

Çevrim İçi Nefreti Önlemeye Yönelik Düzenlemelerin Sosyal Ağların Teknolojik Biçim ve İşleyişi Bağlamında Değerlendirilmesi

Abdulkadir ATİK*

Öz

Kullanıcı sayıları milyarları aşan sosyal ağlar hem kamusal bir tartışma platformu hem de başat bir sosyal etkileşim ortamına dönüşmüş durumdadır. İletişim teknolojilerinin sağladığı olanaklar aynı zamanda bazı toplumsal gerilim ve çatışmaların da her zamankinden daha fazla görünür olmasına neden olmuştur. Sosyal ağların dışı vurumu kolaylaştıran yönleri, anonimliği, dolaysızlığı ve küresel doğası, aşırılık yanlıları ve nefret tacirleri için son derece elverişli bir ortam oluşturmaktadır. Ağlardaki nefret gruplarının sayısının katlanarak artması nefret söyleminin toplumsal barış ve huzuru bozarak bireysel hak ve özgürlükleri tehdit edecek aşamaya gelmesine neden olmuştur. Bu durum kurum ve kuruluşları daha kapsamlı, çok taraflı ve uzun vadeli düzenlemeler yapmaya zorlamaktadır. Bu çalışmada çevrim içi nefret söylemini önlemeye yönelik güncel dijital düzenlemelerin neler olduğu ortaya konularak bu düzenlemelerin ağların dijital döngüsü içinde uygulanabilirliği sorgulanmaktadır. Literatür taraması yöntemiyle kavramsal düzeyde ele alınan çalışma; teknoloji, insan ve toplum arasındaki olgusal ilişkileri yorumsamacı bir yaklaşımla ele almaktadır. Günümüzde hâlihazırda nefret içeriklerinin tespit edilmesi, yayından kaldırılması ve suç teşkil eden içeriklerin adli mercilere bildirilip suçun ve suçlunun kovuşturulması için geliştirilmiş teknolojik denetleme ve gözetim algoritmaları yaygın olarak kullanılmaktadır. Fakat bu algoritmalar gerek teknolojinin kendi doğasından kaynaklanan yetersizlikler gerekse de nefret söyleminin hukuki ve kültürel anlamda henüz daha netleşmemiş kavramsal yapısından dolayı soruna çözüm olamamaktadır.

Anahtar kelimeler: Nefret Söylemi, Sosyal Ağ, Kamusal Alan, Yabancı Düşmanlığı, Ayrımcılık

* Doç. Dr., Atatürk Üniversitesi İletişim Fakültesi Radyo, Sinema ve Televizyon Bölümü, kadiratik@atauni.edu.tr

Atik, A. (2022). Çevrim içi Nefret Düzenlemelerinin Sosyal Ağların Teknolojik Biçim ve İşleyişi Bağlamında Değerlendirilmesi. TRT Akademi, 7 (15), 434-449. DOI: 10.37679/trta.1097019

Derleme Makalesi

Geliş Tarihi: 01.04.2022

Revizyon Tarihi: 20.05.2022

Kabul Tarihi: 27.05.2022

ORCID ID: 0000-0002-3843-3472 DOI: 10.37679/trta.1097019

Evaluation of Regulations for the Prevention of Online Hate in the Context of the Technological Form and Operation of Social Networks

Abdulkadir ATİK

Abstract

This study aims to research on the concepts of hate speech and hate crime that can be viewed in social life and in the media. A conceptual perspective will be presented on what kind of expression hate speech, which has just been introduced to terminology, is, what it aims at and how it is related to freedom of expression, and these areas of discussion will be analyzed in the context of the film Wolves Empire. In the first part, the conceptual analysis of the concepts of hate speech and hate crime, the points that separate hate speech and freedom of expression as it constitutes a frequently encountered area of discussion, and the role of media tools in hate speech will be examined. In the context of the movie, hate speech in the cinema will be discussed. The study will be examined with the method of critical discourse analysis.

Keywords: Hate Speech, Hate Crime, Hate Speech in the Media, Racist Hate Speech, Empire of the Wolves

Review Paper

Recieved: 01.04.2022

Revised: 20.05.2022

Accepted: 27.05.2022

1. Giriş

Nefret söylemi, temel insan hak ve ideallerinden biri olan ifade özgürlüğü ile çoğulcu demokratik bir toplumun eşitlikçi ve insan onuruna yakışır düzenlemelerine ve bireylerin, grupların, azınlıkların diğer haklarına zarar verme potansiyeli arasında sert tartışmalara konu olan, çağımızın önemli sorunlarından biridir. Kavram; şiddeti, düşmanlığı ve ayrımcılığı teşvik etmesi bağlamında ulusal ve uluslararası hukukta net ve ortak vurgulara sahip olsa da ön yargı ve hoşgörüsüzlük iklimi oluşturma noktasında belirsizleşerek, geçerliliğini yitirebilmektedir (Gagliardone, Gal ve diğerleri, 2015, s. 10). Bilgi ve iletişim teknolojilerinin milyarlarca insanı bir araya getirebilecek bir aşamaya ulaşması ve sosyal etkileşimin ana bileşenine dönüşmesiyle birlikte nefret söylemi, her zamankinden daha fazla tartışılan, çözüm öncelikli bir meseleye dönüşmüştür. Medyanın hem yeni söylemler oluşturabilme hem de var olan söylemleri yeniden üretip meşrulaştırabilme gücü, medya merkezli çözüm önerilerini gündeme getirmiştir.

Öte yandan nefret söylemi; hukuk, iletişim, sosyoloji, siyaset bilimi gibi çeşitli bilim alanlarında çok sayıda araştırmaya konu olmuştur. İlgili literatürde konunun ilk dönem çalışmalarında sosyolojik, felsefi ve politik bağlamda daha çok hukuki düzenlemeler çerçevesinde ele alındığı görülmektedir. Son dönem çalışmalarda ise kavram; sosyal ağlar ve dijital düzenlemeler eksenindeki araştırma ve incelemelere dâhil edilmiştir. Bu eğilimde sosyal ağların, her zamankinden daha fazla etkileşime olanak sağlayan ve kamusal tartışmalara zemin oluşturan ortamlara dönüşmesinin etkili olduğu söylenebilir. Bunun yanı sıra ağların, toplumsal sorunlardan ve gerilimlerden beslenen ve bir arada yaşamayı sorgulayan grupların ve sosyal yapıların kavşak noktası hâline gelmesi de bilimsel araştırmaların bu alanda yoğunlaşmasına neden olabilmektedir (legifrance.gouv.fr, 2015).

