir [17]. Bir başka deyişle, aynı metindeki kelimeler benzer anlamlara sahip olma eğilimindedir ve temelde yapay sinir ağı ile iki farklı model kullanarak kelimeleri eğitmeyi amaçlamaktadır.



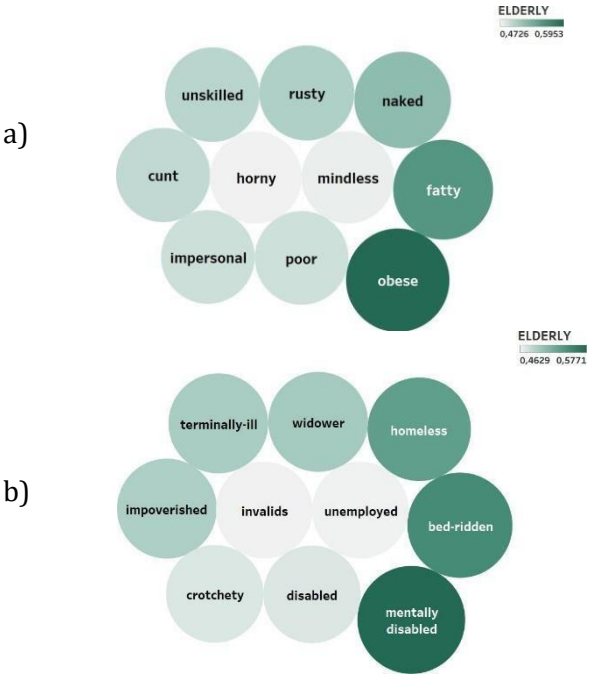
Şekil 1. Word2Vec tabanlı yönteme genel bakış

Word2Vec'in kullandığı iki model (Continuous Bag of Words (CBOW)) ve Skip-Gram Modelidir. CBOW (Continuous Bag of Words) modelinde, pencere boyutu merkezinde olmayan kelimeler girdi olarak alınır ve merkezdeki kelimeler çıktı olarak tahmin edilir. Skip Gram: Skip Gram modelinde pencere boyutu merkezinde bulunan kelimeler girdi olarak alınır ve merkezde bulunmayan kelimeler çıktı olarak tahmin edilir. Hiper parametrelerden biri pencere boyutudur (bkz. Şekil 1). Pencere boyutu, bir cümledeki mevcut ve tahmin edilen kelime arasındaki maksimum mesafedir. Diğer parametre ise gizli katmandaki düğüm sayısıdır ve bize kelimelerin kaç boyutlu bir uzayda temsil edildiğini gösterir.

Bu modeli kullanarak Facebook tarafından oluşturulan veri kümelerini kullanarak savunmasız ve dezavantajlı grupları tanımlayan belirli kelimeleri aradık. Araştırma bulguları aşağıda detaylandırılmıştır.

3.2 Analiz

Makalede, Facebook'un algoritmalarında kullandığı veri seti, kırılgan ve dezavantajlı gruplarla ilişkilendirilebilecek bazı kelimeler açısından incelenmiştir. Savunmasız ve dezavantajlı gruplar ifadesinin kapsamı kültürden kültüre, coğrafyadan coğrafyaya ve tarihsel perspektife dayanmaktadır. Ancak bu çalışmada, son zamanlarda yapılan bazı çalışmalara dayanarak nispeten yaygın ve genel olarak Kabul edilmeye yakın bazı kelimeler belirlenmiştir. Benzer bir sorunun ortaya çıkıp çıkmayacağını görmek için kelimeler hem Türkçe hem de İngilizce veri kümelerinde analiz edilmiştir. En genel haliyle ortaya çıkan sonuç, söz konusu kelimelerin her iki veri setinde de bazı sorunlu kelimelerle birlikte yer aldığıdır. Bu genel sonucun yol açabileceği toplumsal sorunlara kısaca değinmeden önce, özellikle bu kelimeler için elde edilen sonuçları detaylı olarak ele almakta fayda vardır.

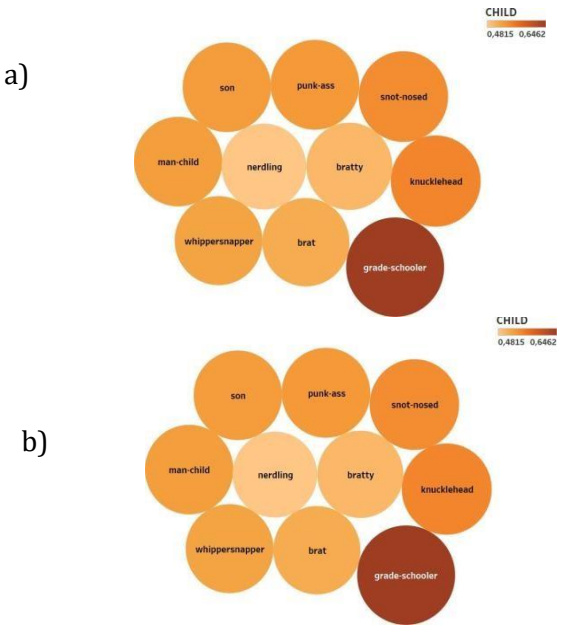


Şekil 2. "Elderly" kelimesinin Türkçe ve İngilizce veri kümelerindeki karşılaştırması
(a) Türkçe veri kümesinde yaşlı
(b) İngilizce veri kümesinde yaşlı

Türkçe veri kümesinde "yaşlı" kelimesinin en çok kullanıldığı en yakın kelimeler taranmıştır. En yaygın kelimeler şişman, yağlı, çıplak, paslı, vasıfsız gibi son derece olumsuz etiketlerden oluşmaktadır (Şekil 2a). Yaşlı bireylere yönelik bu olumsuz atıfların sosyo-kültürel açıdan yaşlanmaya ilişkin kalıp yargıları ortaya koyduğunu söylemek yanlış olmayacaktır. Kalıp yargılar, tanımladıkları olgular hakkında toplumdaki olumsuz ön kabuller hakkında ipuçları içermektedir. Bu bağlamda elde edilen sonuçların yaşlılıkla ilgili toplumsal ön kabulleri içerdiği açıktır. Ancak bu ön kabullerin yapay zekâ algoritmalarındaki varlığı, sosyal sorunların devamı açısından son derece sağlıksız sonuçlara yol açabilmektedir.

