



BMO YAPAY ZEKÂ ETİK İLKELERİ

TMMOB
BİLGİSAYAR MÜHENDİSLERİ ODASI
BİLİŞİM POLİTİKALARI KOMİSYONU
YAPAY ZEKÂ, ETİK VE DÜZENLEME ÇALIŞMA GRUBU



TMMOB
BİLGİSAYAR MÜHENDİSLERİ ODASI

BMO YAPAY ZEKÂ ETİK İLKELERİ

Bu çalışma ile ilgili istek, görüş ve önerileriniz için;

TMMOB BİLGİSAYAR MÜHENDİSLERİ ODASI

Necatibey Cad. No. 88/7 Çankaya – ANKARA

Telefon: 0312 230 31 45

Belgeç (Faks): 0312 230 31 46

e-Posta: iletisim@bmo.org.tr

www.bmo.org.tr

1.Basım Ocak 2026

TMMOB
BİLGİSAYAR MÜHENDİSLERİ ODASI
BİLİŞİM POLİTİKALARI KOMİSYONU
YAPAY ZEKÂ, ETİK VE DÜZENLEME ÇALIŞMA GRUBU

BMO YAPAY ZEKÂ ETİK İLKELERİ

Katkıda Bulunanlar
Ezgi Ekiz Emiroğlu
Deniz Günalan Duru
İzlem Gözükeleş
Taylan Özgür Yıldırım

SUNUŞ

Değerli Meslektaşlarımız;

“Yeryüzünde bir hayalet dolaşüyor: Yapay zekâ hayaleti.”

Bu hayalet, yalnızca mühendislerin laboratuvarlarında değil; emek süreçlerinde, kamu yönetiminde, eğitim ve sağlık hizmetlerinde, güvenlik politikalarında ve gündelik hayatın en mahrem anlarında karşımıza çıkıyor. Kimin işe alınacağına, kimin ameliyat olması gerektiğinden, kimin kredi alacağına hatta kimin suç potansiyeli taşıdığına kadar pek çok kritik karar giderek insan aklından alınıp algoritmalara devrediliyor.

Yapay zekâ, sınırsız bir ilerleme vaadiyle sunulurken, bu ilerlemenin kimin yararına, kimin aleyhine işlediği sorusu maalesef çoğu zaman cevapsız kalıyor. Ancak bu hayaletle ilişkimizi korku ya da kör bir hayranlık üzerinden kurmak zorunda değiliz. Yapay zekâ, insanlığın ortak yararına hizmet edecek biçimde tasarlanabilir, geliştirilebilir ve kullanılabilir. Bunun için gerekli olan şey, teknolojiyi kendi başına bir kader olarak görmek yerine, insan aklı, etik sorumluluk ve toplumsal denetimle birlikte ele almaktır.

Bilgisayar Mühendisleri Odası olarak bizler de, mühendisliğin yalnızca “ne yapılabilir” sorusuna değil, aynı zamanda “ne yapılmalıdır” sorusuna da yanıt verme yükümlülüğü taşıdığını inanıyoruz. Teknolojik gelişmelerin insan onurunu, temel hak ve özgürlükleri, adaleti ve toplumsal faydayı gözetken bir anlayışla ele alınması gerektiği açıktır. Bu bağlamda yapay zekâ sistemlerinin tasarımından geliştirilmesine, uygulanmasından denetlenmesine kadar tüm yaşam döngüsünde etik ilkelerin yol gösterici olması hayati önem taşımaktadır.

Bu doküman, Bilgisayar Mühendisleri Odası’nın yapay zekâ alanındaki etik sorumluluğunu somut bir çerçeveye oturtma çabasının ürünüdür. Hazırlanan Yapay Zekâ Etik İlkeleri; insan merkezlilik, güvenlik, şeffaflık, hesap verebilirlik, ayrımcılığın önlenmesi, veri mahremiyeti ve sürdürülebilirlik gibi temel değerleri esas almaktadır. Amaç, yalnızca teknik uzmanlara değil; politika yapıcılara, akademiye, sektör bileşenlerine ve topluma yol gösteren ortak bir referans oluşturmaktır.

Bu ilkelerin bağlayıcı bir hukuk metni olmanın ötesinde, meslektaşlarımız için bir mesleki duruşu ve toplumsal sorumluluğu temsil ettiğini özellikle vurgulamak isterim. Yapay zekâ çağında bilgisayar mühendislerinin rolü, sadece yenilik üretmek değil; bu yeniliğin kime, nasıl ve hangi sonuçlarla hizmet edeceğini de sorgulamak olmalıdır.

Bu çalışmanın hazırlanmasında emeği geçen tüm çalışma grubu üyelerimize içten teşekkürlerimi sunuyorum. Yapay zekânın insanlığın ortak yararına hizmet eden bir araç olarak gelişmesi yönünde bu dokümanın önemli bir adım olmasını diliyor, tüm ilgililer için yol gösterici olmasını umut ediyorum.

Saygılarımla,

Cem Nuri ALDAŞ
TMMOB Bilgisayar Mühendisleri Odası
Yönetim Kurulu Başkanı
Ocak 2026

İÇİNDEKİLER

1. GİRİŞ	1
2. İLKELER.....	5
2.1. NESNELLİK	5
2.1.1. YZ Abartılarının Yayılma Mekanizmaları.....	5
2.1.2. Karar Vericiler Nelere Dikkat Etmeli?	7
2.1.3. Teknoloji Haberciliği	8
2.2. ADALET VE EŞİTLİK	9
2.2.1. Bulut İmparatorluklarının Yükselişi	9
2.2.2. Yapay Zekânın Endüstrileşmesi.....	11
2.2.3. Yeni YZ Girişimlerinin Bulut Bilişim Platformlarına Bağımlılığı	12
2.2.4. YZ’de Söz, Yetki, Karar Hakkı	13
2.3. AÇIKLANABİLİRLİK.....	14
2.3.1. Açıklanabilir Yapay Zekâ’nın Tanımı ve Önemi	14
2.3.2. AYZ’nin Amaçları	15
2.3.3. Kimlerin Neden AYZ’ye İhtiyacı Var?	15
2.3.4. Düzenleme ve Politika	16
2.3.5. AYZ Açıklama Tipleri	18
2.4. SÜRDÜRÜLEBİLİRLİK.....	22
2.4.1. Sürdürülebilirlik için YZ	22
2.4.2. YZ’nin Sürdürülebilirliği.....	23
2.5. ÖNYARGI.....	25
2.5.1. Geliştiricilerden Kaynaklı Ön Yargılar	26
2.5.2. Veriden Kaynaklı Ön Yargılar	27
2.5.3. Ön Yargıya Karşı Ne Yapılabilir?.....	28
2.5.4. Tahminler ve Gerçekler	29
2.6. MAHREMİYET	30
2.6.1. Hukuksal Önlemler	34

2.6.2. Teknik Çözümler.....	34
2.6.3. Özel Sorun Alanları	35
2.7. HESAP VEREBİLİRLİK	36
2.7.1. Hesap Verebilirlik İçin Gerekli Koşullar.....	37
2.7.2. Hesap Verebilirliğin Başlıca Özellikleri.....	38
2.7.3. YZ’de Hesap Verebilirliğin Hedefleri	40
2.7.4. Ön alıcı ve Tepkisel Hesap Verebilirlik	41
2.8. ZARAR VERMEMEK	41
2.9. ÖZERKLİK	45
2.10. SONUÇLAR.....	48
KAYNAKÇA.....	50

1. GİRİŞ

YZ (yapay zekâ) teknolojileri; sağlık, eğitim, tarım, imalat, adalet, iş yeri yönetimi gibi çeşitli alanlarda giderek daha fazla kullanılıyor. Bu alanlarda insanların işleriyle ve birbirleriyle olan ilişkilerini değiştiriyor ve dönüştürüyor. Söz konusu değişimin insanlık için iyi mi yoksa kötü mü sonuçlanacağı henüz belirsizliğini korusa da YZ'nin insanlığa ve çevreye etkileri hakkında ciddi endişeler var. Üretken YZ teknolojisiyle beraber bu endişeler daha da arttı.

Örneğin, FLI'nın (Yaşamın Geleceği Enstitüsü – Future of Life Institute) 22 Mart 2023'de yayımladığı "Dev Yapay Zekâ Deneylerini Duraklatın" başlıklı açık mektupta, dev YZ deneylerinin duraklatılması isteniyordu. İmzacılar arasında Yoshua Bengio, Stuart Russell, Gary Marcus gibi YZ dünyasının önemli isimlerinin yanında Elon Musk, Steve Wozniak, Yuval Noah Harari, Daron Acemoğlu gibi isimler de vardı. Mektupta YZ sistemlerinin artık genel görevlerde insanlarla rekabet edebilir hale geldiğine dikkat çekiliyor ve şu sorular soruluyordu:

- Makinelerin bilgi kanallarımızı propaganda ve gerçek olmayan şeylerle doldurmasına izin vermeli miyiz?
- Tatmin edici olanlar da dahil, tüm işleri otomatikleştirmeli miyiz?
- Eninde sonunda bizi geçebilecek ve yerimizi alabilecek zihinler geliştirmeli miyiz?
- Medeniyetimizin kontrolünü kaybetme riskine girmeli miyiz?

Ayrıca açık mektupta bu gibi kararların seçilmemiş teknoloji liderlerine bırakılamayacağı belirtiliyor; doğruluk, şeffaflık ve yorumlanabilirlik vurgusu yapılıyordu.^[1]

Daha sonra, 30 Mayıs 2023'te, CAIS'in (Center for AI Safety – Yapay Zekâ Güvenliği Merkezi) web sitesinde yayımladığı bildiride ise şöyle yazılıyordu:^[2] Yapay zekâ kaynaklı yok olma riskinin azaltılması, salgın hastalıklar ve nükleer savaş gibi diğer toplumsal ölçekli risklerle birlikte küresel bir öncelik olmalıdır.

YZ, hızla yaşamın ayrılmaz bir parçası haline gelirken beraberinde YZ'nin nasıl tasarlandığı, geliştirildiği, konuşlandırıldığı ve kullanıldığı giderek daha fazla önem kazanıyor. Buna karşın; nasıl sorusuna yanıt verebilen, kurumların ve bireylerin sorumluluklarını belirten, hesap verebilirlik mekanizmalarını düzenleyen ulusal ve uluslararası yasalar son derece yetersiz.