Bu kapsamda çalışmada çevrim içi nefret söylemini önlemeye yönelik güncel düzenlemelerin neler olduğu ortaya konularak bu düzenlemelerin ağların dijital döngüsü içinde uygulanabilirliği sorgulanmaktadır. Literatür taraması yöntemiyle kavramsal düzeyde ele alınan çalışma; teknoloji, insan ve toplum arasındaki olgusal ilişkileri yorumsamacı bir yaklaşımla ele almaktadır. Literatür taraması, konuyla ilgili temel ve güncel çalışmaların bulgularını, yaklaşım ve bakış açılarını bir araya getirip yeni bir senteze ulaşmayı hedefleyen derleme türündeki çalışmalarda tercih edilen bir yöntemdir (Fink, 2010, s. 3). Yöntem özellikle disiplinler arası çalışmalarda genel bir bakış açısıyla yeni teorik çerçeveler ve kavramsal modeller oluşturabilme olanağı sağlayabilmektedir (Tranfield, Denyer ve Smart, 2003, s. 207-208). Yalnızca kamu otoritelerini, servis sağlayıcı şirketleri, kullanıcı-

ları veya sivil toplum kuruluşlarını baz alan tek taraflı düzenlemelerin üstesinden gelemeyeceği ölçekte bir sorun olan nefret söylemi (Banks, 2010, s. 234), tüm tarafların katılımının esas alındığı, uluslararası uzantılıları olan ve ifade özgürlüğünü gözetemeyen mücadele yöntemlerini zorunlu kılmaktadır. İnternetin ulusal müdahaleleri yetersiz kılan küresel yapısı, hayatın her alanına yansıyan kuşatıcı ve baskın teknolojik doğası, sorunun çözümünde adli denetim mekanizmalarının yanı sıra teknolojik çözümleri de gündeme getirmektedir. Günümüzde hâlihazırda nefret içeriklerinin tespit edilmesi, yayından kaldırılması ve suç teşkil eden içeriklerin adli mercilere bildirilip, suçun ve suçlunun kovuşturulması için geliştirilmiş teknolojik denetleme ve gözetim algoritmaları yaygın olarak kullanılmaktadır. Fakat bu algoritmalar gerek teknolojinin kendi doğasından kaynaklanan yetersizlikler gerekse de nefret söyleminin hukuki ve kültürel anlamda henüz daha netleşmemiş kavramsal yapısından dolayı soruna çözüm olamamakta ve eleştirilmektedir. Meseleyi ele alırken hesaba katılması gereken unsurlardan biri de teknoloji ile insan ve toplum arasındaki ilişkidir. Bu kapsamda çevrim içi nefret söylemi konusunda sosyal ağların rolünün ve etkisinin çok iyi analiz edilmesi gerekmektedir. Çalışmada teknoloji ile insan ve toplum arasındaki etkileşimde, insan ilişkilerini ve ihtiyaçlarını öne alan bir yaklaşım tercih edilmiştir. Dolayısıyla nefret söyleminin önemli ölçüde toplumsal ve kültürel bir olgu olduğu, sosyal medyanın ise dışı vurumu kolaylaştırarak bu sorunu görünür kıldığı öngörülmektedir.

2. Özgürlük ve Eşitlik İkileminde Nefret Söylemi

Uluslararası kabul görmüş, standart bir tanımı olmayan nefret söylemi (Nardocci); yabancı düşmanlığı, ayrımcılık ve şiddete teşvik noktasında ortak referanslara ve hassasiyetlere sahip bir kavram olarak mevzuatta yerini almaktadır. Avrupa Konseyine göre nefret söylemi, “basit bir bireysel düşünce veya irade beyanı olmayıp mağdur üzerinde somut bir etki oluşturmaya ve ayrımcılık riski altındaki gruplara yönelik, yayma, teşvik etme veya kışkırtma yönünde eylemler” olarak kabul edilmekte ve suçla bağlantısı kurulmaktadır. Buna göre nefret söylemi, ırkçı nefreti, yabancı düşmanlığı, Yahudi düşmanlığını veya hoşgörüsüzlüğe dayalı diğer nefret biçimlerini yayan, kışkırtan, teşvik eden veya haklı çıkararak tüm ifade biçimleri olarak el alınmaktadır (rm.coe.int). Birleşmiş Milletlerin her türlü ırk ayrımcılığının ortadan kaldırılmasına ilişkin sözleşmesine göre nefret söylemi “Kanun tarafından cezalandırılabilir suçları, üstünlüğe veya ırksal nefrete dayalı her türlü fikri yaymayı, ırk ayrımcılığına teşviki ayrıca herhangi bir ırk, farklı renk veya bireye yönelik herhangi bir şiddet eylemini veya bu tür eylemlere yönlendiren söylemleri” ifade etmektedir (ohchr.org). Yine Birleşmiş Milletlerin Uluslararası Medeni ve Siyasal Haklar Sözleşmesi’nin 20. maddesine göre ayrımcılık, düşmanlık ve şiddete teşvik eden ırksal ve dini nefrete dayalı atıflar ve her türlü savaş

yanlısı propaganda yasaklanmaktadır (ohchr.org).

Twitter, YouTube, Instagram gibi sosyal ağlarda ise nefret söylemine karşı caydırıcı ve engelleyici politikalar geliştirilmiştir. Bu korumacı politikaların, genel itibarıyla uluslararası kural ve ilkelerle paralel olmakla birlikte daha katı olduğu görülmektedir. Örneğin tartışma ve fikir alışverişine çok daha elverişli olan Twitter, ifade özgürlüğünü insani bir hak olarak tanımlayıp kendi misyonunu “herkese içerik oluşturma, görüşlerini ve bilgilerini paylaşma, görüşlerini ve inançlarını sınırsızca ifade etme” şeklinde belirlemektedir. Söz, yazı veya görsel unsurlara dayalı her türlü nefret söylemini, ifade özgürlüğünü kısıtladığı ve kuruluş amacına ters düştüğü gerekçesiyle reddetmektedir. Özellikle nefret söyleminin hedefinde olan; “din, sınıf, yaş, engellilik, ciddi hastalık, ulusal köken, ırk veya etnik köken nedeniyle belirli insan gruplarını aşağılama”yı ve ciddi zarar görmelerini dilemeyi, ummayı ve istemeyi yasaklamaktadır (twitter.com.tr). Çevrim içi nefret söylemini engellemek amacıyla 31 Mayıs 2016’da Avrupa Konseyi ile Facebook, Microsoft, YouTube ve Twitter bir araya gelerek ifade özgürlüğünün desteklenmesi, ayrımcılık ve yabancı düşmanlığına dair nefret söylemiyle mücadele konusunda davranış kurallarının belirlenmesine hususunda anlaşmaya varmışlardır (Alkiviadou, 2019, s. 31).

Hukuk felsefesinin kadim tartışmalarından biri olan nefret söylemi, özellikle pek çok insanın temel sosyal ilişki dinamikleri hâline gelen sosyal ağlardaki hareketlilikle birlikte, çözüm öncelikli bir meseleye dönüşmüştür. Tartışmaların genellikle demokratik bir toplumun temel unsurları arasında yer alan ifade özgürlüğü, düşünce, din, vicdan özgürlüğü ve ayrımcılığa uğramama hakları ekseninde gündeme geldiği görülmektedir (Weber, 2009, s. 1). Nefret söyleminin bu hak ve özgürlüklerle bağdaşık yönleri, bu konuda alınabilecek kalıcı ve bağlayıcı politik kararları erteleyebilmekte, hukuki ve sosyal düzenlemeleri geciktirebilmektedir.