Türk veri setindeki çocuk kelimesine ilişkin sonuçlar (Şekil 3a), "yaşlı" kelimesindeki olumsuz stereotiplerin aksine, sosyal bir örüntü ortaya koymuştur: erkek evlat ve anne vurgusu. Ne yazık ki erkek evlat, bazı kültürlerde hala yüzyıllar öncesinin erkek egemen vurgusunu içermektedir. Burada karşımıza çıkan tablo, oğul, oğlum, oğluşum gibi vurgular nedeniyle oğula karşı pozitif ayrımcılığı çağrıştırması açısından bu durumun somut bir örneği olarak yorumlanabilmektedir. Çünkü bu kelimelerin sıklığı Kız, kızım, kızın gibi kelimeler oğluma kıyasla belirgin bir şekilde daha azdır. Benzer bir cinsiyet vurgusu, babadan çok çocukla

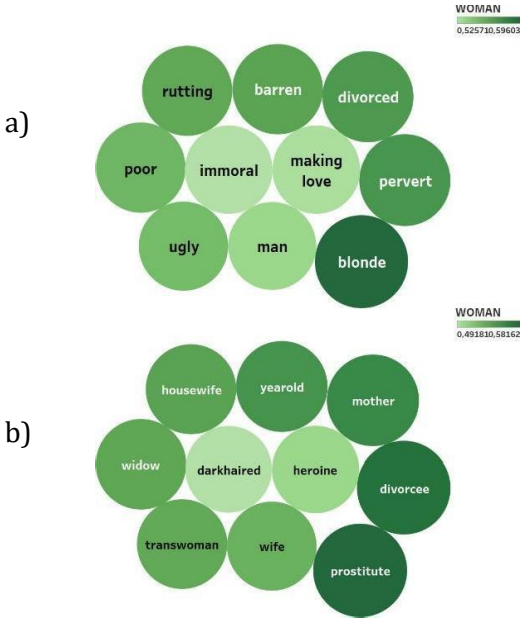
ilişkilendirilen anne kelimesinin kullanımında da görülmektedir. Mevcut tablo, çocuk ile anne arasındaki ilişkinin özdeşleştiğini ve baba rolünün geri planda kaldığını gösterir.



Şekil 3. "Child" sözcüğünün Türkçe ve İngilizce veri kümelerindeki karşılaştırması
(a) Türkçe veri setinde çocuk
(b) İngilizce veri setinde çocuk

Kadın kelimesine bakıldığında (Şekil 4a) hem olumsuz stereotiplerle hem de stereotipik cinsiyet vurgusuyla karşılaşılmaktadır. Boşanmış, kısır, azmış, sapkın gibi kelimeler kadının medeni durumuna ve kadın cinselliğine dair olumsuz vurgular içermektedir. Sarışın, çirkin, ahlaksız, fakir gibi olumsuz nitelemeler ise kadına yönelik fiziksel anlayışlar hakkında ipucu vermekte ve kadının anlamını ikili cinsiyet rollerine indirgeyerek ötekine, erkeğin olumsuzluğuna vurgu yapmaktadır.

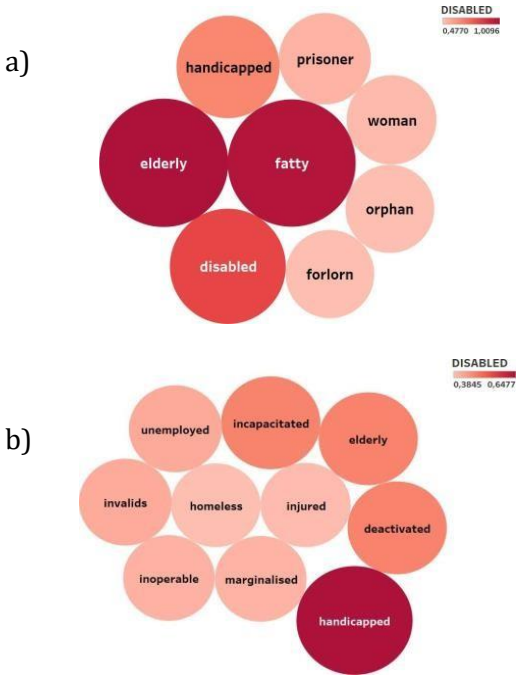
Yukarıda detaylı olarak ele alınan başlıklara ek olarak, Türkçe veri setinde incelediğimiz "engelli" kelimesi sıklıkla kusurlu ifadesiyle birlikte karşımıza çıkmıştır (Şekil 5a). Öte yandan yaşlılar, yetimler, evsizler gibi kişilere yapılan atıflar dikkat çekicidir. "Göçmen" kelimesine bakıldığında ise daha çarpıcı sonuçlarla karşılaşılmıştır. Irkçı, çete, ayyaş, serseri gibi son derece olumsuz nitelemelerin yanı sıra Müslüman, haçlı, ISIS (Ortadoğu'daki bir terör grubunu ifade eden IŞİD) gibi karmaşık din temelli tanımlamalara da rastlanmıştır. Söz konusu tablonun toplumun zimmilere bakış açısına dair rahatsız edici ipuçları içerdiğini söylemek mümkündür. Çok benzer bir durum, mülteci kelimesi için de geçerlidir.



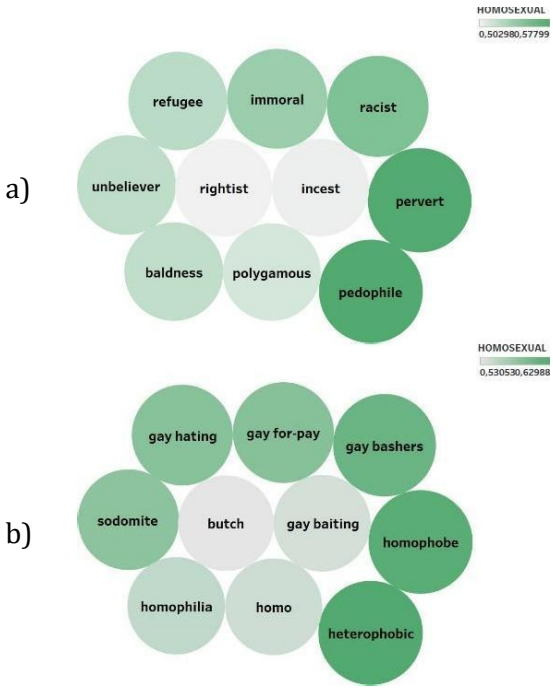
Şekil 4. "Woman" sözcüğünün Türkçe ve İngilizce veri kümelerindeki karşılaştırması
(a) Türkçe veri kümesinde kadın
(b) İngilizce veri kümesinde kadın