Büyük Teknoloji şirketleri, sorunların ve sorumluluklarının farkındalar. Fakat dışarıdan ve bağımsız bir gözetim yerine şirket içinde kurdukları etik kurulları ve kendi hazırladıkları ve uymayı taahhüt ettikleri etik çerçevelerle kamuoyunu ikna etmeye çalışıyorlar.

Şirketler YZ etiğini, kurlsızlaştırmayı, kendi kendini düzenlemeyi veya piyasa güdümlü yönetişimi haklı çıkarmak için kullanma eğilimindedir. Teknoloji şirketleri tarafından etik dilin bu şekilde artan şekilde araçsallaştırılması "etik göz boyama" (ethics washing) olarak adlandırılıyor ve genellikle teknoloji şirketlerinin kurumsal iletişim stratejisinin parçası olarak ortaya çıkıyor. Şirket içinde geliştirilen teknikler ve ürünler hakkında etik düşünceyle meşgul olabilecek uzmanlardan oluşan kurullar oluşturuyorlar. Hesap verebilirlik ve yönetim için bünyelerine yeni bir katman ekleyerek kendilerini dizginleyici kontrollere ve dahili kısıtlamalara tabi tutmaya istekli olduklarını göstermeye çalışıyorlar. Fakat şirketlerin veya etik kurullarının felsefi sorgulamaları, şirketlerin çitlerle çevrilmiş özel mülkiyetli alanlarında devam ettiği sürece bundan asıl yarar sağlayan toplum değil şirketin kendisi oluyor. Kendi bünyelerinde kurdukları YZ etik komitelerinin kararları, üst yönetimin onayına ve şirket finansmanına bağlı olduğundan şirket çıkarlarına aykırı bir durumda kurul çalışanlarının işine son verilebiliyor. YZ etiği ile şirketler, etik konuların önem ve önceliğini belirliyor ve bazı sorunları daha görünür kılarken bazılarını kamuoyundan saklamaya çalışıyorlar.

Son yıllarda ise YZ etiğine ilişkin tartışmalar, şirketlerin koyduğu çerçeveyi aşarak yeni bir evreye girdi. Van Wynsberghe'in^[3] vurguladığı gibi, YZ'nin sürdürülebilirliği etrafında şekillenen bu tartışmalar, üçüncü dalga YZ etiğini ortaya çıkardı.

YZ etiğinin birinci dalgası, henüz var olmayan bir teknolojiyi tartışıyor, YZ'nin neler yapabileceğine (örneğin süper zekâ) odaklanıyor ve robot ayaklanmalarına ilişkin hayali senaryolar üzerinde duruyordu. İkinci dalga, birinci dalga yaklaşımların spekülâtif tezlerine tepki olarak ortaya çıktı. Henüz var olmayan bir teknoloji hakkında hayali senaryolar üretmek yerine YZ teknolojisinin halihazırda oluşturduğu risklere odaklanmanın daha doğru olacağını savundu. Bu nedenle ikinci dalga YZ etiği, yapay öğrenme tekniklerine ilişkin sorunları gündemine aldı. Açıklanabilirlik, şeffaflık, önyargı ve vatandaşların haklarını ihlal eden yüz ve duygu tanıma sistemleri gibi konular tartışılmaya başlandı.

Veriye dayalı YZ yaklaşımlarının içerdği sorunların görünürlüğü arttı. Ancak ikinci dalga çerçevesindeki tartışmalarda açıklanabilirlik, önyargı, mahremiyet vb. konular salt teknolojik çözümlerle aşılabilecek sorunlar olarak görülüyordu. Bunun yanında teknolojik çözümlerin uygulanamayacağı tekelleşme gibi sorunlar, sorun olarak değerlendirilmiyordu.

Ancak ikinci dalga YZ etiğinin ortaya çıkışı birinci dalgaya özgü spekülâtif yaklaşımları ortadan kaldırmadı. Bunlar kimi zaman varoluşsal riskler kimi zaman da özgür iradeye sahip YZ'ler biçiminde gündeme gelmeye devam ettiler. Örneğin, hâlâ etik teorilerin önerdiği türden ahlaki muhakemeler yapabilen ve bunun sonuçlarına göre hareket eden YZ sistemleri tartışılıyor. Mevcut YZ, medyadaki abartılı haberlerin aksine bunun çok uzağında. YZ'lerin öncelikle tepki verilecek bir olay olduğunun farkına varabilmeleri ve çeşitli parametreleri hesaba katmaları gerekiyor. Örneğin, meşhur tramvay problemi için YZ; cinsiyet, yaş, statü vb. faktörleri çok kısa bir sürede belirleyebilmeli. İkincisi, karşılaşılan durumun etik boyutlarını ve verilecek kararın olası sonuçlarını öngörebilmeli. Üçüncüsü, mevcut durumu çözenin kendi sorumluluğunda olup olmadığına karar verebilmeli. Dördüncüsü, soruna hangi soyut etik kural(lar)ın uygulanabileceğine ve değerler birbirleriyle çeliştiğinde nasıl hareket edilebileceğine karar vermeli. Ayrıca kararın bazı önyargılardan veya bilişsel engellerden etkilenip etkilenmediğini belirlemek de önemli. Beşincisi, somut bir çözüm önermek için soyut etik kuralları mevcut soruna uygulayabilmeli. Henüz böyle bir YZ yok.^[4]

Üçüncü dalga ise YZ etiğini ikinci dalgayı içeren ama ondan farklı bir paradigma çerçevesinde ele alır. İkinci dalga yaklaşımların YZ modellerini teknik uzmanlar tarafından optimize edilecek izole varlıklar olarak görmesini eleştirir ve daha iyi tasarım yaratmak için sistemik veya yapısal değişime gerek olduğunu savunur. Temel iddiası, belirli sorunların sadece belirli teknolojilerin merceğinden görülmemesi gerektiği ve görülemeyeceğidir. Teknolojiler, daha büyük bir sosyoteknik sistemin parçası olarak algılanmalıdır.

Üçüncü dalga YZ etiğinde, hem çözüm hem de kapsam bakımından ikinci dalganın çizdiği çerçevenin dışına çıkarılır: Sürdürülebilirlik ve YZ sistemlerinin çevreye etkisi; gündelik yaşamın bir parçası haline geldikçe YZ'nin altyapılaşması, endüstrileşmesi ve gücün birkaç şirketin elinde toplanması; YZ hakkındaki abartılı (hype) iddia ve haberlerle kamuoyunun yanıltılması; YZ'nin insanların yerini alan değil onları güçlendiren biçimlerde kullanımı gibi konular gündeme getirilir.

İkinci dalga, birinci dalgaya; üçüncü dalga da ikinci dalgaya bir tepki olarak ortaya çıkmıştır. Ancak ikinci dalganın ortaya çıkışı birinci dalgayı ortadan kaldırmadığı gibi üçüncü dalga da ikinci dalgayı ortadan kaldırmaz. Ama başta sürdürülebilirlik olmak üzere ikinci dalganın ele aldığı konuları farklı bir bakış açısıyla (sürdürülebilirlik için YZ'den YZ'nin sürdürülebilirliğine geçilir) tartışır, şeffaflık ve önyargıyı toplumsal bir bağlamda değerlendirir, mevcut güç ilişkilerini dikkate alır.

BMO (Bilgisayar Mühendisleri Odası) tarafından hazırlanan etik ilkeler, üçüncü dalga YZ etiği kapsamında değerlendirilebilir.

2. İLKELER

2.1. NESNELLİK

Her yeni teknolojiye çeşitli vaatler ve felaket senaryoları içeren mitlerin eşlik etmesi olağan bir durumdur. Ama telgraf, telefon, elektrik ve son olarak da internet örneklerinde olduğu gibi yeni teknolojiler hakkındaki abartılı iddialar sonsuza kadar devam etmez. Teknoloji, gündelik hayatın sıradan bir parçası haline gelmeye başladığında teknolojinin sosyoekonomik ilişkilere etkileri daha somut hissedilmeye başlar.

Günümüzde YZ teknolojileri için de benzer bir süreç söz konusudur. Ancak bu süreçte özellikle iki konuya dikkat etmek gerekiyor. Birincisi, YZ teknolojileri hakkındaki yayılan abartılı söylemlerin ortaya çıkışında ve yayılmasında teknoloji sektörünün önde gelen aktörleri büyük bir rol oynamaktadır. Bir diğer deyişle, önceki teknolojilerde olduğu gibi kendiliğinden ve çeşitli aktörlerin katkılarıyla ilerleyen heterojen bir süreç yoktur. İkincisi, teknoloji şirketlerinin başat rolü nedeniyle YZ teknolojisi hakkındaki abartılı iddialar şirketlerin çıkarlarına hizmet edecek biçimde şekillenmektedir. Ayrıca "varoluşsal riskler" gibi felaket senaryoları toplumun karşı karşıya olduğu asıl sorunları gölgeleyebilmektedir.

2.1.1. YZ Abartılarının Yayılma Mekanizmaları

❖ Antropomorfizm

Antropomorfizm (ister gerçek ister hayali olsun) hayvanlar, doğal güçler, tanrılar ve makineler gibi insan olmayan varlıkların davranışlarına insani özellikler (örneğin niyetler, motivasyonlar ve duygular) atfetmeyi içerir. Günümüzde YZ aracılığıyla bir kazanç elde etmek isteyen aktörler, antropomorfizmden yoğun bir şekilde yararlanırlar. Antropomorfizm, sosyal etkileşimi kolaylaştırmak, kullanıcı deneyimini iyileştirmek veya pazarlama amacıyla gündeme gelebilir. Örneğin şirketlerin, karar vericileri, YZ'nin gerçek bir 'anlayış' geliştirebileceğine veya sorunları doğru bir şekilde 'yorumlayabileceğine' ikna ettiğini varsayalım.

Bu durumda YZ, sosyal yardımlar hakkında karar vermek, eğitime rehberlik etmek, davranışsal uyumu ölçmek veya mahkeme davalarını yargılamak gibi hassas görevlerde kullanılabilecektir.