Özgürlükçü/eşitlikçi ikilemde gerçekleşen tartışmaların Amerikan liberal demokrasi anlayışı tarafında yer alan Lewis (2009), nefret söylemini önlemeye yönelik yasaların, demokratik toplumsal yapıya ve bu yapının temel taşlarından biri olan ifade özgürlüğüne zarar verebileceğini ileri sürmektedir. Nefret söylemi toplumsal grupların birbirleri hakkındaki düşüncelerinin görünür olmasını sağlamakta ve dışı vurumu kolaylaştırmaktadır. Dolayısıyla sorunların tespiti ve demokratik yapının iyileştirilmesi için söylemlerin engellenmemesi gerektiğini düşünmektedir. Lewis’e göre bireyler daha yaşanılabilir bir dünya için, başkalarının kendileri hakkında gerçekte ne düşündüklerini öğrenebilecek cesaretle olmalı ve bu gerçeklikle yaşayabilmeyi öğrenmelidir (Akt. Karadağ Thorpe, 2019, s. 72-73). Daha çok özgürlükçü saiklerle yola çıkan bu görüş, sağlıklı bireysel ve toplumsal refleks-

lerin ortaya çıkabilmesi için kısıtlama ve engellemelerin olmaması gerektiği savunmaktadır. Özgürlükçü görüş, ifade özgürlüğünü mutlak bir hak olarak tanımlama eğilimindedir. Sınırları ne kadar geniş olursa olsun bu hakkın kullanımından kaynaklanan sorunlu durumlar olabileceği unutulmamalıdır. İfade özgürlüğü (dış dünyaya açıklama veya forum externum) hakkının düşünce özgürlüğü (iç inanç dünyası veya forum internum) hakkından farklı olarak başkalarının hakları, görev ve sorumluluklar bağlamında sınırları olduğu kabul edilmektedir. Avrupa İnsan Hakları Sözleşmesi'nin 10'uncu maddesinin ikinci fıkrası, ifade özgürlüğü hakkının "ulusal güvenlik, toprak bütünlüğü veya kamu güvenliği yararına, kargaşa veya suçun önlenmesi için, sağlığı veya genel ahlakın korunması amacıyla başkalarının saygınlık veya haklarının korunması için gizlilik koşullarında alınan bilginin ifşa edilmesinin önlenmesi amacıyla veya yargının otoritesinin ve tarafsızlığının korunması için kanunda öngörülen ve demokratik bir toplumda ihtiyaç duyulan formalite, koşul, kısıtlama veya cezalara" tabi olarak düşünülmesi gerektiğinin altını çizmektedir (Weber, 2009, s. 1-31).

Tartışmaların karşı tarafında olanlar ise demokratik yapı içinde varlığını sürdüren insanların eşit statüde olduğu ve bu statüyü bozabilecek her türlü tutum ve davranışın yasal çerçevede sınırlandırılması gerektiğini düşünmektedirler. Eşitlikçi yaklaşım, sosyal bir varlık olan insanı, toplumsallaşırılığı içinde ele alma çabası içindedir ve bu çevrede tutum ve davranışlarını değerlendirme eğilimindedir. Ayrıca bu yaklaşımı savunanlar, söylem ile davranış arasındaki ilişkiye dikkat çekmektedir. Buna göre nefret söylemi yalnızca söylemsel düzeyde kalan ifade biçimleri olarak düşünülmemelidir. Bu tarz söylemlerin bireysel ve toplumsal alanda somut yansımaları bulunmaktadır. Nefret söylemine maruz kalan kişi veya gruplar; vatandaşlık haklarından ve toplumsal fırsatlardan mahrum kalma tehlikesiyle karşı karşıya kalabilmektedir. Kendilerini huzur ve güvende hissedemeyen insanlar demokratik tepki, tavır ve katılımı isteksiz davranabilmektedirler. Hatta nefret söylemi uzun vadede hedefindeki insanların can ve mal güvenliğini tehlikeye sokmakta ve şiddete maruz kalmalarına yol açabilmektedir. Dolayısıyla nefret söylemiyle zarar gören en önemli unsur toplumsal düzen, huzur ve barıştır. Nefret söylemi tartışmalarında eşitlikçi tarafta yer alan Jeremy Waldron (2012, s. 4) nefret söyleminin, etkisini yavaş yavaş gösteren bir zehir gibi 'toplumsal iyi'yi yok ederek toplumsal barışı tehlikeye soktuğunu iddia etmektedir. Waldron'un bir toplumda yaşayan her bireyin kendini o yapının eşit ve değerli bir üyesi olarak görmesi, rahat ve güvende hissetmesi olarak tanımladığı toplumsal iyi, kapsayıcı olmalıdır (Akt. Karadağ Thorpe, 2019, s. 77). Nefret söylemini, insanları nefret suçuna yönlendiren tahammülsüzlüğün ve hoşgörüsüzlüğün bir tür dışavurumu

olarak da değerlendirmek mümkündür. Söylemlerin hedefindeki gruplara “toplumda size yer yok” mesajı verilerek pasifize olmaları sağlanmakta ve “yaşama ve katılım hakkı” ihlal edilmektedir (İnceoğlu, Sözeri, 2012, s. 24).