Eşcinsel, terörist, işgalci, yağmacı, engelli, faşist, barbar, köpek gibi tüm ifadeler mülteci kelimesiyle birlikte en sık kullanılan kelimeler arasında yer almaktadır. Mevcut durumda toplumun mültecilerle ilgili ciddi sorunları olduğu gözlemlenebilmektedir. Bu tablonun AI algoritmalarında kullanımında karşılaşılabilecek sorunların boyutu, verilerin herkese açık olduğu düşünüldüğünde endişe vericidir.

Hakaret içeren benzer bir tabloya "homoseksüel" kelimesinin metninde de rastlanmıştır (Şekil 6). Ayrıca "yetim, evsiz, hasta, yoksul" gibi kelimeler incelendiğinde kalıplaşmış vurguların hakaret içeren kelimelerden daha yaygın olduğu görülmektedir.



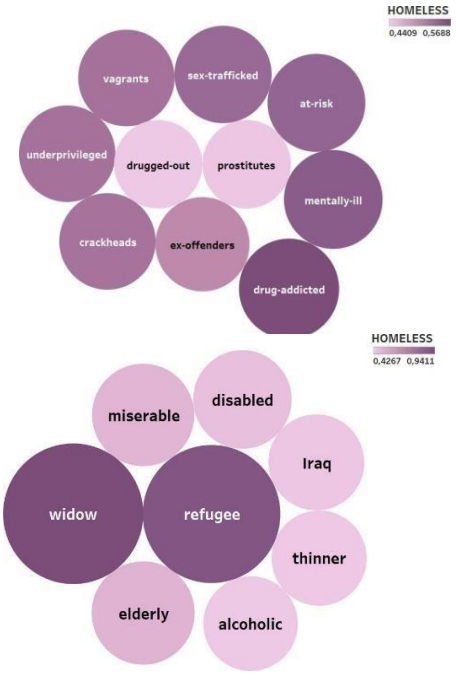
Şekil 5. "Disabled" sözcüğünün Türkçe ve İngilizce veri kümelerindeki karşılaştırması
(a) Türkçe veri setinde engelli
(b) İngilizce veri setinde engelli



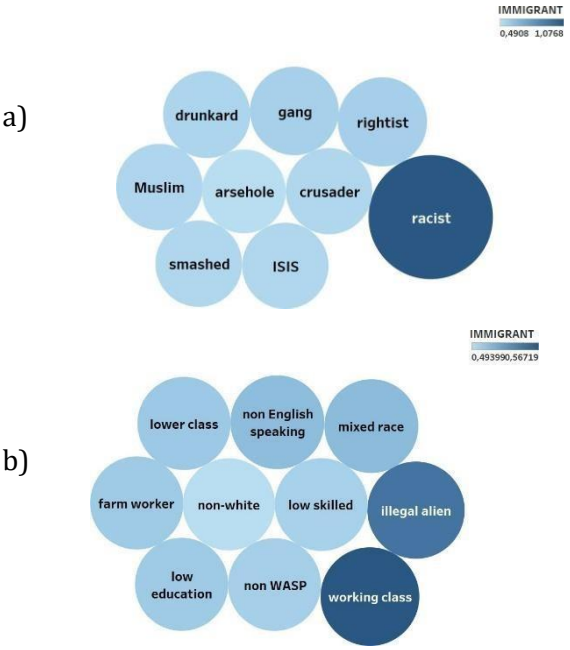
Şekil 6. "Homosexual" sözcüğünün Türkçe ve İngilizce veri kümelerindeki karşılaştırması
a) Türkçe veri setinde eşcinsel
b) İngilizce veri setinde eşcinsel

İngilizce veri setine bakıldığında (Şekil 2b), "yaşlı" kelimesi Türkçe veri seti ile benzer olumsuz kalıp yargılar sunmaktadır (Şekil 2). Benzer şekilde kadın kelimesi için de olumsuz kalıp yargıların yanı sıra cinsiyet vurguları açısından İngilizce veri seti ile Türkçe veri setinin çok benzer çıktılar ürettiği görülmüştür. Ancak İngilizce veri setindeki çocuk kelimesi, "oğlan ve oğul" gibi eril vurgular dışında Türkçe veri setinden farklılaşmaktadır. İngilizce veri setinde "child" kelimesi Türkçe veri setinden farklı olarak aşağılayıcı ve hakaret içeren nitelemeler içermektedir. Öte yandan "engelli" kelimesi her iki veri setinde de hakaret olarak tanımlanamasa da (Şekil 5) kalıplaşmış etiketlerden oluşan ifadeler içermektedir. "Göçmen ve mülteci" kelimeleri, Türkçe veri setindeki hakaret vurgusunun aksine (Şekil 8a ve Şekil 9a), İngilizce veri setinde sıklıkla aşağılayıcı tanımlamalarla kullanılmaktadır (bkz. sırasıyla Şekil 8b ve Şekil 9b).