❖ **YZ Uzmanlarının (!) Artması**

ChatGPT gibi teknolojiler, YZ alanının erişilebilirliğinde bir genişlemeye yol açtı. Böylece sadece ChatGPT'yi veya üretken YZ API'lerini kullanmak dışında fazladan bir bilgisi olmayan YZ uzmanları türedi. Bu süreçte, toplumun YZ etrafında şekillenen güçlü anlatıya kendini adapte etmeye çalışmasının önemli payı var. Fakat YZ uzmanlarının (!) sayısı arttıkça YZ teknolojisini bir çekiç, her sorunu da teknolojik çözümlere ihtiyaç duyan bir çivi olarak görme eğilimi de güçlenmektedir.

❖ **Fırsatları Kaçırma Korkusu**

"Teknolojinin hızına yetişemiyorum", "acaba geride mi kaldım?" vb. korkular bireyleri olduğu kadar devletlerin YZ politikalarını da etkilemektedir. Büyük teknoloji şirketleri ve geliştirdikleri YZ teknolojileri, giderek daha çok ulusların güvenliği ve egemenliği için vazgeçilmez olarak görülmektedir. Yaşamın Geleceği Enstitüsü'nün GPT-4'ten daha güçlü YZ sistemlerinin eğitime en az 6 ay ara verilmesi gibi çağrılar bir ara verme ve önlem çabasından çok YZ yatırımlarını artırmaya hizmet etmekte; bunun sonucunda da şirketler ve devletler "öyle büyük bir potansiyel güçten" geri kalmama endişesiyle YZ yatırımlarını artırmaktadır.

❖ **YZ Teriminin Aşırı Kullanımı**

Son yıllarda, firmalar ürün ve hizmetlerinin YZ içerdiğini daha çok ifade etmeye başladılar. Şirketler abartılı ve kimi zaman asılsız iddialarla ürünlerini YZ ile ilişkilendirerek pazarlamaya gayreti içindeler.

2.1.2. Karar Vericiler Nelere Dikkat Etmeli?

YZ hakkında yayılan abartılı söylemlerin (örneğin YZ'nin içerdği varoluşsal riskler), diğer YZ ilkelerimizde ele aldığımız sorunların göz ardı edilmesinde büyük payı vardır. Dolayısıyla en başta kamu kurumlarındaki karar vericilerin, teknoloji şirketlerinin yarattığı abartılı söylemlerin etkisinde kalmadan, YZ'nin mevcut potansiyelini nesnel ve gerçekçi olarak değerlendirebilmeleri diğer etik ilkelerde tartışılan sorunların çözümünde önemli bir adım olacaktır. Aşağıdaki soruların sorulması, YZ hakkında nesnel bir bakış açısı geliştirilmesine yardımcı olabilir:

- ❖ YZ teknolojilerini neden kullanıyoruz? Kurumumuzun bu teknolojilere gerçekten ihtiyacı var mı, yoksa sadece güncel olanı takip etme kaygısıyla mı hareket ediyoruz?
- ❖ Kurumumuzda kullanılan YZ uygulaması neden insan biçimci bir tarzda tasarlanmış? Bu tarz, gerekli bir strateji mi?
- ❖ YZ teknolojisini uygun şekilde uygulamak için yardım alacağımız YZ uzmanı (!) gerçekten ilgili teknoloji için gerekli bilgi derinliğine sahip mi?
- ❖ YZ'ye yatırım yaparken fırsat maliyetlerini hesaba kattık mı?
- ❖ Doğal kaynaklar, insan emeği ve enerji kullanımı da dahil olmak üzere tüm YZ altyapısının sürdürülebilirliğini dikkate alıyor muyuz? YZ tedarik zincirinin tamamının ve her bir bileşendeki etik, ekonomik, sosyal ve ekolojik hususların farkında mıyız?
- ❖ Bu soruların yanında karar vericilerin YZ hakkındaki abartılı anlatıların farkında olması, kurumun YZ ihtiyacını daha gerçekçi değerlendirmesini sağlayacaktır. YZ'yi ilerlemenin tek yolu olarak sunma konusunda çıkarları olan aktörlerin bulunduğu hakkındaki farkındalık, YZ'nin gerekliliği konusunda daha ölçülü bir yargıya varılmasına yardımcı olacaktır.

2.1.3. Teknoloji Haberciliği

Kamuoyunu yanıltan abartılı iddiaların önüne geçebilmek için teknoloji habercilerine de önemli görevler düşmektedir. Teknoloji sektöründeki insanlar genellikle YZ modellerini ve bunların etkilerini anlayabilecek ve açıklayabilecek tek kişilerin kendileri olduğunu iddia ederler. Ancak bu iddialar özellikle şirket yetkilileri veya sözcülerinden geldiğinde iddialara şüpheyile yaklaşmak gerekir.

Teknoloji habercilerinin YZ modelleri hakkındaki iddiaları değerlendirirken aşağıdaki konuları dikkate alması kamuoyunun bilinçlenmesine yardımcı olabilir:

- ❖ **Veri:** Bir amaç için ve belirli bir veri setinden yararlanılarak eğitilmiş bir model, bu amaca uygun olarak mı konuşlandırılmıştır, yoksa yeni bir amaç mı söz konusudur?
- ❖ **Mahremiyet:** Veriler toplanırken yurttaşların rızası alınmış ve yurttaşlar, verinin nasıl kullanılacağı hakkında bilgilendirilmiş midir?
- ❖ **Emek süreçleri:** Şirket, verilerin etiketlenme sürecinde yararlandığı iş gücü hakkında şeffaf davranmakta mıdır?
- ❖ **Doğruluk iddiaları:** Modellerini tanıtmaya çalışan şirketler sık sık “yüzde 95 doğruluk” gibi iddialarda bulunurlar. Haberciler burada daha derine inmelidir. Yüksek skor, yalnızca orijinal verilerin bir örneğinden mi geliyor? Model, gerçek dünyadaki örneklerle kontrol edildi mi? Herhangi bir gerçek dünya doğrulaması olmadığında şirketin iddialarına inanılmamalıdır.
- ❖ **Ölçüt:** Şirket, modeli değerlendirmek için hangi ölçütü kullanmaktadır? Kullandığı ölçüt uygun mudur? Doğruluk iddiası, yaşamsal bir konudur. Çünkü yanlış pozitifin mi yoksa yanlış negatifin mi daha kötü olduğuna karar verilmelidir. Örneğin, bir kanser tarama aracında, yanlış pozitif insanların gereksiz bir test yaptırmasına neden olabilirken, yanlış negatif bir tümörün tedavi edilebilir olduğu erken evrede gözden kaçmasına yol açabilir. Bir sanığın iki yıl içinde suç işleme riskini tahmin eden modeller, siyah ve beyaz sanıklar arasında benzer doğruluk oranına sahip olabilir. Ancak algoritmanın siyah sanıklar için beyaz sanıklara göre iki kat daha fazla yanlış pozitif sonuç vermesi adil olmayacaktır.

2.2. ADALET VE EŞİTLİK

YZ, küresel adalete ve faydalara eşit erişime katkıda bulunmalıdır.

YZ hakkında sorulması gereken soruların başında YZ'den kimin nasıl etkileneceği gelmektedir. Belirli bir etki bazıları için diğerlerine göre daha olumlu olabilir. Çalışanlar, hastalar ve tüketicilerden hükümetlere, yatırımcılara ve işletmelere kadar pek çok paydaş vardır ve bunların hepsi YZ'den farklı şekillerde etkilenebilir. Bu farklılıklar küresel düzeyde daha belirgindir. YZ, gelişmiş ve gelişmemiş ülkeleri nasıl etkileyecektir? Daha az eğitilmiş ve düşük gelirli insanlar, YZ'den yararlanabilecek midir? Teknolojiye kimler erişebilecek ve faydalarından yararlanabilecektir? YZ'yi kullanarak kimler daha da güçlenecektir? Kimler bu YZ'nin sağlayacağı avantajlardan mahrum kalacaktır?

Bu sorulara yanıt verebilmek için YZ'nin üzerinde yükseldiği üretim ilişkilerine bakmak gerekir. Son yıllarda, YZ'deki gelişmeler, YZ'nin insanlık için varoluşsal bir risk içerdiği tezlerini gündeme getirir de insanlık için gerçek varoluşsal bir risk, YZ'nin kendisinden değil YZ'nin arkasındaki üretim ilişkilerinden kaynaklanmaktadır. Son yıllarda YZ bağlamında üç önemli gelişme yaşanmaktadır ve bu gelişmelerin etkileri YZ ile sınırlı kalmayacaktır.

2.2.1. Bulut İmparatorluklarının Yükselişi

PCWorld'de 31 Aralık 2008'de yayımlanan bir makalede tüm zamanların en kötü yedi teknoloji tahmini listelenmişti. Listenin başında IBM'in kurucusu Thomas J. Watson'ın 1943 yılındaki tahmini yer alıyordu (https://www.pcworld.com/article/532605/worst_tech_predictions.html): "Bence dünyada belki beş bilgisayarlık bir pazar var." Kişisel bilgisayarların çıkışı Watson'un öngörüsünü alt üst etti. Dizüstü bilgisayarlar, tablet ve akıllı telefonlar derken dünya bilgisayarla doldu. Ama Watson'un tahmini yıllar sonra doğru çıktı!

Bilgisayarlarımızın (tablet, akıllı telefon, TV gibi internete bağlı tüm cihazlar) dünyadaki birkaç bulut bilişim platformuna bağımlılığı giderek artıyor. Telefonlarımızın birçok özelliği ancak internete bağlı olduğumuzda etkin oluyor. Sosyal medya platformlarına giriyor, anlık mesajlaşma uygulamalarını kullanıyor, müzik dinliyor, film izliyor, yol tarifi alıyor, bankacılık işlemlerini yapıyor, alışveriş yapıyoruz ve fotoğraflarımızı yedekliyoruz. Televizyonlarımız, medya platformlarına bağlı. Bilgisayarlarımıza kurduğumuz yazılımların sayısı azaldı. Şirketlerin yazılımlarına abone oluyoruz. Daha da önemlisi bireylerden çok kuruluşların platformlara bağımlılığı artıyor. Eğer bir zekâdan bahsedeceksek, gerçekte zeki olan (yüksek işlemci hızlarına ve bellek kapasitelerine sahip olmalarına rağmen!) telefonlarımız değil, YZ modellerini oluşturan, eğiten ve YZ hizmeti sunan dev bilgisayarlar, yani bulut platformlarıdır.