3. Kamusal Alan ve Etkileşim Ortamı Olarak Sosyal Ağlar ve Çevrim İçi Nefret

Artık hukuki merciler tarafından da bir kamusal alan olduğu kabul edilen (Marantz, 2018) sosyal ağların ifade özgürlüğünde temel ortamlardan birine dönüştüğü görülmektedir. İletişim ve fikir alışverişi ortamı olma misyonlarını kısa sürede gerçekleştiren sosyal medya platformları, geleceğin küresel toplumunu inşa etme iddiasıyla yollarına devam etmektedir. İki milyarın üzerinde kullanıcısı olan Facebook, aile ve arkadaşları birbirine bağlamanın yanı sıra fikirlerin özgürce tartışıldığı bir ortam oluşturma felsefesiyle yola çıktığını belirtmektedir. Bugün ise bu iddianın bir adım ötesine geçerek geleceğin küresel toplumunun sosyal altyapısını oluşturmak, refah ve özgürlüğü yaymak, barışı ve anlayışı teşvik etmek, insanları yoksulluktan kurtarmak ve bilimi hızlandırmak gibi hedeflerle varlığını gerçekleştirmektedir (Zuckerberg, 2021). Facebook gibi diğer popüler küresel ağlar ve her biri kendi kurallarını oluşturmaktan sorumlu olan yüz milyonlarca WEB sitesi, topluluk altındaki insanları birbirine bağlayacak merkezi olmayan bir ağ olarak ortaya çıkmıştır. Ancak internet geliştikçe, Facebook, YouTube, Instagram, Reddit, Tumblr ve Twitter gibi şirketler, dünya çapında milyarlarca insan için birincil konuşma kanalları olarak bireysel WEB sitelerinin ve forumların yerini alarak daha fazla kurumsallaşmışlardır. Sosyal medya platformlarının büyüklüğü ve üstlendiği misyon, bugün her zamankinden daha fazla bu ağları tartışmaların ve eleştirilerin hedefi hâline getirmiştir. Hükümetler ve kurumsal haber kuruluşları bu ağları; asılsız haberler ve dezenformasyona yol açtığı için insan hakları örgütleri; toplumsal cinsiyete dayalı tacizi kolaylaştırdığı ve intihar ve cinayeti canlı yayınlayıp özendirdiği, çeşitli gruplar; terörizmin örgütlenmesine, ırkçılığa, şiddet ve nefrete zemin hazırladığı için bu ağlar eleştirilmekte ve çeşitli davalara konu olmaktadır (Koebler, Cox, 2018). Sosyal ağlarda yükselen nefret söyleminin toplumsal barış ve huzuru bozarak bireysel hak ve özgürlükleri tehdit edecek aşamaya gelmesi kurum ve kuruluşları daha kapsamlı, çok taraflı ve çok boyutlu uzun vadeli düzenlemeler yapmaya zorlamaktadır. Gelenen noktada sorunun teknolojik düzenlemelerle mi yoksa hukuki denetimle mi çözülmesi gerektiği sorunsalı henüz cevaplanabilmiş değildir. Ağların baskın teknolojik doğası, yüksek yoğunluklu veri akışı ve hızı ve uluslararası boyutu, sorunla yalnızca geleneksel adli ve polisiye önlemlerle mücadele edilemeyeceğini göstermektedir. Teknoloji eksenli çözümler ise ulusal hukuku ve kamu otoritelerini yetkisizleştirebilmekte, bireysel hak ve özgürlükleri küresel servis sağlayıcı şirketlerin insaf ve inisiyatifine bırak-

makta ve yerel kültürel değerler ve hassasiyetleri devre dışı bırakabilmektedir. Bu bağlamda çevrim içi nefret söylemiyle mücadeleye ilişkin en kapsamlı düzenlemelerden biri, Fransa Ulusal İnsan Hakları Danışma Komisyonu (CNCDH)'nin 12 Şubat 2015 tarihli genel kurulunda kabul edilen düzenleme kararlarıdır (legifrance.gouv.fr, 2015). Çalışmada öne çıkartılan tartışmalara genel bir bağlam oluşturması açısından bu politikalara kısaca değinilecektir. "İnternette Nefret Söylemiyle Mücadeleye İlişkin Görüş" başlıklı düzenleme dört aşamalı bir mücadele planı üzerine şekillenmektedir. Devletin dijital egemenliğinin öne çıkartıldığı ilk aşamada; WEB üzerindeki temel hak ve özgürlükleri garanti altına almada devletin rolünü güçlendirmek, siber suçlarla mücadelede diplomatik girişimleri artırmak ve dijital ekosistemdeki yabancı aktörlerin Fransız hukuku ve endüstri politikalarına entegrasyonu için çalışmalar yapmak hedeflenmektedir. Düzenlemeler özü itibarıyla servis sağlayıcı şirketlerin faaliyet ve sorumluluk alanlarında insan haklarına ve ifade özgürlüğüne saygı göstererek ve yerel kamu otoriteleriyle iş birliği içinde hareket etmesini içermektedir. Doğrudan çevrim içi nefret söylemini hedef alan ikinci aşamada basın, düşünce ve ifade özgürlüğü ile bu özgürlüklerin kötüye kullanımına yönelik mevzuatın güncellenmesi ve mevcut mekanizmaların güçlendirilmesi öngörülmektedir. Bu kapsamda sosyal ağlarda kamusal alan ve özel alan ayırımının netleştirilmesinin önemine dikkat çekilmekte ve nefret suçlarıyla mücadelede daha proaktif politikalar için servis sağlayıcı şirketlerin sorumluluklarına dikkat çekilmektedir. Buna göre kullanıcıların sahip olduğu teknik olanaklardan daha fazlasına sahip olan şirketler, sorunlu içeriği hızlı bir şekilde belirleyerek kamu otoritelerine bildirmek ve failerin tespiti ve yakalanmasında iş birliğine gitmekle yükümlü tutulmuşlardır.

Düzenlemenin üçüncü aşaması sosyal ağlardaki etkileşimin bütün taraflarını organize eden ve gözetken duyarlı ve yenilikçi bir üst kurulun gerekliliğine işaret etmektedir. Bu kurulun hem kurumsal düzensizliğe son vermesi hem de gözlem ve araştırmalarla nefret söylemiyle ilgili eğilimleri belirleyip politikalar üretmesi beklenmektedir. Düzenlemenin son maddesi ise daha kapsayıcı ve uzun vadeli bir çözüm olan dijital eğitim ve vatandaşlık konusunda ulusal eylem planı belirlemeye yöneliktir. Dijital platformlara erişim ve etkileşim hakkı temel bir hak olmakla birlikte bireylere sorumluluklar da yüklemektedir. Bu kapsamda kullanıcıların belli bir yetkinlik düzeyine ve sorumluluk bilincine sahip olmalarını sağlayacak bilgilendirme ve bilinçlendirme politikaları uygulanmalıdır.

4. Çevrim İçi Nefret Söylemini Önlemeye Yönelik Teknoloji Eksenli Düzenlemeler

Çağımızın yaygın etkileşim ortamları olan sosyal ağlardaki içeriğin etkili bir şekilde denetimi önemli olduğu kadar zor bir konudur. Milyonlarca kullanıcı tarafından oluşturulan küresel ölçekli ağlardaki veri hacmi, bu verilerin kültürel anlamda hassas ve bağlamsal doğası ve insan iletişimindeki çok boyutluluk ve nüanslar sorunun büyüklüğünü ortaya koymaktadır (Cobbe, 2020, s. 739). Yeni medyanın anonimliği, dolaysızlığı ve küresel doğası, aşırılık yanlıları ve nefret tacirleri için son derece elverişli bir ortam oluşturmaktadır. Teknolojinin küreselleşmesi, ağlardaki nefret gruplarının sayısının ve nefret faaliyetlerini katlanarak artmasına olanak sağlamaktadır (Banks, 2010, s. 233). Toplumsal sorun ve gerilimlerden beslenen, her zamankinden daha fazla bir arada yaşamayı sorgulayan grupların bir araya gelerek yeni çatışma alanları oluşturdukları sosyal ağlardaki nefret eğilimi endişe verici boyutlara ulaşmıştır (legifrance.gouv.fr, 2015). Buna rağmen siber özgürlükçüler, yeni medyanın bu sınırsız doğasına vurgu yaparak devletin siber uzayı denetleme meşruiyetini ve kabiliyetini sorgulayarak bilgi, fikir ve enformasyonun serbestçe dolaşımını savunmaktadırlar (Banks, 2010, s. 233). Sorunun ulaştığı boyutlar kamu otoritelerini, servis sağlayıcı şirketleri ve ulusal ve uluslararası kuruluşları önlem almaya itmiştir. Bu kapsamda geliştirilen çevrim içi nefretle mücadele düzenlemelerinin teknolojik, yasal ve etik alanlarında ilerlediği görülmektedir. Ulusal sınırları aşan dijital döngüsellik ve ağlardaki yüksek veri hacmi, nefret söylemiyle mücadele için tek taraflı düzenlemeleri yetersiz kılmaktadır. Ayrıca küresel bir ortam olan internete yönelik geliştirilen ulusal düzeydeki denetleme ve gözetim düzenlemeleri tutarlı olamamaktadır (Alkiviadou, 2019, s. 20). Teknolojik düzenlemeler ise her ne kadar etkili bir mücadele yöntemi olarak düşünülse de kendi içindeki yetersizlikleri ve uygulanım kaynaklı yeni sorunlara yol açabilme potansiyeliyle soru işaretleri oluşturmaktadır.