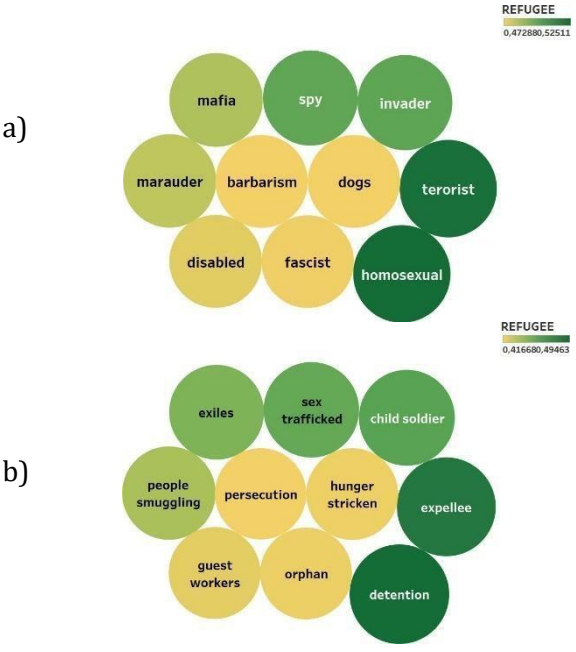
İngilizce veri setinde homoseksüel kelimesi sıklıkla homoseksüelliğe ilişkin basmakalıp ifadelerle birlikte kullanılmaktadır (Şekil 6b). Bununla birlikte, İngilizce veri setinde "yetim, evsiz ve yoksul" kelimeleri, Türkçe veri setinden farklı olarak (bkz. sırasıyla Şekil 11b, Şekil 7b ve Şekil 10b) son derece aşağılayıcı ifadelerle yer almaktadır (bkz. sırasıyla Şekil 11a, Şekil 7a ve Şekil 10a). Bu tabloya göre toplumun "evsizlere" karşı tutumunun son derece endişe verici olduğunu söylemek mümkündür. Benzer bir belirgin farklılık "hasta" kelimesi için de geçerlidir. İngilizce veri setinde, Türkçe veri setinden farklı olarak hasta kelimesi ağır hakaret içeren ifadeler içermektedir (Şekil 12). İngilizce veri setinde Müslüman (Şekil 13a) ve Afro-Amerikan (Şekil 13b) kelimelerine baktığımızda, İslam'a yönelik hakaret içeren ifadelere sıklıkla rastlanırken, Afro-Amerikan'a yönelik daha az saldırgan ifadeler gözlemlenmektedir. Son olarak, "terör" kelimesi çok açık bir şekilde İslam ile ilişkilendirilmekte ve bu kelime İslam'a karşı saldırgan ifadelerle birlikte kullanılmaktadır.



Şekil 7. “Homeless” sözcüğünün Türkçe ve İngilizce veri kümelerindeki karşılaştırması
(a) Türkçe veri setindeki evsiz
(b) İngilizce veri setindeki evsiz

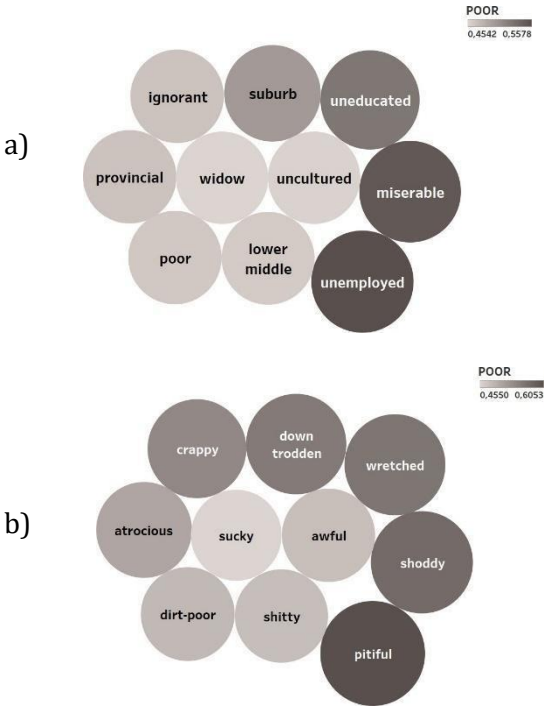


Şekil 8. "Immigrant " sözcüğünün Türkçe ve İngilizce veri kümelerindeki karşılaştırması
(a) Türkçe göçmen veri kümesi
(b)İngilizce göçmen veri kümesi



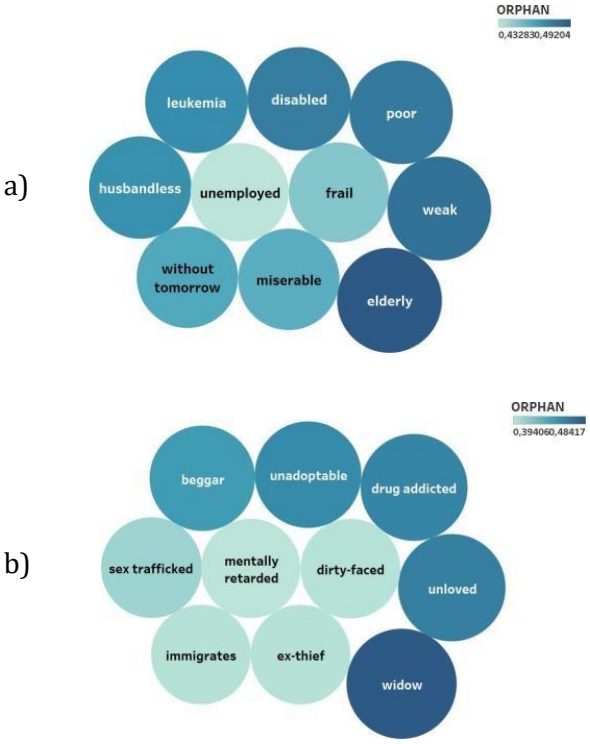
Şekil 9. "Refugee" sözcüğünün Türkçe ve İngilizce veri kümelerindeki karşılaştırması
a) Türkçe veri setinde mülteci
b) İngilizce veri setinde mülteci

Özetlemek gerekirse, Türkçe veri kümesi özellikle göçle ilgili terimler açısından son derece saldırgan içerikler barındırmaktadır. Öte yandan, kelimelerin hastalık veya yoksulluk gibi sıkıntılı durumları çağrıştırdığı durumlarda hakaretten ziyade basmakalıp ifadeler içerdiği gözlemlenmiştir. İngilizce veri setinde de benzer bir özetleme yapacak olursak, hakaretten ziyade azınlık gruplara yönelik aşağılayıcı bir yaklaşım göze çarpmaktadır; evsiz, hasta, yetim gibi yoksunluk ve sıkıntı anlatan kelimelerde son derece rencide edici ve aşağılayıcı ifadelerle karşılaştığımızı söyleyebilmekteyiz. Başka bir deyişle, Türkçe veri setinin hasta, yaşlı, evsiz vb. kırılgan durumlar için biraz daha az saldırganlık içerdiğini göstermektedir.