Bu durum, çeşitli riskleri de beraberinde getiriyor. Bu risklerin başında da bilgisayarlarımızın bu dev bilgisayarlara bağımlılığının giderek artması var. Ayrıca bu dev bilgisayarların sadece birkaç büyük şirketin elinde toplanmış olması karşı karşıya olduğumuz riski daha da artırıyor. Örneğin, ana bilgisayarlardaki güncellemeler veya teknik sorunlar, bulut platformlarına yapılan saldırılar veya bulut platformlarındaki güç kesintileri küresel krizlere neden olabiliyor.

Son yıllarda YZ altyapılarının ve hizmetlerinin geliştirilmesinde üç büyük teknoloji şirketi öne çıkıyor: Amazon, Microsoft ve Google. Bu şirketlerin gücü üç etkene dayanıyor. İlk olarak, YZ gelişiminin anahtarı olan büyük miktarda değerli veriye sahipler ve veri birikiminin sürekliliğini sağlayabiliyorlar. İkincisi, bu şirketler, yüksek vasıflı YZ araştırmacılarını istihdam etmek ve onları elde tutmak için çok daha fazla finansal kaynağa sahipler. Üçüncüsü, GPU'lar (Grafik İşlem Birimleri) gibi ileri teknoloji hesaplama kaynaklarına sahipler. Bir diğer deyişle, bu üç Büyük Teknoloji şirketi; veri, YZ uzmanlığı ve bilgi işlem altyapıları üzerindeki artan ekonomik ve siyasi güçleri ile YZ teknolojilerinin geliştirilmesinde, YZ hakkındaki küresel anlatının şekillendirilmesinde ve YZ'nin yönelimlerinin belirlenmesinde merkezi bir role sahipler.

2.2.2. Yapay Zekânın Endüstrileşmesi

On beş yıl önce, Microsoft'un bir işletim sistemi üreticisi, Google'ın bir arama motoru ve Amazon'un da bir alışveriş sitesi olduğunu söyleyebilirdik. Oysa bu üç şirket günümüzde çok çeşitli hizmetler sunuyorlar ve aslında internetteyken çoğu zaman doğrudan veya dolaylı olarak^[5] bu şirketlerin bulut platformlarına bağlanıyoruz. AWS (Amazon), Azure (Microsoft) ve GCP (Google), YZ'nin endüstrileşmesi sürecinin temel omurgasını oluşturuyor. Günümüzde internetin kalbinin attığı bu üç platform, farklı sektörlerdeki işletmeler üzerinde derin bir etkiye sahipler. Pazar araştırması tahminlerine göre AWS'yi internetin birincil işletim sistemi olarak değerlendirebiliriz. Dünyadaki tüm bulut hizmetlerinin üçte biri AWS (%32) üzerinde çalışıyor. Onu Microsoft Azure (%22) ve Google Cloud (%11) takip ediyor.^[6]

Fakat YZ'deki gelişmelerle beraber büyük teknoloji şirketleri son yıllarda daha iddialı adımlar atıyorlar. Üçüncü taraf geliştiricileri ve işletmeleri çekmeyi amaçlayan, sektör odaklı çözümler ve pazar yerleri geliştirerek daha geniş bir YZ ekosisteminin büyümesini teşvik ediyorlar. Örneğin sağlık sektörü her üç platformun da yer aldığı bir sektör. GCP, tıbbi teşhis ve veri analizine yönelik YZ çözümlerini kolaylaştırmak için 'Cloud Healthcare API', 'Medical Imaging Suite' ve 'Healthcare Natural Language AI' gibi sağlık hizmeti odaklı ürünler sunuyor. Azure'un sağlık hizmetleri 'Health Data Services' ve 'Azure Data Lake Storage'. Bunların yanında YZ destekli bir sağlık botu hizmeti de var. AWS, yapay öğrenme modellerini kullanarak tıbbi ve sigorta verilerini dönüştürmek için 'AWS Health', 'HealthLake' hizmetlerini sağlıyor.^[6]

Perakende sektörü, Amazon'un uzmanlık alanı olmasına karşın diğer iki bulut şirketinin de perakende sektörünü hedefleyen çözümleri var. Örneğin, Google perakende sektöründe, 'Discovery AI for Retail' ve 'Recommendations AI' gibi çözümlerle dönüşüm oranlarını iyileştirmeyi ve kişiselleştirilmiş ürün önerileri sunmayı amaçlıyor. Müşteri etkileşimi ve iletişim merkezleri için AWS'nin 'Amazon Connect' ve GCP'nin 'Contact Center AI' hizmetleri var. Veri yönetimi ve analizi için 'BigQuery', 'Amazon DocumentDB' ve 'Azure Synapse Analytics' kullanılabiliyor.

YZ'nin endüstrileşmesi^[7] olarak adlandırılan bu süreç, Büyük Teknoloji şirketlerinin daha gelişmiş ve kapsamlı YZ hizmetleri sağlamasıyla şekilleniyor. Şirketler, söz konusu hizmetleri sağlarken YZ için kritik bir altyapı hizmet sağlayıcısı haline geliyorlar.

2.2.3. Yeni YZ Girişimlerinin Bulut Bilişim Platformlarına Bağımlılığı

Microsoft'un OpenAI'ye yaptığı yatırımlar sadece nakit değildi. Paranın bir kısmı işlem gücü için bulut kredisiydi. 2019'dan beri OpenAI modelleri, Microsoft'un Azure platformunda eğitiliyor ve çalıştırılıyor. Mevcut YZ; veri, kod ve hesaplamanın bir araya getirilmesinin bir sonucu. Bu üç parçadan biri eksikse YZ projelerinin hayata geçirilmesi olanaksızlaşmaktadır.

Veri ve özellikle hesaplama gücü üç şirketin en önemli bir avantajı. Bu avantaj sayesinde söz konusu şirketler, bulutlarında YZ alanına özgü uygulamaların geliştirilmesine yardımcı olan ve giderek büyüyen YZ hizmeti setleri sunuyorlar. Bu hizmetler, YZ'nin benimsenmesini daha hızlı ve kolay hale getirirken yeni girişimlerin hareket alanını kısıtlıyor ve bir bulut platformuna bağımlılıklarını artırıyor. Bir zamanlar, IBM gibi bilgisayar üreticilerinin yeni yazılımları teşvik ederek bilgisayarlarını daha işlevsel hale getirmesi gibi bulut platformları da platformları üzerinde uygulama geliştirmeyi teşvik ederek platformların değerini artırıyorlar. Bunun yanında yeni uygulamalar, platformları veri gücünü nicelik ve nitelik yönünden artırarak bulut devlerinin konumunu güçlendiriyor.

Platformlar, kendi modellerini eğitemeyen ya da eğitmeyi tercih etmeyen kuruluşlar için YZ algoritmalarını eğitmek üzere önceden yapılandırılmış sanal makine şablonları da sunuyor. Microsoft, OpenAI'nin hesaplama ihtiyaçlarından yola çıkarak YZ sistemleri için standart bir çözüm geliştirmeye çalıştı. Şimdi bu çözüm, sohbet robotlarını ve diğer özel YZ çözümlerini eğitmek ve kullanmak için bir hizmet olarak diğer müşterilere sunuluyor.

Yeni YZ firmaları için bulut platformları özellikle ekonomik belirsizlikler karşısında bir esneklik sağlıyor. Ayrıca bir veri merkezine yatırım yapmak için harcanabilir fonlara ve kendi özel çözümlerini çalıştırmak için gereken daha genel bileşenlerin bakımını ve kodlamasını yapacak çalışanlara sahip değiller. Dolayısıyla bulut platformlarının bilgi işlemin farklı unsurlarını bir hizmet olarak sunarak yeni girişimlerin önündeki engelleri azalttığını söyleyebiliriz. Diğer yandan yeni girişimler için tam bir bağımlılık ilişkisi söz konusu ve bir çıkış yolları bulunmuyor.

Kısacası, YZ birkaç şirketin egemenliğinde ve onların çıkarları doğrultusunda gelişiyor. Facebook'un eski çalışanlarından Jeff Hammerbacher'in dediği gibi bir kuşağın en iyi beyinleri vakitlerini reklamları daha fazla tıklatmaya harcadılar. Şimdi de durum pek farklı değil, teknoloji insanlar arasındaki eşitsizliği artıran, Büyük Teknoloji şirketlerini güçlendiren ve toplumun büyük bir kesimini yoksullaştıran bir biçimde kullanılıyor.

2.2.4. YZ'de Söz, Yetki, Karar Hakkı

Büyük Teknoloji şirketleri kendiliklerinden ayrıcalıklı konumlarından vazgeçmeyeceklerine göre yurttaşların daha adil ve eşit bir dünya için harekete geçmesi gerekiyor. Bu doğrultuda, bir birbirini tamamlayacak üç strateji uygulanmalıdır.^[8]

- ❖ **Yurttaşların YZ okur yazarlığı artırılmalıdır.** Okuryazar vatandaşlar YZ hakkında yeterli bilgi ve anlayışa sahip olacak, gündelik hayatlarında bunları kullanmanın olağan yollarını belirleyebilecek ve şirket çıkarlarına karşı toplumsal çıkarları öne çıkarabileceklerdir.
- ❖ **Yurttaşların, gündelik hayatlarını etkileyen YZ sistemlerinin düzenleme süreçlerine aktif katılımı sağlanmalıdır.** Başta kamu kurumlarında ve işyerleri olmak üzere, YZ teknolojilerinin hayata geçirildiği yerde bu teknolojilerden etkilenen yurttaşlar ve çalışanlar söz, yetki ve karar hakkında sahip olmalıdır. Teknolojinin kullanımının bazı sosyal ve ekonomik sonuçları olduğu hiçbir zaman unutulmamalı; bu teknolojileri düzenleyen kuralların yukarıdan aşağıya talimatlarla değil halkın katılımıyla oluşturulması için mücadele edilmelidir.

- ❖ **Yurttaşlar, YZ oluşturma sürecine dahil olmalıdır.** Yurttaşların YZ'nin kolektif geleceğini (ve dolayısıyla kendi geleceklerini) şekillendirmek ve potansiyel sonuçlarını ifade etmek için kendi bakış açıları ve mevcut uzmanlıkları doğrultusunda müdahalelerde bulunabilmesinin koşulları yaratılmalıdır.