Servis sağlayıcı firmalar tarafından gündeme getirilen teknolojik denetim ve yasakların işe yarayıp yaramadığı konusunda çeşitli araştırmalar yapılmaktadır. Bu çalışmalardan elde edilen genel sonuç, yasakların önemli ölçüde amacına ulaştığı ve platformlardaki nefret söylemini azaltabildiği yönündedir. Bu konudaki en kapsamlı araştırmalardan biri üç Amerikan üniversitesinin ortaklığında yürütülen ve Reddit platformunun 2015 yılında uygulamaya başladığı nefret söylemi yasaklarının sonuçlarının incelendiği çalışmadır. Yasağın hem katılımcı kullanıcılar hem de etkilenen topluluklar üzerindeki nedensel etkilerinin araştırıldığı çalışmadan 100 milyon Reddit gönderisi incelenmiştir. Araştırma sonucunda yasaklanan alt dizinlere katılan kullanıcıların ya siteyi terk ettikleri ya da nefret söylem ve uygulamalarını azalttıkları ortaya çıkartılmıştır (Chandrasekharan, Pavalanathan ve diğerleri 2017, s. 31-19).

Nefret söylemini önlemeye yönelik teknolojik düzenlemelerin en belirgin örneklerini servis sağlayıcı şirketlerin kullanıcılar için geliştirdikleri topluluk kuralları oluşturmaktadır. Müşterilerin sisteme dâhil olurken kabul ettikleri gönüllü davranış kurallarını içeren bu kodlar, gerekli hâllerde servis sağlayıcı firmalara içeriği silme veya hizmeti iptal etme olanağı sunmaktadır. Bu kurallar çevrim içi nefret düzeyinin azaltılmasında önemli rol oynayabilmektedir. İngiltere’de bu kuralların uygulamasını sağlamak amacıyla endüstri birlikleri kurulmuştur (Banks, 2010, s. 237). Bu kurallar bir yandan her bir kullanıcıyı genel ilkelere uymaya zorlarken diğer yandan zararlı içeriği servis sağlayıcıya bildirmeye teşvik etmektedir.

Sosyal ağlardaki istenmeyen içeriğin tespit edilmesine ve silinmesine yönelik olarak geliştirilmiş algoritmik düzenlemeler yaygın ve etkili denetim yollarından biridir. Teknoloji bazlı, otomatikleştirilmiş ve diğer denetleme yöntemlerine göre daha meşru bir yöntem olan algoritmik uygulamalar, sosyal ağ sağlayıcıları tarafından giderek daha fazla benimsenmektedir. Algoritmik uygulamalar önemli ölçüde kullanıcılardan gelen bildirimler ve ağ moderatörlerinin incelemesine dayansa da çok daha kapsayıcı ve hızlı bir denetime olanak sağlamaktadır (Cobbe, 2020, s. 739). Dört buçuk milyardan fazla internet kullanıcısının her gün milyarlarca içerik paylaştığı dijital dünyada, tüm içeriğin taranması ancak yapay zekâ tabanlı algoritmalarla mümkün olabilmektedir. Algoritma ön moderasyon aşamasında tüm içeriği denetleyerek zararlı olmayan içerikleri serbest bırakmakta, zararlı olan içeriği ise otomatik olarak silmektedir. Emin olunamayan içerikler ise insan incelemesi için moderatörlere gönderilmektedir. Algoritmanın doğru tahminler yapabilmesi için kelime eşleme tabanlı doğal dil işleme teknolojisi, özellikle alay ve öfke tonlarını anlamaya yönelik duygu analizi ve veri tabanlarıyla eşleme teknikleri kullanılmaktadır (appen.com, 2021).

Sağladığı tüm avantajlara rağmen algoritmaların üstesinden gelemediği ve hatta sorunu daha da içinden çıkılmaz hâle getirdiği durumlar da söz konusudur. Bu sorunların başında nefret söyleminin tespitinde öncelikli unsur olan dilin kullanım biçimleri, kültürel bağlamlar ve sosyal ağların çok dilli yapısı gelmektedir. Facebook yaptığı bir açıklamada yapay zekâ algoritmaları sayesinde istenmeyen e-postaların yüzde 100’ünü, terörle ilgili içeriklerin yüzde 99.5’ini, şiddet içerikli mesajların yüzde 86’sını tespit edebildiğini bildirmiştir. Bu başarı oranları nefret söylemiyle ilgili denetlemelerde yüzde 38’e düşmektedir. Nefret söylemi konusundaki en önemli sorunlardan biri de kavramın henüz tam olarak tanımlanmamış ve sınırları çizilmemiş doğasıdır (Koebler ve Cox, 2018). Popüler sosyal ağların çoğu Amerikan orijinli olduğu için Türk mevzuatında suç sayılabilecek sövme, iftira, hakaret gibi söylemler, ifade özgürlüğünü daha geniş bir perspektiften ele

alan ABD hukukunda eleştiri olarak kabul edilebilmektedir (diabgm.adalet.gov.tr). Öte yandan nefret söylemi olarak kabul edilen ifadelerin hangilerinin ciddi bir sosyokültürel bağlama işaret eden yapısal ifadeler, hangilerinin güncel tartışmalara yönelik anlık ifadeler olduğu noktasında da algoritmaların hata yapabileceği göz ardı edilmemelidir. Yapısal söylemler çoğunlukla politize olmuş gruplar tarafından gündeme getirilen sistematik, planlı ve maksadını aşabilecek nitelikte olabilmektedir. Sıradan nefret ifadeleri ise sosyal ağlardaki anonimliğin yol açtığı gruba uyma, katılma gibi eğilimlerin sonucunda ortaya çıkabilen geçici ve zamanla değişebilme potansiyeli olan durumları ifade etmektedir. Dolayısıyla algoritmik uygulamaların bu iki söylem biçimini birbirinden ayırt etmesi oldukça zor gibi görünmektedir (Conseil Supérieur de l'audiovisuel Rapor, 2009'dan akt. legifrance.gouv.fr, 2015).