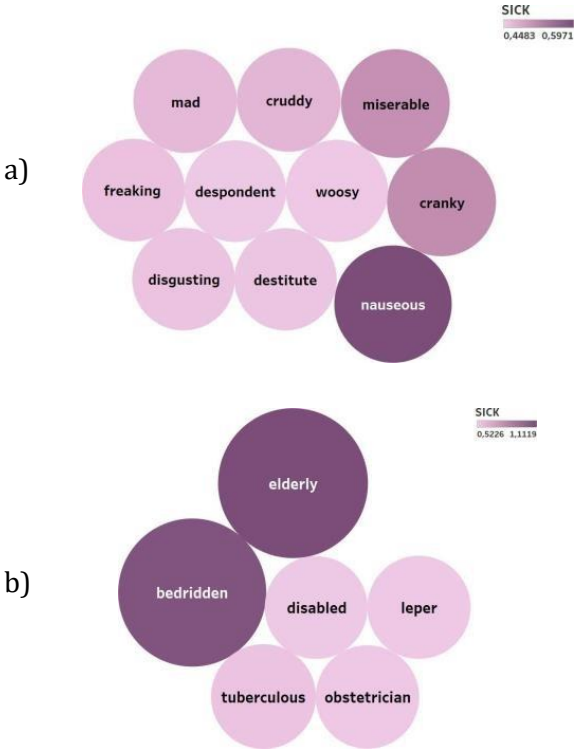


Şekil 10. "Poor " sözcüğünün Türkçe ve İngilizce veri kümelerindeki karşılaştırması
a) İngilizce veri setinde fakir
b) Türkçe veri setinde fakir

Bu durum, toplumsal düzeyde bu tür kırılgan durumlar ve insanlar için daha az öfke ve daha fazla acıma olduğunu gösterebilmektedir. Ancak Türkiye veri setinde göçmen azınlıklar için durum çok daha tedirgin edicidir. Bu durum göçün ve toplumsal huzursuzluğun etkilerinin toplumsal düzlemde güçlü bir şekilde hissedildiğine işaret etmektedir. Öte yandan, İngiliz veri setine göre, azınlık unsurları sosyal düzeye karşı nispeten daha az tedirginliğe neden olmaktadır.

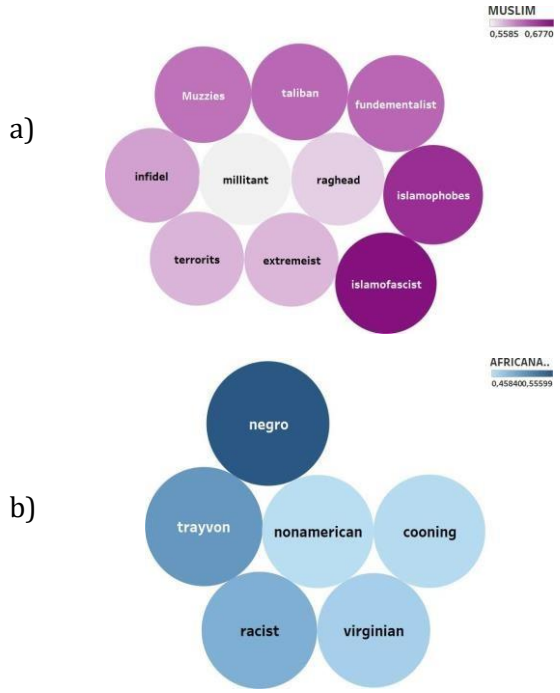


Şekil 11. "Orphan" sözcüğünün Türkçe ve İngilizce veri kümelerindeki karşılaştırması
(a) Türkçe veri setinde yetim
(b)İngilizce veri kümesinde yetim



Şekil 12. "Sick" sözcüğünün Türkçe ve İngilizce veri kümelerindeki karşılaştırması

a) Türkçe veri setinde hasta
b) İngilizce veri setinde hasta



Şekil 13. İngilizce veri kümelerinde "Muslim" ve "Afro-American" kelimelerinin karşılaştırması

a) İngilizce veri kümesinde Müslüman kelimesi
b) İngilizce veri kümesinde Afro-Amerikalı kelimesi

Bu durum, daha önce de belirtildiği üzere, yapay zekâda nispeten yaygın olarak kullanılan veri setlerinin, özellikle savunmasız ve dezavantajlı gruplar için genel olarak tehlikeli olabilecek önyargılı ifadeler içerdiğini göstermektedir. Ancak bunlar kültürden kültüre değişiklik gösterebilmektedir. Beklenen şudur ki mevcut durum, bu veri setleri kullanılarak yapay zekâ kullanımı açısından toplumun entegrasyonunda karşılaşılabilecek sorunların öngörülmesine katkı sağlayacaktır. Bu bölümde ortaya koyduğumuz söz konusu tehlike, bir sonraki bölümde mevcut verilerle oluşturulan akıllı sistem ile televizyon verilerini alan kitlelerde meydana gelen sosyal yıkım arasında bir analoji kurularak etik yönleriyle açıklanacaktır.

4 Gerbner'in Yetiştirme Perspektifinden Yapay Zekâ'da Yanlılık Problemi

Bu bölümde, gelmiş geçmiş en yaygın kitle iletişim teorisi olan Gerbner ve arkadaşlarının yetiştirme teorisi, yapay zekâ perspektifinde yeniden yorumlayarak, makine öğreniminin kullandığı

veriler nedeniyle yol açabileceği olası sosyal ve kitlesel sorunlara dikkat çekilmektedir [18]. George Gerbner tarafından 1960'lı ve 1970'li yıllarda kavramsallaştırılan yetiştirme teorisi, televizyonun bireylerin sosyal gerçeklik algıları üzerinde önemli etkileri olduğunu savunmaktadır. Teori, 1956-2000 yılları arasında iletişim bilimlerinde en çok atıf yapılan üç kaynaktan biridir [19]. Braynt ve Miron 1993-2005 yılları arasında yayınlanan 16 akademik dergide de en çok atıf yapılan teoridir [20]. Teori, medyadaki her teknolojik gelişmeyle birlikte sorgulanmıştır. Son altmış yılda, kitle iletişim araçları alanı kablo, uydu, video oyunları ve en son olarak da sosyal medyanın yayılmasına tanık olmuştur. 2000 yılından bu yana 125'in üzerinde çalışma, teorinin sürekli değişen medya ortamına uyum sağlama kabiliyetine işaret ederek teoriyi desteklemiştir. Ancak bugüne kadar, yetiştirme kuramının mevcut kitle iletişim teknolojileri üzerindeki etkilerine dair yapılan tüm çalışmalar, bu teknolojilerin bireyler üzerindeki etkilerine odaklanmıştır. Bu çalışmada söz konusu kurama farklı bir bakış açısı tartışmaya açılmaktadır.