2.3. AÇIKLANABİLİRLİK

2.3.1. Açıklanabilir Yapay Zekâ'nın Tanımı ve Önemi

Açıklanabilir Yapay Zekâ (AYZ-Explainable Artificial Intelligence/XAI) terimi, karmaşık yapay zekâ sistemlerinin karar verme gerekçelerinin insan tarafından anlaşılabilir çıktılar kümesine dönüştürülmesini ifade eder. Bu tür çıktılar, kural kümelerinden girdi görüntülerinin ısı haritalarına kadar değişebilir ve örneğin görseller için genellikle bir YZ modelinin karar vermek için kullandığı görsel bölümlerini belirtir.

Kullanıcı dostu olma ve şeffaflık gibi AYZ ile ilgili terimler onlarca yıldır tartışılrsa da derin öğrenmeye dayalı YZ algoritmalarının hızlı gelişmesiyle birlikte, daha karmaşık açıklanabilirlik yöntemlerine ve bununla ilgili terimlerin daha net tanımlarına ihtiyaç duyulmuştur. Bu gereksinimin arkasındaki neden, derin sinir ağlarının içsel yapısıyla güçlü bir şekilde ilişkilidir. Bu tür ağlar, çok sayıda veriyi birçok iterasyonda işleyerek ağ ağırlıklarını (girdiye dayalı olarak ağın amacını gerçekleştiren ağ parametreleri) değiştiren algoritmalar temelinde eğitilir. Parametre sayısı genellikle milyonlardan milyarlara kadar değişmektedir.

2.3.2. AYZ'nin Amaçları

Yukarıda bahsedilen nedenlerden ötürü, derin sinir ağlarının davranışını insanlar tarafından açıklamak neredeyse imkansızdır. İçsel karar verme kuralları insan anlayışına kapalı olan bu tür açıklanması zor (veya neredeyse imkânsız) YZ modelleri, kara kutu modelleri olarak adlandırılır. YZ günlük hayatımızda giderek daha etkili hale geldikçe, finans, sağlık hizmetleri ve otonom sürüş gibi kritik alanlarda kullanılan uygulamaların sayısı artmaktadır. Bu durum, hem kullanıcıların hem de geliştiricilerin YZ tarafından verilen kararların altında yatan gerekçeleri açıklamak istemesi nedeniyle AYZ'ye olan ihtiyacı daha da önemli hale getirmektedir.

2.3.3. Kimlerin Neden AYZ'ye İhtiyacı Var?

AYZ özellikle üç grup için önemlidir.^[9]

- YZ sistemini kullanan kişiler
- YZ kararlarından etkilenen kişiler
- YZ algoritmalarının geliştiricileri

Birinci gruba örnek olarak, otomatik olarak oluşturulan tanı raporları öneren bir YZ sistemini kullanan tıp doktorları verilebilir. Alan uzmanı olarak bir tıp doktoru, genellikle daha ileri incelemeler yapmak için tanının arkasındaki gerekçeyi anlamak isteyecektir. İkinci grup, yukarıdaki senaryoda hasta olan veya YZ tabanlı bir karar verme sistemi tarafından kredisi reddedilen kişi gibi YZ uygulamasının son kullanıcılarından oluşur. Bu kullanıcılar da kararın hayatlarını doğrudan etkilemesi nedeniyle gerekçesini anlamak isteyeceklerdir. Son olarak, üçüncü gruptaki YZ sistemlerinin geliştiricileri, modellerinin sağlamlığını artırmak ve zayıf yönlerini tespit etmek için davranışlarını anlamak isteyeceklerdir. Geliştiriciler tarafından bir diğer kullanım senaryosu ise verilerdeki ön yargıyı tespit etmek ve nasıl ortadan kaldırılacağını bilmektir.

2.3.4. Düzenleme ve Politika

❖ DARPA'nın AYZ Programı

2017'de önerilen program, 2017-2021 yılları arasında gerçekleştirilmiştir. Programın temel amacı, "daha açıklanabilir modeller üreten" ve "insan kullanıcıların ortaya çıkan yapay zekâlı ortakları anlamasını, uygun şekilde güvenmesini ve etkili bir şekilde yönetmesini sağlayan" makine öğrenimi teknikleri geliştirmektir. Program temel olarak açıklanabilir öğreniciler geliştirmeye, açıklanabilirliği iyileştirmek için insan psikolojisini anlamaya ve AYZ tekniklerini değerlendirmeye odaklanmıştır. Çeşitli araştırma gruplarının katılımı ve değerlendirme prosedürlerine katkıda bulunan birçok kullanıcı ile yapılan çalışmalarla, AYZ hakkında önemli iç görüler elde edilmiştir. (Ayrıntılı bilgi için bkz.^[10])

❖ Avrupa Birliği (AB)

AB'nin AYZ konusundaki düzenlemeleri, YZ sistemlerinin güvenilir ve güvenli olmasını sağlayan, AB'nin temel haklar değerlerine bağlı yasal bir çerçeveye olan bağlılığını yansıtmaktadır.^[11] AB'nin 2018 YZ stratejisinde vurgulanan yaklaşımı, başlangıçta vatandaş haklarını korurken inovasyona da vurgu yapmış, YZ şeffaflığını, hesap verebilirliğini ve insan gözetimini savunmuştur. Genel Veri Koruma Yönetmeliği (GDPR), otomatik sistemler tarafından alınan kararlar için bir "açıklama hakkı" kavramını ortaya koymuş, ancak uygulanması belirsizliği ve pratik uygulanabilirliği konusunda tartışmalara yol açmıştır.

2019'da Yapay Zekâ Üzerine Yüksek Düzeyli Uzman Grubu'nun (HLEG) Güvenilir YZ için Etik Yönergeleri, kullanıcı güveni için temel olan "açıklanabilirlik" kavramını daha da geliştirmiş, teknik açıklanabilirlik ile YZ sistemlerinin amaçlarının ve kararlarının daha geniş anlaşılması arasında ayırım yapmıştır.

12 Temmuz 2024 tarihinde AB Resmî Gazetesi'nde yayımlanan ve 1 Ağustos 2024 tarihinde yürürlüğe giren Avrupa Birliği Yapay Zekâ Yasası (EU AI Act), açıklanabilirlik ve şeffaflık ilkelerini ön planda tutarak, yapay zekâ sistemlerinin güvenli ve etik bir şekilde kullanılmasını hedefleyen kapsamlı bir düzenlemedir. Yasa, AI sistemlerini risk seviyelerine göre sınıflandırarak yüksek riskli uygulamalarda açıklama yükümlülükleri getirmektedir. Örneğin, kredi değerlendirmeleri veya sağlık teşhis sistemleri gibi kritik alanlarda kullanılan yüksek riskli modellerin, karar alma süreçlerinin şeffaf olması ve açıklanabilirlik mekanizmalarının sağlanması gerekmektedir. Bu bağlamda, yasa yalnızca sistemin çıktısını değil, aynı zamanda karar süreçlerini ve kullanılan verilerin model üzerindeki etkisini de açıklayabilmeyi zorunlu kılar. EU AI Act, açıklanabilirliği artırmak için geliştiricilere raporlama, risk değerlendirme ve insan denetimi gibi sorumluluklar yükleyerek, bireylerin yapay zekâ kararlarına güven duymasını sağlamayı amaçlamaktadır.^{[12],[13]}

❖ Amerika Birleşik Devletleri (ABD)

AYZ'nin önemi, ABD başkanı tarafından Ekim 2023'te imzalanan "Yapay Zekâ Üzerine Yürütme Emri"nde de vurgulanmış, iki ana alana odaklanılmıştır: YZ modellerinin şeffaflığı ve açıklanabilirliği. Düzenleyici kurumlar, kuruluşlar tarafından kullanılan YZ modellerinin şeffaf olmasını sağlamaya odaklanmaktadır; bu, bu modellerin iç işleyişlerinin ve veri işleme yöntemlerinin açık ve anlaşılır olması gerektiği anlamına gelir.

❖ Birleşik Krallık (UK)

İlk YZ Güvenlik Zirvesi Kasım 2023'te BK'da gerçekleşmiş, 28 ülkeden yetkililer, YZ endüstrisi liderleri ve önde gelen araştırmacılar bir araya gelerek YZ teknolojilerinin geliştirilmesi ve denetlenmesine odaklanmıştır. Zirve sonrası yayınlanan bildirgede^[14] AYZ'nin gerekliliği şöyle belirtilmiştir: "YZ sistemlerinin potansiyel etkisini mevcut forumlarda ve diğer ilgili girişimlerde inceleme ve ele alma yönündeki uluslararası çabaları ve insan haklarının korunması, şeffaflık ve açıklanabilirlik, adillik, hesap verebilirlik, düzenleme, güvenlik, uygun insan gözetimi, etik, ön yargı azaltma, gizlilik ve veri korumasının ele alınması gerektiğinin tanınmasını memnuniyetle karşılıyoruz."

2.3.5. AYZ Açıklama Tipleri

❖ Sayısal Açıklamalar

Sayısal çıktıların tipik bir örneği özellik öneminin (feature importance) belirlenmesidir. Makine öğrenimi ve yapay zekâ bağlamında özellik önemi, bir yemek tarifindeki hangi malzemelerin yemeğin başarısı için en kritik olduğunu anlamaya benzer. Bir pasta yaptığınızı düşünün. Un ve yumurta gibi belirli malzemeler temeldir- bunlar "önemlilik" derecesi yüksektir çünkü bunların değiştirilmesi pastanın sonucunu önemli ölçüde etkiler.

Evinizi satmak istediğinizi ve piyasa fiyatını tahmin eden bir uygulama kullandığınızı düşünün. Bu uygulama evinizin büyüklüğü, yaşı, konumu, yatak odası ve banyo sayısı gibi ayrıntıları dikkate alır. Bu bilgileri analiz ettikten sonra, uygulama size tahmini fiyatı ve her ayrıntının bu fiyatı basit sayılarla nasıl etkilediğini söyler.

❖ Kural Tabanlı Açıklamalar

Yoga duruşlarınızı fotoğraflar aracılığıyla değerlendiren bir fitness uygulaması örneğinde olduğu gibi, uygulama karmaşık bir algoritma kullanarak postürünüzü, hizalamanızı ve formunuzu analiz eder. Uygulama, analizi 1'den 10'a kadar bir ölçekte basit bir sayısal puana dönüştürür ve "Puanınız yedi, çünkü daha iyi denge için kollarınızın daha düz ve ayaklarınızın daha geniş olması gerekiyor" gibi kural tabanlı açıklamalar sunar.