Algoritmik uygulamalar adli ve sosyopolitik bağlamlarda da eleştirilmektedir. Fransa'da faaliyet gösteren Toplum İnternet Hizmetleri Derneği, çevrim içi nefret düzenlemeleri kapsamında kamu otoritelerince alınan bazı kararların servis sağlayıcı şirketlere ağır bir sorumluluk yüklediğini belirtmektedir. Dernek, Fransa İnsan Hakları Ulusal Danışma Komisyonu'nun 2015 yılında karara bağladığı çevrim içi nefretle mücadele düzenlemelerini "ağır ceza tehdidi altında, tüm sorumlulukları zincirdeki tek bir aktöre yüklediği" gerekçesiyle eleştirmiştir. Şirketlerden, herhangi bir kamu otoritesi veya yargıç tarafından inceleme yapılmaksızın, zarar verici içeriği tespit etmelerini ve çok kısa sürede silmelerini istemek gerçekçi görünmemektedir. Diğer yandan bir şirketin tamamen yargı otoritesinin yerine geçmesinin adalet mekanizmasının işleyişi açısından sakıncaları olabileceği de göz ardı edilmemelidir (lasic.fr, 2020). Hizmet sağlayıcının içerikten doğrudan sorumlu tutulması onlara dijital kamu düzenini belirlemede çok merkezi bir rol atfederek, fiili olarak gözetim ve denetim yetkisi vermektedir (legifrance.gouv.fr, 2015).

Sosyal ağların içerik denetleme yollarından biri de kullanıcı bildirimleridir. Bazı sorunlu yanlarına rağmen bu uygulama; uygunsuz içeriğin doğru tespiti, zamanında müdahale ve karar almayı kolaylaştırma açısından diğer uygulamalara göre daha az tartışmalıdır. YouTube, Twitter, Facebook gibi popüler sosyal ağların uzun zamandır kullandıkları bu uygulama, uygunsuz olduğu düşünülen içeriklerin işaretlenerek moderatörlere bildirilmesi esasına dayanmaktadır. YouTube bildirimlerde Topluluk Kuralları'na aykırılığı ve küçük yaşta kitlelere uygunsuz içerikleri hedeflerken şikâyet nedeninin ayrıca yazılabilmesine olanak sağlamaktadır (support.google.com). Twitter ise bildirim kategorilerine gizli bilgi paylaşma, taciz, nefret davranışı, şiddet içeren tehditler, kendine zarar verme unsurlarını dâhil etmektedir (help.twitter.com). Sosyal ağların bildirim uygulamalarını yaygınlaştırabilmek

amacıyla bildirenin kimliğinin gizli tutulması, pratiklik, bildirim sonucunu takip edebilme gibi özendirici uygulamaları da devreye soktukları görülmektedir. Bunlara ek olarak YouTube, Güvenilir İşaretleyici Programı'nı geliştirmiştir. Bu program tek seferde birden fazla video bildirmeyi, işaretlenen içerikle ilgili kararları takip edebilmeyi, sürecin hızlandırılmasını ve içeriklerle ilgili çeşitli tartışmalara katılma olanakları sağlamaktadır. Ayrıca program yalnızca sivil toplum kuruluşları için çevrim içi eğitim programlarına katılabilmeyi sağlamaktadır (support.google.com). Program kullanıcılara gönüllü yarı moderatörlük yetkisi vermektedir. Belirlenen kodlara göre işaretlenen içerikler moderatörlere gönderilerek işleme alınması sağlanmaktadır. Çoğu zaman moderatörlerin içerikleri ne şekilde işleme aldıkları, sonuçlarının ne olduğu bilinmemektedir.

Sosyal ağlardaki bildirimler yalnızca teknik bir uygulama olarak değerlendirilmelidir. Bildirimler; kullanıcılar ve platformlar, insanlar ve algoritmalar ile sosyal medyanın yapısal düzenlemeleri ve normları arasındaki karmaşık etkileşimi ifade etmektedir (Crawford ve Gillespie, 2016, s. 411). İşaretle ve sil politikası, bildirimlerin yoğunluğu ve hızına bağlı olarak sorunu tespit etmek ve en uygun müdahale yolunu seçmek için ideal olsa da kullanıcıların bireysel veya toplu istismarlarına maruz kalabilmektedir. Bazı gruplar manipülatif amaçlarla açıktan veya gizli olarak bildirimde bulunma organizasyonları düzenleyerek baskı oluşturabilmektedir (Crawford ve Gillespie, 2016, s. 422-423).

Nefreti önlemeye yönelik düzenlemelere ilişkin olarak servis sağlayıcı firmalar tarafından geliştirilen ve belki de çoğu kullanıcının bilmediği bir uygulama da 'gölge yasaklama' (shadow-banning) uygulamasıdır. Bir sosyal ağ kullanıcısının servis sağlayıcı tarafından zımmen engellenmesi ve sessize alınmasını ifade eden bu kısmi yasak, uzun süredir Twitter ve Reddit gibi sosyal ağlarda teknolojik bir denetleme unsuru olarak kullanılmaktadır. Bu uygulamada kullanıcının içerik oluşturmaya ve ekleme yapmasına izin verilmekte ama bu içeriğin başkaları tarafından görülmesi kısıtlanmakta veya tamamen engellenmektedir (Young, 2022, s. 15). Sosyal ağlardaki bot hesapların ve nefret içeriğinin engellenmesi amacıyla gündeme gelen bu uygulama; kötüye kullanım ve manipülasyona açıktır. Nitekim 2018 yılında Amerikan Kongresi'nde, Twitter'ın muhalif hesapları kasten engellediğine dair tartışmalar gündeme gelmiş ve Twitter CEO'sunun açıklama yapması istenmiştir. Bu olaydan hemen sonra Twitter, kullandığı algoritmayı değiştirerek önceden görünmeyen hesapların içeriğini kullanıma açmıştır (Romano, 2018).

Kullanıcıları çevrim içi düzenlemelere dâhil etme konusunda istekli olan servis sağlayıcı şirketler, kamu otoritelerini sürece eklemlemede çekimser davranmaktadır. Siber suçların soruşturulması için gerekli olan IP ve trafik bilgisi an-

cak ilgili şirketten elde edilebilmektedir. Çoğu ABD orijinli olan sosyal ağlar Türk kamu otoritelerince kendilerine doğrudan yapılan bilgi, belge ve ceza istinabe taleplerini kabul etmemektedir. ABD ile Türkiye arasında yapılan adli yardımlaşma anlaşmasına göre suça ilişkin başvurular ve talepler doğrudan ABD'deki adli makamlara iletilmekte oradan da şirketlere talepte bulunmaktadır. Örneğin Twitter, kullanıcı başvuruları doğrultusunda içeriği kaldırabiliyorken Telekomünikasyon İletişim Başkanlığı (TİB)'ndan yapılan IP ve trafik bilgisi taleplerini kabul etmemektedir. Facebook ise sadece cinayete ve intihara teşebbüs, kayıp bir şahsın bulunması ve çocuklara yönelik cinsel istismar gibi acil müdahale gerektiren durumlarda adli makamlardan iletilen doğrudan talepleri kabul etmektedir. Diğer ceza istinabe taleplerini geri çevirmektedir (diabgm.adalet.gov.tr).