Televizyon çağında kitlesel mesajlar karşısında pasif olarak konumlandırılan bireyler, yapay zekâ çağında kitlesel mesajlardan bizzat sorumlu konumdadır. Öte yandan bir bakıma televizyon çağındaki pasif izleyicilerin yerini yapay zekâ teknolojileri almış durumdadır. Başka bir deyişle, televizyon mesajlarıyla kitleler üzerine inşa edilen kültürel ekosistem Gerbner'in deyimiyle "ekilerek" günümüzün yapay zekâ modellerinin eğitimiyle inşa edilmektedir.

Yetiştirme teorisi, televizyon izlemenin insanların dünyaya bakışını etkilediğini ve tipik olarak risk ve güvensizlik duygusunun artmasına neden olduğunu iddia etmektedir [21]. Teori birbiriyle ilişkili iki önermeye dayanmaktadır: (1) televizyon programları gerçek dünyanın tutarlı ancak çarpıcı biçimde çarpıtılmış bir görünümünü temsil eder (2) bu tutarlı ve formüle edilmiş çarpıtılmış temsillerin aşırı izlenmesi izleyicilerin dünya görüşlerinin şekillenmesine aracılık eder [22],[23],[24]. Başka bir deyişle, Gerbner araştırmasında içerik analizi yöntemini kullanarak medya üzerinden bir kültürü analiz etmeye çalışmakta ve televizyondan aktarılan kültürel değerlerin topyekûn bir yanılgı yarattığını iddia etmektedir. Ona göre televizyon; tutumları eker, geliştirir ve popülasyonun günlük kültürünün ortak kaynağını oluşturur. İstenilen ortamın ya da bireyin oluşması uzun zaman alır. Sürecin sonunda televizyon aracılığıyla yapay bir kültür yaratılmış olur.

Tüm bunlardan yola çıkarak, çalışmamız özelinde verinin kültür ve siyaset ile iki şekilde ilişkilendirilebileceği düşünülmektedir. Birincisi, toplumun gündelik hayatında ve kültüründe var olan şeylerin internet ortamında tekrar tekrar üretilerek veri yığınlarına dönüşmesi ve bu verilerle beslenen yapay zekâ algoritmaları aracılığıyla tekrar toplumun kullanımına sunulması. İkincisi ise gündelik hayatın ve kültürün içinde olmayan şeylerin internet ortamının sağladığı özgürleşme ve sanallık aracılığıyla kurgusal formlarda veri yığınlarına dönüşmesi ve bu verilerle beslenen yapay zekâ algoritmaları aracılığıyla toplumsal sürece dahil edilerek kültürel dönüşüme yön vermesidir. Bu, veri merkezli bir toplumsallaşma biçimi oluşturmaktadır.

Gerbner'in araştırmasını televizyon etrafında şekillendirmesi de benzer bir nesnel nedenden kaynaklanmaktadır. Dönemin tekno-kültürel yapısı dolayısıyla diğer kitle iletişim araçlarından farklı olarak bireylerin araştırma yapmalarına ya da dışarı çıkmalarına, iletişim ve bilgi edinme için efor sarf etmelerine gerek yoktur, çünkü televizyon onları evin tam ortasında beklemektedir. Televizyon ailenin bir üyesi olarak sürekli sabırla ve ısrarla hikâyelerini anlatır. Günlük yaşamlarında fiziksel ve sosyal çevrelerinin sadece küçük bir kısmıyla temas halinde olan bireylerin sosyal gerçeklik algısı, doğrudan deneyimsel uyaranlar olmaksızın duyduklarına, gördüklerine ve okuduklarına dayanan temsili deneyimlerden büyük ölçüde etkilenir [7]. Bireylerin gerçeklik imajı medyanın sembolik ortamına tabi hale geldikçe, sembolik ortamın bireyler üzerindeki etkisi de artmaktadır [25]. Özetle Gerbner'in teorisi, döneminin tekno-kültürel yapısı ve bunun artan şiddet ve ayrımcılık gibi somut göstergeleri ile yakından ilişkilidir.

Günümüze bakıldığında, benzer tekno-kültürel göstergeler dikkat çekicidir (Bkz. Coded Bias Belgeseli, Amazon işe alım robotu, vb.) Dünya veri merkezli bir sosyalliğe yöneldikçe, bir "ağ toplumuna" dönüştükçe, medyanın kamusal söylemi şekillendirmede oynadığı rol giderek daha önemli hale gelecektir [26]. Çevrimiçi kanallar ve 24 saat yayın yapan kablolu haber kanalları sayesinde medyanın yaygınlaşması da bu rolün önemini etkilemektedir. 'Ağ toplumu'nun

en etkili ve faydalı tanımlarından biri, bu terimi ayrılmaz bir şekilde dijital ağların yükselişiyle özdeşleştirmekte ve 'ağ toplumunun oluşumunu' ifade etmektedir. Güç ilişkilerinin uygulanması, küresel dijital iletişim ağlarının, zamanımızın temel gücü ve sembol işleme sistemi olarak yükselişiyle kararlı bir şekilde dönüşüme uğramıştır [27]. Sonuç olarak, medya önyargılarının söylemi etkileme kapasitesi her zamankinden daha güçlüdür. Yani internet ortamı bir yandan tutumlarımızı oluşturup pekiştirirken [28]; diğer yandan, özellikle "anonim" olma lüksü ile kendimizi sınırlandırmaya gerek duymadan bu tutumlar doğrultusunda dijital ayak izi bırakabileceğimiz bir zemin sağlamaktadır. Anonimliğin siber saldırganlık için bir tetikleyici olduğu bilinmektedir [29]. Dolayısıyla dijital ekosistemde bıraktığımız ayak izi, gerçek hayatta olabileceğinden daha radikal ve saldırgan tutumlar içerebilmektedir.