❖ Metinsel Açıklamalar

Sosyal medyada maksimum etkileşim için fotoğraf paylaşmanın en iyi zamanını tahmin eden bir uygulama örneğinde, karmaşık bir makine öğrenimi modeli takipçilerin en aktif olduğu zamanlar ve hangi içerik türünün daha fazla beğeni aldığı gibi örüntüleri analiz eder. Uygulama, verileri analiz ettikten sonra "Takipçileriniz hafta sonları akşam saatlerinde en aktif durumda. Açık hava aktiviteleriyle ilgili fotoğraflar %20 daha fazla beğeni alıyor. Buna dayanarak, en iyi etkileşim için doğa yürüyüşü fotoğraflarınızı bu Cumartesi akşam saat 19.00'da paylaşmanızı öneriyoruz" gibi anlaşılır metinsel açıklamalar sunar.

❖ Görsel Açıklamalar

Örnek 1

Uyku düzeniniz, fiziksel aktiviteniz, kalp atış hızınız ve günlük programınıza dayanarak stres seviyenizi tahmin eden bir sağlık uygulaması, karmaşık veri analizini görsel bir grafikte sunar. Bu grafik, farklı faktörleri temsil etmek için renkleri ve boyutları kullanır: uyku için mavi, fiziksel aktivite için yeşil, kalp atış hızı için kırmızı ve programınız için sarı. Her faktörün stres seviyeniz üzerindeki etkisi, grafikteki renkli bölümünün büyüklüğüyle gösterilir.

Örnek 2

İçeriklerine göre fotoğraflarınızı otomatik olarak kategorilere ayıran bir fotoğraf yönetim uygulaması, görüntüleri analiz etmek için karmaşık bir makine öğrenimi modeli kullanır. Uygulama bir fotoğrafı neden belirli bir şekilde kategorize ettiğini göstermek için, kararını etkileyen görüntü alanlarını vurgular. Örneğin, bir fotoğrafı manzara olarak kategorize ederken gökyüzünü ve dağları vurgulayabilir.

❖ Karma Açıklamalar

Evinizdeki malzemelere göre tarifler öneren akıllı bir mutfak uygulaması örneğinde, uygulama önerisini karma bir yaklaşımla açıklar:

- Görsel olarak: mutfak dolabında tanınan malzemeleri vurgulayarak
- Metinsel olarak: malzemeleri listeleyerek ve tarifi neden iyi bir seçim olduğunu açıklayarak
- Sayısal olarak: tarifi tercihlerinize ve malzeme mevcudiyetine uygunluğunu gösteren bir puan sunarak

❖ Teknik Yaklaşımlar ve Algoritmalar

AYZ, kritik uygulamalarda kara kutu tabanlı makine öğrenimi algoritmalarının artan kullanımı nedeniyle son dönemde öne çıkan bir alan olduğundan, hem teknik detaylar hem de AYZ algoritmalarının sınıflandırılması konusunda çok sayıda yayın bulunmaktadır. Son zamanlarda birçok yayın, AYZ metodolojileri için bir sınıflandırma sistemi önermiştir, ancak bunlara dayanarak ortak bir hiyerarşi oluşturmak oldukça karmaşık bir görevdir. Bu nedenle, AYZ algoritmalarının farklı özelliklerini ele alan çok yönlü bir sınıflandırma oluşturmak anlamlı bir yaklaşımdır. Aşağıda, önerilen araştırma ve öneriler temelinde^[15], bu çok yönlü sınıflandırmanın çeşitli bölümlerini ve alt sınıfları özetlenmiştir.

❖ İşlev Tabanlı Yaklaşımlar

- **Yerel Bozunum ile (Pertürbasyonlarla) Açıklama:** ML modelinin giriş verilerini bozarak verilerdeki en anlamlı özellikleri bulmaya çalışır
- **Yapıyı Kullanma:** ML modellerinin içsel özelliklerini davranışlarını açıklamak için kullanır
- **Meta Açıklamalar:** Diğer açıklama yöntemlerinin çıktılarını birleştirerek modeli daha iyi anlamayı sağlar
- **Mimarinin Değiştirilmesi:** ML modellerinin karmaşık mimari yapılarını basitleştirerek anlaşılabilirliği artırır
- **Örnekler:** Test edilen verilerden seçilen ve modelin davranışını açıklayan örneklerdir

❖ Sonuç Tabanlı Yaklaşımlar

Sonuç tabanlı yaklaşımlar üç temel kategoride incelenir:

Özellik Önemi Derecelendirilmesi: Bu yaklaşım, sonuçların hangi giriş verisi özellikleri ile ilişkili olduğunu belirler. Önceki bölümlerde belirtildiği gibi, birçok AYZ algoritması özelliklerin önem derecesini puanlayarak sonuç üretir.

Vekil Modeller: Bu modeller, belirli bir ML modelinin giriş-çıkış çiftlerini kullanarak vekil model eğitir. Bu çiftler, modelin davranışını açıklamak amacıyla kullanıldığından, doğru sonuçlar vermesi bile açıklama amaçlı kullanılırlar. Vekil modeller genellikle yüksek derecede yorumlanabilir olup, açıklamaların çıkarılmasını kolaylaştırır.

Temsili Örneklemeler: Bu yöntemde, bir modelin eğitim setindeki veriler, güvenilirliği yüksek veya düşük olarak sınıflandırılarak seçilir ve açıklanabilirlik için temsili örnekler olarak kullanılır.

❖ Kavramsal Yaklaşımlar

Kavramsal yaklaşımlar üç ana boyutta ele alınır:

Evre: AYZ algoritmaları, ante-hoc veya post-hoc olarak sınıflandırılır. Ante-hoc yöntemler, modelin eğitim süreci sırasında açıklamalar oluştururken, post-hoc yöntemler yalnızca daha önce eğitilmiş modeller üzerinde çalışır.

Uygulanabilirlik: AYZ algoritmaları, modelden bağımsız veya modele özgü olarak iki şekilde sınıflandırılır. Modelden bağımsız yöntemler tüm modellere uygulanabilirken, modele özgü yöntemler yalnızca belirli model türleriyle çalışır.

Kapsam: AYZ algoritmaları yerel veya evrensel olarak ikiye ayrılır. Evrensel yöntemler tüm veri noktalarını kapsayan genel açıklamalar sağlarken, yerel yöntemler belirli veri örnekleriyle sınırlı açıklamalar sunar.

❖ Karma Yaklaşımlar

Karma yaklaşımlar, yukarıda belirtilen kavramların hibrit versiyonlarını kullanarak daha kapsamlı bir sınıflandırma hiyerarşisi oluşturur. Bu yaklaşım, farklı yöntemlerin güçlü yanlarını birleştirerek daha etkili açıklamalar sunmayı hedefler.

2.4. SÜRDÜRÜLEBİLİRLİK

YZ sürdürülebilirlik hedeflerini desteklediği gibi kendisi de sürdürülebilir olmalıdır.

Sürdürülebilir kalkınma, bugünün ve yarının toplumunu bir bütün olarak değerlendirir. Kalkınma, günümüzün ihtiyaçlarını karşılarken gelecek neslin ihtiyaçlarını karşılama yetisine zarar vermemelidir. Sürdürülebilir kalkınma kavramının başlıca üç boyutu vardır: ekonomik sürdürülebilirlik, sosyal sürdürülebilirlik ve çevresel sürdürülebilirlik. Sürdürülebilir kalkınma, kalkınma ile çevre, ekonomi ve toplumun ihtiyaçlarına hizmet etme arasındaki gerilimi aşmaya çalışır.

Bu ilkede odak noktamız YZ'nin çevresel sürdürülebilirliğidir ve sürdürülebilirlik, iki açıdan ele alınmaktadır:

- Sürdürülebilirlik için YZ'nin Kullanımı
- YZ'nin Sürdürülebilirliğinin Sağlanması

2.4.1. Sürdürülebilirlik için YZ

Sürdürülebilirlik için YZ tartışmalarında YZ, sürdürülebilirlik hedeflerini gerçekleştirebilmek için kullanılabilecek bir araç olarak değerlendirilir. Sürdürülebilirlik hedefi doğrultusunda kullanılabilen YZ araçlarını iki başlık altında inceleyebiliriz. Birincisi, takip ve tahmin yapan YZ sistemleridir: Güneş ışınımı tahmini, rüzgâr hızı tahmini, ülkeye özgü enerji ithalat tahminleri, kısa vadeli güç yükü tahmini, bina enerji kullanımı ve güç tüketimi tahminleri, aşırı hava olayları tahmini, uygun habitat tahmini ve koruma planlaması vb. İkincisi, bir eylemi gerçekleştiren ajansal YZ sistemleridir: Tarımda pestisit azaltımı, akıllı şebekelerde optimal (yenilenebilir) enerji yönetimi ve mikro şebekelerde güç akışı optimizasyonları, optimal trafik çizelgeleme teknikleri vb.

YZ, özellikle enerji sektöründe sürdürülebilirlik hedeflerinin gerçekleştirilmesinde etkili olabilir. Örneğin bireysel hanelerin elektrik yükünün kısa vadeli tahmini ve akıllı şebeke yönetimi/planlaması zor bir iştir. Elektrik tüketim profili, farklı çevresel ve mevsimsel etkilere bağlı olarak saatlik, günlük, haftalık ve mevsimsel olarak değişebilmektedir. YZ, bireysel hanelerin elektrik tüketim örüntüsündeki öngörülemezlik ve belirsizlikleri azaltarak planlama sürecini iyileştirebilir.

YZ'nin sürdürülebilirlik için kullanımı enerji sektörüyle sınırlı değildir. Örneğin sürdürülebilir tarımda, mahsul koruma optimizasyonunda kullanılan YZ çözümlerinin mantar ilacı kullanımını ve tank sızıntılarını büyük oranda azaltarak çevre kirliliğini azalttığı görülmektedir. Brezilya'da da bilgisayar görüşü kullanılarak yabancı ot ilaçlamasında, herbisit kullanımında ve su tüketiminde kayda değer bir tasarruf sağlanmıştır.