İletişim teknolojilerinin oluşturduğu olanak ve fırsatlara paralel olarak büyüyen çevrim içi nefret sorununa yine teknoloji odaklı çözümler aramak son derece gerekli görünse de yeterli değildir. Çok güçlü toplumsal ve bireysel bağlamları olan nefret söylemi, ifade özgürlüğüne zarar vermeden çok taraflı ve çok boyutlu düzenlemelerle ele alınması gereken bir meseledir. Burada çevrim içi düzenlemelerle birlikte, nefret söylemine zemin oluşturan toplumsal koşulların da iyileştirilmesi zorunluluğunun altını çizmek gerekmektedir. Çünkü teknolojik olanakların kullanılma biçimi önemli ölçüde toplumsalla bağlantılıdır.

5. Sonuç

Nefret söylemiyle mücadele neden ve sonuçları açısından çok taraflı müdahale ve düzenlemeleri gerektiren karmaşık ve zor bir mesele olarak karşımızda durmaktadır. Zamanla kendi kültürünü oluşturan ayrı bir teknolojik biçim olarak işleyen sosyal ağlar, güncel tartışmaların merkezinde yer almaktadır. Konunun sağlıklı bir zeminde tartışılması, anlamlandırılması ve çözüm yolları geliştirilebilmesi için öncelikle teknoloji ile insan ve toplum arasındaki etkileşimin doğasını anlamaya çalışmak gerekmektedir. Sosyal ağlarla birlikte nefret söyleminin boyut değiştirdiği, yaygınlaştığı ve ciddi bir soruna dönüştüğü ortada olan bir gerçektir. Sorunun bu noktaya gelmesinin asıl nedeni sosyal ağların sunduğu teknolojik olanaklar mıdır yoksa zaten reel hayatta var olan bireysel ve toplumsal tartışma ve gerilimler midir? Bu sorunun cevabı teknik ve teknolojinin kendi doğasında gizlidir. Daha geniş bir perspektiften, dijital teknolojiyi üreten sosyoekonomik yapıya, ağlardaki etkileşimin döngüsellğine ve insan davranışlarına yön veren sosyopsikolojik etkenlere bakıldığında sosyal ağların dışı vurumu kolaylaştıran ortamlar olduğu ve bu kapsamda kullanıcılara sonsuz fırsat sunduğu görülecektir. Dolayısıyla sosyal ağlar aynı düşünceleri paylaşan insanların birbirini bulmasını, kendileri gibi düşü-

nen insanlarla duygu ve düşüncelerini paylaşmasını kolaylaştırmaktadır.

Nefret söylemi, sosyal ağlarda grupları bir arada tutan ve benzer düşüncedeki insanları çeken son derece güçlü sosyopsikolojik motivasyonlar içermektedir. Bu durum sosyal medyanın heyecan verici deneyimler yaşatma vaadiyle de paralel ilerlemektedir. Sosyal ağlardaki cezazırlık ve anonimlik hissi ise dışa vurumu kolaylaştırarak nefret söylemine uygun bir zemin hazırlamaktadır (Charpenel, 2014'ten akt. legifrance.gouv.fr, 2015). Bu nedenle kullanıcılardan kaynaklanan güçlü sosyo-psikolojik eğilimler nefret söylemiyle mücadeleyi sonuçsuz bir çabaya dönüştürebilmektedir. Ayrıca söylemin kimlik, aidiyet ve sosyal katılım olgularıyla olan derin bağlantıları da ifade özgürlüğünün sınırlarını çizmeyi zorlaştırmaktadır.

Göz ardı edilmemesi gereken konulardan biri de yeni medyanın küresel tüketim kapitalizmine göre işleyen ve gelişen bir olgu olduğu gerçeğidir. Bu noktada teknolojinin doğasına ilişkin çözümlerin yanı sıra bu teknolojiyi ortaya çıkartan, besleyen ve yönlendiren pazar koşullarının da hesaba katılması gerekmektedir. Sektörün başat firmaları olası kullanıcı tepkilerini ve kayıplarını göze alabilecek durumdayken pazara yeni girmiş firmalar; engel ve kısıtlamalara sıcak bakmamakta ve hatta kısıtlamalardan dolayı popüler platformlardan kaçan kullanıcıları kendilerine çekmeye çalışmaktadırlar (Young, 2022, s. 14).

Çevrim içi nefreti önlemeye yönelik düzenlemeleri yetersiz kılan unsurlardan biri de ağların sürekli gelişen ve değişen teknolojik yapısıdır. Yeni medyanın WEB 2.0 uygulamalarıyla birlikte sosyal bir ağa dönüşmesinin üzerinden yaklaşık çeyrek asır geçmiş olmasına rağmen tam anlamıyla kendi teknolojik formunu bulduğu söylenemez. Sürekli gelişen ve dönüşen bir ortam için kalıcı ve geçerli kurallar koymak, uzun vadeli politikalar üretmek çok mümkün görünmemektedir. Bu nedenle yeni düzenlemelerin dinamik ve gelişmelere hızlı bir şekilde tepki veren bir yapıda olması gerekmektedir.

Çevrim içi nefreti önlemeye yönelik düzenlemelerin kullanıcılar, servis sağlayıcı şirketler ve kamu otoriteleri olmak üzere üç temel tarafı bulunmaktadır. Nefret söylemiyle mücadelede istendik sonuçların alınabilmesi bu bileşenlerin ortak çabasını ve koordineli hareketini sağlayacak düzen ve devamlılığa bağlıdır. Bu kapsamda kullanıcıları nefret söylemine iten iletişimsel ihtiyaçların ve sosyal ağların "ortak bir ihtiyaç olan bilinç gelişimine" katkısının çok iyi hesaba katılması gerekmektedir. R. Williams'ın (1962, s. 320) "her gerçek iletişim kuramı aynı zamanda bir topluluk/ortaklık kuramı olmak zorundadır." tespitinden hareketle çevrim içi nefretin önemli ölçüde sosyal ağların dışında gerçekleşen bir sorun ol-

duğunu kabul etmek ve alınacak önlemlerin reel hayata yönelik düzenlemelerle birlikte düşülmesini sağlamak gerekmektedir (Scannell, 2020, s. 141). Çevrim içi nefret, demokratik değerlerin yeniden yorumlanmasını gerektirecek kadar büyük bir tehdit oluşturmaya da bazı insani amaçların aşınmasına yol açmakta ve insanları yeni bir hümanizmi yeniden keşfetmeye zorlamaktadır (Ellul, 2003, s. 449). Fransa’da faaliyet gösteren Toplum İnternet Hizmetleri Derneği (ASIC), çevrim içi nefretle etkili bir şekilde mücadele etmenin tek yolunun “ifade özgürlüğüne saygı duyarak orantılı bir tepkiye izin vermek ve yavaş yavaş, ilgili mesajların ciddiyetine bağlı olarak internet kullanıcılarının davranışlarını değiştirmek ve adalete cezai söylemlerin faillerini tespit etmek, kovuşturmak ve cezalandırmak için yetki vermek ve fırsat oluşturmak” olduğunu ifade etmektedir (Iasic.fr, 2020). Bu kapsamda kullanıcılarda topluluk ve katılım bilincinin oluşturulması, sorumlu davranış kodlarının geliştirilmesi ve kullanıcıların bu davranış kodlarına uymalarının özendirilmesi öncelikli bir konudur.