Ayrımcılık ve şiddet, televizyon mesajlarının bir sonucu olarak toplumsal tutum ve söylemleri şekillendiren sorunlardır. Bu eğilim, akıllı şehirler, araçlar ve fabrikalar gibi otonom sistemleri oluşturan veri ve algoritmalarda da mevcuttur. Bunu bilmemizin sebebi is ağ toplumunun yaygın söyleminde yer alan şiddet unsurlarına ilişkin çok sayıda kanıt bulunmasıdır [30]. Şiddet unsurlarıyla dolu verilerin adil olup olmadığı ya da hangi verinin hangi bağlamda daha doğru kullanılabileceğine ilişkin düzenlemeler yapılmadığı sürece bu tür akıllı ekosistemlerin günümüzün kültürel sorunlarına ev sahipliği yapması kaçınılmaz olacaktır. Çünkü kendi tohumlarımızla büyüttüğümüz bir yapının bizden izler taşımaması mümkün değil. Bu anlamda her zaman bizden bir parça taşıyacak olan yapay zekânın da bir çocuğu doğru yetiştirmek kadar hassasiyetle yetiştirilmesi uzun vadede sağlıklı bir toplumsal yapı için elzem görünmektedir.

Gerbner'den esinlenen yetiştirme teorisinin günümüzdeki versiyonunda insanlık, tohumların ekildiği zihinler değil, dijital ayak izleriyle tohumlarını eken ve saçan konumundadır. Bu benzetme, daha önce ekilen tohumların, yani bireylerin büyümüş, olgunlaşmış olmasını gerektirmektedir. Her ne kadar medyanın yanlı ve sağlıksız mesajları ile şekillenmiş olsa da yetiştirilecek tohumlardan verim alabilmek, büyürken yaşanan sorunların önüne geçebilmek ve verileri güçlendirebilmek için tohumların yani verilerin sağlıksız olanlarını eleyerek ilerlemek büyük önem arz etmektedir. Günümüz tekno-kültürel medeniyetine ekilen tohumların zaman içerisinde istenilen ürünü vermesi ancak böylelikle mümkün olacaktır.

Gerbner'in bu teorisi temelde televizyonun insanların ailesinden bir anlatıcı olduğunu ve buna bağlı olarak televizyondaki yayınların düşünce ve bilinç üzerinde bir tekel oluşturduğunu ifade etmektedir. Kısacası Gerbner, "çok televizyon izlemek insanı aptallaştırır" sözünü bilimsel olarak bu kuramla açıklamaktadır. Aynı durum veriler için de geçerlidir. Çoklu perspektiflerden gelen çok fazla ve organize edilmemiş veri, akıllı olmayan sistemler ve insanlar yaratacaktır. Bu nedenle verilerin düzenlenmesi ve standartlaştırılması bugün ve gelecekte hayati önem taşımaktadır. Dahası, sosyal bilimler ve bilgisayar bilimlerinin bu prosedürlerin uygulanması odağında sıkı bir şekilde entegre edilmesi gerektiği düşünülmektedir. Son olarak, bu standardizasyonun denetlenebilir olması, hayati önem taşıyan bir diğer husustur.

5 Tartışma ve Sonuç

Son yıllarda, algoritmalardaki hatalardan kaynaklanan ırkçılık, cinsiyetçilik, yaş ayrımcılığı gibi sosyal sorunlar dikkat çekici hale gelmiştir, çünkü bu tür sistemler hayat değiştiren, kritik kararların alındığı birçok güvenlik açısından önemli ortamda kullanılabilmektedir. Bu nedenle, birçok yeni yaklaşım önerilmiş ve bazı yaklaşımlar değiştirilmiştir. Bu çalışma, algoritmaların önemli noktalarda önemli kararlar almak için kullanıldığı günümüzde, verilerden bağımsız olarak değerlendirilmesinin sorunlu sonuçlara neden olabileceği fikrini ortaya koymaktadır. Algoritma hatalarından kaynaklanan sorunların çözümü için iki bakış açısı sunulmaktadır. Birincisi, bütünsel odaklı bir yaklaşımın önemi, ikincisi ise disiplinler arası bir değerlendirmenin önemi.

Bu bağlamda, öncelikle fonksiyonun kendisini bir veri olarak tanımlanmış ve her iki perspektifi de ortaya koymak için Facebook veri seti kullanılarak bir deney gerçekleştirilmiştir. Deneyde, sorunu daha kapsamlı bir şekilde ortaya koyabilmek için hem Türkçe veri setinde hem de İngilizce veri setinde kırılgan ve dezavantajlı grupları çağrıştıran bazı kelimeleri araştırılmıştır. Bu noktada amaç, olması gereken veriden ziyade var olan veriyi ve bu verinin potansiyel sorunlarını göstermek olmuştur.

Sonuçlar, veri kümelerinde aranılan anahtar kelimelerin genellikle diğer yüksek derecede saldırgan ve damgalayıcı kelimelerle birlikte bulunduğunu göstermektedir. Sonuçlar, sosyal bilimsel bir yaklaşımla yorumlandığında, mevcut sorunun yeni bir sorun olmadığı sonucuna varılmış ve söz konusu eski sorunun mevcut teknolojik araçlarla farklı bir bakış açısıyla ele alınabileceği gösterilmiştir. Gerbner'in yetiştirme ve kültürleme kuramında ortaya koyduğu insan ve televizyon etkileşimine bağlı olarak ortaya çıkan sorunların günümüz yapay zekâ teknolojilerinde yapay zekâ ve insan etkileşimi için hala güncelliğini koruduğunu, etkileyen ve etkilenen nesneler değiştirilerek ortaya koyulmuştur. Sonuç olarak, olası sosyo-kültürel sorunların gözlemlenebilmesi ve önlenebilmesi için "algoritma ve veri çiftinin bütünsel olarak ele alınması" ve "bu tip araştırmaların disiplinler arası bir bakış açısıyla gerçekleştirilmesinin" önemi vurgulanmıştır.