2.4.2. YZ'nin Sürdürülebilirliği

YZ'nin sürdürülebilirliği, YZ sistemlerinden kaynaklanan karbon ayak izini ölçmenin ve azaltmanın bir yolu olarak sürdürülebilir veri kaynaklarına, güç kaynaklarına ve altyapılara odaklanır. İnternette bir şeyler yapmak, ister insan kullanıcılar tarafından ister YZ tabanlı uygulamalar tarafından yapılsın, elektrik gerektirir. Ne yazık ki bu dijital altyapıyı çalıştırmak için gereken elektrik hâlâ büyük ölçüde fosil yakıt bazlı enerji tedarikinden sağlanmaktadır. Bu nedenle, enerji yoğun YZ tabanlı süreçlerin çevreye etkileri kritik bir sorun olarak karşımıza çıkmaktadır. Bulut altyapıları için harcanan enerji miktarı her geçen gün artmaktadır. Ayrıca veri merkezlerinin inşa edildiği bölgelerde su, ona ihtiyacı olanlardan alınarak sunucuları soğutmak için kullanılır. Böylece YZ sistemlerinin eğitildiği ve çalıştığı bulut altyapıları, su ve elektriği hesaplama gücüne dönüştürürken büyük miktarda çevresel hasara neden olurlar.

YZ modellerinin eğitim süreci binlerce megawatt saat elektrik tüketebilir ve yüzlerce ton karbon salabilir. Ayrıca, YZ model eğitimi, veri merkezlerindeki ısı atımı için çok miktarda tatlı suyun buharlaşmasına yol açar ve zaten sınırlı olan tatlı su kaynaklarına zarar verir. Özellikle fosil yakıtların yakılması yoluyla elektrik üretimi, hava kirliliğine, su kaynaklarında termal kirliliğe ve tehlikeli maddeler de dahil olmak üzere katı atıkların üretilmesine neden olur. Arizona ve Şili örneklerinde görüldüğü gibi, sunucu soğutma ve elektrik üretimi için kullanılan su, kuraklıkların artmasına neden olabilir.

Büyük teknoloji şirketleri son yıllarda sera gazı emisyonları konusunda bazı büyük iddialarda bulunsalar da YZ teknolojisinin (ve özellikle de üretken YZ'nin) yükselişiyle beraber enerji ihtiyacı artmıştır. Ayrıca şirketler, enerji ve su kullanımı konularında şeffaf davranmamaktadır. YZ'nin çevreye etkileri hakkında birçok sorun ancak araştırmacı gazetecilerin gayretleriyle ortaya çıkarılabilmektedir. Guardian'ın analizine göre, 2020'den 2022'ye kadar Google, Microsoft, Meta ve Apple'ın "şirket içi" veya şirkete ait veri merkezlerinden kaynaklanan gerçek emisyonlar muhtemelen resmi olarak bildirilen değerlerin 7,62 katıdır.^[16]

YZ'nin sürdürülebilirliği hakkındaki bir diğer sorun da sürdürülebilirlik tartışmalarında genellikle toplam karbon emisyonu ve su tüketimi gibi kolayca ölçülebilen çevresel ölçütlere odaklanılmasıdır. Buna karşın, YZ'nin çevresel maliyetlerinin farklı bölgeler ve topluluklar arasında eşit olarak dağıtılması zorunluluğu olan çevresel eşitliğe yeterince önem verilmemektedir. Birçok durumda, YZ'nin olumsuz çevresel etkileri özellikle yoksul ve savunmasız bölgelerde daha fazladır. Örneğin, 2022 yılında Google Finlandiya'daki veri merkezini %97 oranında karbonsuz enerji ile işletirken, bu oran Asya'daki veri merkezleri için %4-18'e düşmektedir.^[17]

Küresel kuzeyde enerji maliyetlerinin ve hukuksal düzenlemelerin artması sonucunda veri merkezleri giderek daha fazla hukuksal düzenlemelerin zayıf olduğu küresel güneye yönelmektedir.^[18] Zaten temiz suya erişimde ciddi sıkıntılar yaşayan bölge halkı şimdi de suyunu büyük teknoloji şirketlerinin veri merkezleriyle paylaşmak zorunda kalacaktır.

Veri merkezleri, YZ'nin can damarıdır. Fakat bu merkezler, suyunu ve enerji kaynaklarını paylaşmak zorunda olacak halkın rızası ve çıkarları doğrultusunda inşa edilmeli ve çalışmalıdır. Küresel kuzeyde geçerli hukuk, küresel güneyde de geçerli olmalıdır.

Sürdürülebilirlik hedeflerine ulaşmak için, öncelikle YZ'nin sürdürülebilir veri kaynakları ve enerji altyapılarıyla entegre bir şekilde kullanılması gerekir. Bu bağlamda, enerji yoğun YZ süreçlerinin çevresel etkilerini minimize etmek amacıyla yenilenebilir enerji kaynaklarına geçiş teşvik edilmelidir. Ayrıca, YZ sistemlerinin şeffaflık ve etik çerçeveler içinde çalışması sağlanarak, yurttaşların farkındalığı ve toplumun çeşitli kesimlerinin bu süreçlere katılımı artırılmalıdır. Farklı disiplinlerden uzmanların bir araya gelerek, YZ'nin çevresel, sosyal ve ekonomik sürdürülebilirlik hedefleri doğrultusunda bir araç olarak nasıl kullanılabileceğini belirlemesi de kritik bir adım olacaktır. Böylece, kalkınma ile çevre, ekonomi ve toplum arasındaki gerilimi aşma çabası daha somut ve etkili hale gelecektir.

2.5. ÖNYARGI

Kamuoyu, YZ'nin tarafsız olmadığı ve tam olarak nesnel olamayacağı hakkında bilinçlendirilmelidir. YZ sistemlerinin neden olabileceği veya artırabileceği yanlılıkların (bias) önüne geçebilmek için kullanılan şirketler ve kurumlar, veri setleri, geliştiren ekibin profili ve kullanılan algoritmalar hakkında şeffaf olmalıdır.

YZ tabanlı karar sistemlerinin insan karar vericilere göre daha nesnel olduğu varsayılır. YZ'nin kararlarının nesnel hesaplamalara dayandığı, bilgisayarlar zaman zaman hatalı karar verseler bile söz konusu hataların teknik bir sorundan kaynaklandığı düşünülür. Bilgisayarların kararlarının da ön yargılı (bias) olabileceği ihtimali göz ardı edilir. Nitekim yaygın kanı ön yargının insani bir özellik olduğu ve sayısal doğrularla hareket eden makinelerin ön yargısız kararlar verdiğidir.

Ancak tam da ön yargının insani bir özellik olmasından dolayı ön yargı, YZ'nin (özellikle derin öğrenme sistemlerinin) içsel bir özelliğidir ve YZ etiği tartışmalarında en çok üzerinde durulan konularından biridir. YZ sistemlerinde ön yargı başlıca iki nedenden kaynaklanabilir.

2.5.1. Geliştiricilerden Kaynaklı Ön Yargılar

Teknolojiyi geliştiren aktörlerin değer yargıları, geliştirilmesinde rol aldıkları teknolojiyi şekillendirir. Teknoloji, iyi ya da kötü olmadığı gibi tarafsız da değildir. Teknoloji, onun şekillenmesinde etkili olan ve belirli tarihsel koşullar altında hareket eden aktörlerin çıkarlarını ve değer yargılarını içerir. Örneğin, YZ sistemlerinin geliştirilmesinde etkili olan uzmanlar, ağırlıklı beyaz erkeklerden oluşur. Teknoloji sektöründe çalışan kişilerin eğitimleri, sosyal geçmişleri, ilgi alanları, rol modelleri ve yaşam tarzları birbirine yakındır. Bu da ister istemez bir sınırlılığa ve çeşitli ön yargıların modellerin içine sızmasına neden olur. Herkes gibi YZ modellerini tasarlayanlar da hata yapabilir. Ama daha da önemlisi dünyaya belirli bir sosyal geçmişin ve profesyonel ortamın kültürel merceklerinden bakar. Bu nedenle, belirli hesaplama hedeflerinin peşinden gitmenin veya belirli parametreleri ve değişkenleri seçmenin sonuçlarını genellikle görmezden gelirler veya hafife alırlar.

Bir eğitim seti oluşturmak, neredeyse sonsuz derecede karmaşık ve çeşitli ilişkiler içeren bir dünyayı soyutlamayı gerektirir. Problemin çerçevesinin çizilmesi, sistemin hedeflerinin belirlenmesi tarafsız bir eylem değildir. Örneğin, bir kredi kartı şirketinin müşterilerin kredi itibarını tahmin eden bir sistem geliştirirken verdiği kararlar sistemin gelişimine yön verir. Çünkü en başta kredi itibarının ne olduğu sorusunu yanıtlanması ve bu yanıtın sayısallaştırılması gerekir. Şirketin hedefi nedir? Kârını mı yoksa geri ödenebilen kredi miktarını mı en üst düzeye çıkaracaktır? Şirketin istediği kârını artırmaksa, algoritma da buna en uygun fonksiyonu modelleyecektir. Hedef, geri ödenebilen kredi miktarını artırmaksa başka bir modelleme gündeme gelecektir.

2.5.2. Veriden Kaynaklı Ön Yargılar

Ön yargı, YZ sistemlerinin eğitimlerinde kullanılan verilerden ve seçilen öz niteliklerden de kaynaklanabilir. Derin öğrenme sistemleri, geniş bir eğitim veri setinden genelleme yapabilecek ve eğitim veri setinde yer almayan yeni gözlemleri doğru bir şekilde sınıflandırabilecek şekilde tasarlanırlar. Başka bir deyişle, derin öğrenme sistemleri, yeni örneklerde (yeni başvuru sahiplerinin özgeçmişlerindeki kelime grupları gibi) hangi veri noktalarının aranacağına karar vermek için belirli örneklerden (geçmiş iş başvuru sahiplerinin özgeçmişleri gibi) öğrenerek bir tür tümevarım gerçekleştirebilir. Bu gibi durumlarda, ön yargı terimi, bu tahmin edici genelleme süreci sırasında ortaya çıkabilecek bir hata türünü, yani sistemin yeni örneklerle karşılaştığında sergilediği sistematik veya tutarlı bir şekilde tekrarlanan bir sınıflandırma hatasını ifade eder.

Algoritmanın eğitildiği veri gerçekliğin zayıf bir temsiliye veya mevcut toplumsal önyargıları barındırıyorsa algoritma da yanlış çıktılar üretecektir. Örneğin, krediler için risk puanlaması yapan bir YZ modeli, modelini eğitirken çoğunlukla erkeklerin onaylanmış kredilerini içeren bir veri setinden yararlanmışsa, erkeklerin daha risksiz müşteriler olduğuna meyilli olacaktır.