Çıkar Çatışması Beyanı

Makale yazarı herhangi bir çıkar çatışması olmadığını beyan etmiştir.

Kaynakça

- Alkiviadou, N. (2019). Hate Speech on Social Media Networks: Towards a Regulatory Framework?, *Information & Communications Technology Law* 28(1), (s.19–35).
- Avis Sur La Lutte Contre Les Discours de Haine Sur İnternet, (10 juillet 2015) <https://www.legi.france.gouv.fr/jorf/id/JORFTEXT000030862432> (Erişim Tarihi: 02.02.2022).
- Banks, J. (2010). Regulating Hate Speech Online, *International Review of Law, Computers & Technology*, 24(3), (s.233–239).
- Chandrasekharan, E., Pavalanathan, U., Srinivasan, A., Glynn, A., Eisenstein, J & Gilbert, E. (2017). You Can’t Stay Here: The Efficacy of Reddit’s 2015 Ban Examined Through Hate Speech, *PACM on Human-Computer Interaction*, 1(31), (s.31-22)
- Cobbe, J. (2021). Algorithmic Censorship by Social Platforms: Power and Resistance, *Philosophy & Technology*, (s.34:739–766).
- Council Of Europe Committee Of Ministers Of The Committee Of Ministers To Member States On “Hate Speech” Recommendation No. R (97) 20, <https://rm.coe.int/1680505d5b> (Erişim Tarihi: 14.02.2022).
- Crawford, K. & Gillespie, T. (2016). What is a Flag For? Social Media Reporting Tools and The Vocabulary of Complaint, *New Media & Society*, 18(3) (s.410–428).
- Ellul, J. (2003). *Teknoloji Toplumu*, (Çev. Musa Ceylan), Bakış Yayınları, İstanbul.
- Gagliardone, I., Gal, D., Alves, T. & Martinez, G. (2015). *Combattre Les Discours De Haine Sur Internet*, Collection UNESCO Sur La Liberte De L’Internet, Paris.
- Fink, A. (2010). *Conducting Research Literature Reviews*, SAGE Publications, United States of America.
- International Convention on the Elimination of All Forms of Racial Discrimination, <https://www.ohchr.org/Documents/ProfessionalInterest/cerd.pdf> (Erişim Tarihi: 03.02.2022).
- International Covenant on Civil and Political Rights, <https://www.ohchr.org/EN/ProfessionalInterest/Pages/CCPR.aspx> (Erişim Tarihi:03.02.2022).

- İhlalleri Bildirme, <https://help.twitter.com/tr/rules-and-policies/twitter-report-violation> (Erişim Tarihi: 10.03.2022).
- İnceoğlu, Y. ve Sözeri, C. (2012). Nefret Suçlarında Medyanın Sorumluluğu: “Ya Sev Ya Terk Et Ya Da...”, Nefret Söylemi ve/veya Nefret Suçları, (Der. Yasemin İnceoğlu), Ayrıntı Yayınları, İstanbul.
- İnternet Ortamında İşlenen Suçlarda Uluslararası Ceza İstinabe İşlemleri, Dış İlişkiler ve Avrupa Birliği Genel Müdürlüğü, <https://diabgm.adalet.gov.tr/Home/SayfaDetay/internet-ortaminda-islernen-suclarda-uluslararası-ceza-istina-be-islemleri14022020041756> (Erişim Tarihi: 20.02.2022).
- Koebler, J. & Joseph Cox, The Impossible Job: Inside Facebook’s Struggle to Moderate Two Billion People (August 23, 2018), <https://www.vice.com/en/article/xwk9zd/how-facebook-content-moderation-works>
- Leveraging AI and Machine Learning for Content Moderation, (February 04, 2021) <https://appen.com/blog/content-moderation/> (Erişim Tarihi: 17.02.2022).
- Marantz, A., Reddit and the Struggle to Detoxify the Internet: How do we fix life online without limiting free speech? (March 12, 2018) <https://www.newyorker.com/magazine/2018/03/19/reddit-and-the-struggle-to-detoxify-the-internet> (Erişim Tarihi: 14.03.2022).
- Nardocci, C. Hate Speech: Le Indicazioni Del Diritto Internazionale, <https://controlodio.it/hate-speech-indicazioni-diritto-internazionale/> (Erişim Tarihi: 11.03.2022).
- Nefret Davranışı Politikası, <https://help.twitter.com/tr/rules-and-policies/hateful-conduct-policy> (Erişim Tarihi: 07.03.2022).
- PPL Avia: L’association Des Services Internet Communautaires Salue La Décision Du Conseil Constitutionnel (18 Juin 2020), Association des Services Internet Communautaires (lasic.fr) (Erişim Tarihi: 03.02.2022).
- Romano, A. How Hysteria Over Twitter Shadow-banning Led to a Bizarre Congressional Hearing, (Sep 6 2018), <https://www.vox.com/2018/9/6/17824652/twitter-dorsey-energy-and-commerce-hearing-shadow-banning> (Erişim Tarihi: 04.02.2022).
- Scannell, P. (2020). Medya ve İletişim, (Çev. Oğuzhan Taş, Burcu Sümer), Ütopya Yayınevi, Ankara
- Tranfield, D., Denyer, D. ve Smart, P. (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review, *British Journal of Management*, 14, (207-222).
- Uygunsuz İçeriği Bildirme, <https://support.google.com/youtube/answer/2802027?hl=tr&co=GENIE.Platform%3DAndroid#zippy=%2Cvideo-bildirme%2Coyunatma-listesi-bildirme%2Ck%C3%BC%C3%A7%C3%BC-resim-bildirme> (Erişim Tarihi: 14.03.2022).
- Weber, A. (2009). Nefret Söylemi El Kitabı, (Çev. Metin Çulhaoğlu), Avrupa Konseyi Yayınları, Strazburg.
- YouTube Güvenilir İşaretleyici Programı, https://support.google.com/youtube/answer/7554338?hl=tr&ref_topic=9387085 (Erişim Tarihi: 13.03.2022).
- Young, G.K. (2022). How Much is too Much: The Difficulties of Social Media Content Moderation, *Information & Communications Technology Law*, 31(1), (s.1-16).
- Zuckerberg, Z., Building Global Community, (5 Mayıs 2021) <https://www.facebook.com/notes/3707971095882612/> (Erişim Tarihi: 23.02.2022).