Dolayısıyla çalışmanın katkılarından biri de tespit edilen sorunun anlaşılması ve çözülmesi noktasında mevcut toplumsal bağlamla uyumlu bir bakış açısı geliştirmenin önemini ortaya koymasıdır. Son olarak, mevcut durumun ciddiye alınmaması halinde, aşılması zor büyük ölçekli toplumsal sorunların ortaya çıkabileceği öngörülmüştür. Gelecekte yapılacak bir çalışma olarak, tanımlanan soruna çözüm üretmek amacıyla disiplinler arası çalışmalar yürütebilecek bilim insanlarının akademik üretkenliğini yönlendirmek için böyle bir alanın gerekliliklerinin açıklanması ve ana hatlarıyla ortaya koyulması amaçlanmaktadır. Çalışmanın, sosyal bilimler ve bilgisayar bilimlerinin entegrasyonuna ilişkin önyargıların yıkılmasına aracılık etmesi ve böylelikle bu alana yatırım yapanların çok sayıda maddi ve manevi engelin üstesinden gelmelerine katkıda bulunması beklenmektedir.

6 Bildirimler

Yazarlar herhangi bir çıkar çatışması yaşamadıklarını beyan etmişlerdir.

Kaynaklar

[1] Zhang JM, Harman M. "Ignorance and Prejudice" in Software Fairness", 2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE), IEEE, 1436–1447, 2021.

[2] Aho AV, Lam MS, R. Sethi R, Ullman JD. Compilers: Principles, Techniques, and Tools (2nd Edition), Addison-Wesley Longman Publishing Co., Inc., USA, 2006. Erkek C, Ağıralioğlu N. *Su Kaynakları Mühendisliği*. Altıncı baskı. İstanbul, Türkiye, Beta, 2010.

[3] Wu C, Thompson ME, Sampling theory and practice, Springer, 2020Li RTH, Chung SH. "Digital boundary controller for single-phase grid-connected CSI". *IEEE 2008 Power Electronics Specialists Conference*, Rhodes, Greece, 15–19 June 2008.

[4] C. Hertweck, C. Heitz and M. Loi, "On the Moral Justification of Statistical Parity", Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21), 747–757, 2021a.

[5] Internet: Github, https://github.com/,01.2022.

[6] Bitbucket, https://bitbucket.org/,01.2022.

[7] G. Gerbner, "Cultivation Analysis: An Overview", Mass Communication and Society 1(3-4),175–194, 1998.

[8] S. Galhotra, Y. Brun and Meliou A, "Fairness Testing: Testing Software for Discrimination", Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering (ESEC/FSE 2017), 498–510, 2017.

[9] A. Aggarwal, P. Lohia, S. Nagar, K. Dey and D. Saha, "Black Box Fairness Testing Of Machine Learning Models, Proceedings of the 2019 27th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE 2019), 625–635, 2019.

[10] S. Udeshi, P. Arora and S. Chattopadhyay, "Automated Directed Fairness Test- Ing", Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering (ASE '18), 98–108, 2018.

[11] C. Hertweck, C. Heitz and M. Loi, "On the Moral Justification of Statistical Parity", Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21), 747–757, 2021b.

[12] S. A. Friedler, C. Scheidegger and S. Venkatasubramanian, "On the (im) Possibility of Fairness", arXiv preprint arXiv:160907236, 2016.

[13] D. Cirillo, S. Catuara-Solarz, E. Günay, L. Subirats, S. Mellino, A. Gigante, A. Valencia, M. J. Rementeria A. S. Chadha and N. Mavridis, "Sex and Gender Differences and Biases in Artificial Intelligence for Biomedicine and Healthcare", npj Digital Medicine 3, 81, 2020.

[14] S. Benthall and B. D. Haynes, "Racial Categories in Machine Learning", Proceedings of the Conference on Fairness, Accountability, and Transparency

(FAT* '19), 289–298, 2019.

[15] M. Kearns, S. Neel, A. Roth and Z. S. Wu, "An Empirical Study of Rich Subgroup Fairness for Machine Learning", Proceedings of the Conference on Fairness, Accountability, and Transparency, 100–109, 2019.

[16] A. Joulin, E. Grave, P. Bojanowski and M. Douze, "Fasttext.zip: Compressing Text Classification Models", arXiv preprint arXiv:161203651.

[17] E. Grave, P. Bojanowski P, Grupta, A. Joulin and T. Mikolov, "Learning Word Vectors for 157 Languages", Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), 2018.

[18] G. Gerbner, L. Gross, M. Morgan and N. Signorielli, "Living with Television: The Dynamics of the Cultivation Process", Perspectives on Media Effects, 17–40, 1986.

[19] J. Bryant and D. Miron, "Theory and Research in Mass Communication", Journal of Communication, 2004.

[20] W. J. Potter and K. Riddle, "A Content Analysis of the Media Effects Literature", Journalism & Mass Communication Quarterly, 84(1), 90–104, 2007.

[21] G. Gerbner and L. Gross, "Living with Television: The Violence Profile", Journal of Communication, 26(2), 172–199, 1976.

[22] J. Shanaha and M. Morgan, "Television and its Viewers: Cultivation Theory and Research", Cambridge University Press, 1999.

[23] L. J. Shrum, "Media Consumption and Perceptions of Social Reality: Effects and Underlying Processes", Lawrence Erlbaum Associates, Media Effects, Advances in Theory and Research, 69-96, 2002.

[24] M. Morgan, J. Shanahan and N. Signorielli, "Yesterday's New Cultivation, Tomorrow" Mass Communication and Society 18(5), 674–699, 2015.

[25] A. Bandura, "Social Cognitive Theory of Mass Communication" Media Psychology, 3(3), 265–299,
2001.

[26] M. Castells and G. Cardoso, "The Network Society: From Knowledge to Policy", Center for Transatlantic Relations, Johns Hopkins University-SAIS, 2006.

[27] M. Castells, M. Fernandez-Ardevol, J. L. Qiu and A. Sey, "Mobile Communication and Society: A Global Perspective" Mit Press, 2006.

[28] M. Castells, "The Impact of the Internet on Society: A Global Perspective", Article from the book Change: 19 Key Essays on How the Internet Is Changing Our Lives, 127–148, 2014.

[29] P. K. Smith, "Cyberbullying and Cyber Aggression", Handbook of school violence and school safety, Routledge, 111–121, 2012.

[30] D. Helbing, "Societal, Economic, Ethical and Legal Challenges of the Digital Revolution: From Big Data to Deep Learning, Artificial Intelligence, and Manipulative Technologies", arXiv:1504.03751,2015.