Verinin ön yargılara etkisi hakkındaki bir diğer ünlü örnek de Amazon'un iş başvurusu yapan adayların özgeçmişini değerlendirdiği otomasyon sistemidir. 2015 yılında şirket, iş başvurularının cinsiyetçi bir şekilde değerlendirildiğini fark etmiştir. Çünkü sistem, eğitim sürecinde ağırlıklı olarak erkek adayların öz geçmişleriyle eğitilmiş ve bunun sonucunda erkekleri daha tercih edilebilir adaylar olarak belirlemiştir. Bu nedenle, içinde kadın kelimesi geçen ("kadın satranç takımı kaptanı" gibi) özgeçmişler cezalandırılmış, aşağı sıralara itilmiştir.

2.5.3. Ön Yargıya Karşı Ne Yapılabilir?

Ön yargı, bilgisayar bilimcilerin farkında olduğu ve üzerinde çalıştığı önemli konulardan biridir. YZ geliştirme ekiplerinin daha çeşitli kesimlerden oluşmasını sağlamak (örneğin farklı cinsiyet ve etnik kökenden kişileri çalışmalara dahil etmek) veya YZ sistemlerinin eğitiminde kullanılan veri setlerinin seçiminde daha özenli olmak ön yargıyla mücadeleye katkıda bulunabilir. Fakat YZ sistemlerindeki ön yargı sorununun çözümü kolay değildir.

Birincisi, "çöp girer, çöp çıkar" sözü kusurlu istatistiksel modellerin ve girdi verilerinin çıktılarının da kusurlu olacağını ifade eder. Aslında YZ modellerinin girdileri ve çıktıları bizzat toplumun kendisidir. Bu nedenle, "çöp girer, çöp çıkar" yerine "toplum girer, toplum çıkar" daha yerinde bir ifade olabilir. Veri yanlışlığını üreten "çöp", makine tarafından okunabilen dijital izlere dönüşen asimetrik sosyal ilişkiler, kültür ve uygulamalar demeti olan toplumdan başka bir şey değildir.

İkincisi, YZ sistemlerini baştan sona ve farklı yeni bağlamları göz önünde bulundurarak tasarlayanın maliyetli olmasıdır. Algoritmaları tasarlarlarken nitelikli işgücü kullanmak, eğitim verisi bulmak, algoritmanın özelliklerini belirlemek, test etmek, iyileştirmek ve modeli uygulamak için para gerekir.

Bu maddi gereklilikler, bilgisayar bilimcilerin problemleri çözmeye çalışırken farklı bağlamlarda farklı görevleri yerine getirebilecek sistemler tasarlama (soyutlama) yaklaşımı ile birleşince belirli bir bağlama uygun olarak geliştirilen bir algoritmayı farklı bir bağlamda olduğu gibi veya değiştirerek kullanma gibi bir eğilim ortaya çıkmaktadır. Fakat bu yaklaşım özellikle toplumsal sorunların çözümünü için uygulandığında yeni sorunlara neden olur. Belirli bir yer için tasarlanmış bir sistem farklı toplulukların farklı adalet anlayışları nedeniyle başka bir yerde doğrudan uygulanamayabilir. Aynı şekilde belirli bir problem alanına yönelik tasarlanan bir sistem farklı bir problem alanı için yetersiz veya yanlış olabilir. Çünkü belirli bir bağlama göre oluşturulmuş standartlar ve algoritmanın ilgilendiği etik konular yeni uygulamada sorunlara neden olabilir.

Üçüncüsü, adalet kodlanamaz ve adaletin ne olduğu politik bir karardır. Adalet konusu, felsefe, sosyal bilimler ve hukukta yıllardır tartışılmaktadır ve şimdi de bilgisayar biliminin gündemindedir. Ama bilgisayar biliminin yükü daha fazladır; kendisinden adaleti matematiksel terimlerle açıklaması da beklenmektedir.

2.5.4. Tahminler ve Gerçekler

Aşağıdaki tabloda başlıca sorunlardan biri yanlış pozitiflerle yanlış negatifler arasındaki dengeyi bulabilmektir. Yanlış pozitifte, gerçekte negatif olan bir olgunun pozitif olduğunu tahmin etmişsinizdir (Birinin suçlu olduğunu tahmin ettiniz, ama gerçekte suçsuzdu). Yanlış negatifte ise gerçekte pozitif olan bir olgunun negatif olduğunu tahmin etmişsinizdir (Suçlu olan birine suçlu değil demek). Yani insanları haksız yere suçlama ihtimali de var, suçluları fark edemeyerek toplum güvenliğini tehlikeye atma ihtimali de. Algoritma, toplumun iyiliği için arada sırada suçsuz insanları mağdur etmeli midir (yanlış pozitif)? Yine bir denge kurmaya çalışsak bile insanların adalet hakkındaki düşüncelerinin de zaman içinde değiştiğini unutmamak gerekiyor.

GERÇEK DEĞERLER			
TAHMİN EDİLEN DEĞERLER	POZİTİF	POZİTİF	NEGATİF
		Gerçek Pozitif True Positive – TP	Sahte Pozitif False Positive – FP
	NEGATİF	Sahte Negatif False Negative – FN	Gerçek Negatif True Negative – TN

Özetle, ön yargı çeşitli teknik çözümlerle kısmen iyileştirilebilir olsa da YZ sistemlerine içsel bir olgudur. Sistemlerin optimizasyon ölçütleri ve istatistiksel eşitliğin ötesinde değerlendirilmesi ve matematik ve mühendislik çerçevelerinin nerede sorunlara yol açtığına anlaşılması gerekmektedir. Bu aynı zamanda YZ sistemlerinin verilerle, çalışanlarla, çevreyle ve kullanımından hayatları etkilenecek bireylerle nasıl etkileşime girdiğini anlamak ve YZ'nin nerede kullanılmaması(!) gerektiğine karar verebilmek anlamına gelmektedir.

En önemlisi de toplum YZ'nin tarafsız olmadığı ve tam olarak nesnel olamayacağı hakkında bilinçlendirilmelidir. YZ sistemlerinin neden olabileceği veya artırabileceği ön yargılarla mücadele edebilmek için YZ sistemlerini geliştiren şirketlerin veri setleri, geliştiren ekibin profili ve kullanılan algoritmalar hakkında şeffaf olması sağlanmalıdır.

2.6. MAHREMİYET

YZ teknolojilerini geliştiren şirketler insanların mahremiyet hakkına saygı duymalıdır.

Mahremiyetin tanımı ve sınırları konusunda bir ortaklaşma yoktur. Ancak paylaşılan ortak nokta her insanın belirli bir düzeyde ve derecede mahremiyete gereksinimi olduğudur. Mahremiyet hakkındaki yaklaşımlar, iki farklı boyutta ele alınabilir.^[19] Birinci boyut, özel ve kamu ayrımının nasıl yapıldığıdır. Mahremiyet nerede başlayıp, nerede bitmektedir. İkinci boyut ise mahremiyetin neden savunulduğu ve hangi gerekçelerle özel ve kamu arasında bir sınır çekme ihtiyacı duyulduğudur.

Birinci boyuta, özel ve kamu sınırına, baktığımızda farklı yaklaşımlar görürüz. Söz konusu sınırlamaları dört kategori altında inceleyebiliriz.^[19] Mekânsal, davranışsal, kararsal ve bilgisel sınırlamalar.

Mekânsal sınırlamalar: Doğal olarak başlıca sınır koyma, mekânsal alanda ve fiziksel olarak gerçekleşir. Örneğin, “insanın evi kendi kalesidir” ve “devletin yatak odamızda bir işi yoktur” gibi yaklaşımlar bu kapsamda değerlendirilebilir.

Davranışsal sınırlamalar: Belirli insani ilişkilerin, üçüncü şahısların müdahalesinden muaf tutulması üzerinde durur. Sevmek, sevilme, saygı duymak, güvenmek gibi insanlar arası ilişkiler, yanmanın oksijene ihtiyaç duyması gibi mahremiyete ihtiyaç duyar.^[20]

Kararsal sınırlamalar: Sınır, insanların kararları bağlamında da çizilebilir. Özellikle ABD Anayasası, mahremiyeti göz önünde bulundurarak insanların kararlarına müdahale edilemeyeceğinin altını çizer. Kürtaj, doğum kontrolü, belirli bir yaşam tarzı seçme, çocuklarını kendi dinine göre yetiştirme, cinsel yönelimler vb. hakkındaki kararlar bu kapsamda değerlendirilebilir.

Bilgisel sınırlamalar: Bilişim teknolojilerindeki gelişmelerle beraber kişiler hakkında toplanan bilgiler devasa boyutlara ulaşmıştır. Modern hayatın bir gereği olarak, insanlar çeşitli kurumlara ve şirketlere kişisel bilgilerini verirler veya vermek zorunda kalırlar. Burada sınır, verilen bilginin kişinin kontrolü dışında başka yerlerde kullanılmaması ve kişisel bilginin akışının sınırlandırılması şeklinde gerçekleşir.

Bu sınırlamalar mutlak değildir ve harfi harfine değerlendirilmemelidir. Örneğin, iki yetişkin arasındaki cinsel ilişki, özel bir alanda değil de kamuya açık bir alanda gerçekleşiyorsa burada mahremiyetin sınırlarından bahsetmek zorlaşır. Çocuklarını şiddet uygulayarak yetiştirmeyi tercih eden ebeveynlerin mahremiyeti de tartışmalıdır. Bu sınırlamalarda dikkat edilmesi gereken nokta, mahremiyeti tanımlamak için uygulanan sınırlamanın kişinin kendini toplumdan yalıtması ile değil, toplumsallığın sınırlarının çizilmesi ve toplumsal ilişkilerin yeniden düzenlenmesi ile ilgili olduğudur.^[19]

Mahremiyet tartışmasının ikinci boyutu ise mahremiyetin hangi nedenlerle savunulduğudur. Burada üç farklı bakış açısı söz konusudur.^[19] İnsani, siyasi ve araçsal nedenler.

İnsani nedenler: Mahremiyet, kendi başına insani bir anlam taşır. İnsanın kendi alanına zorla girilmesi, davranışlarına ve kararlarına müdahale edilmesi ve kişisel bilgileri hakkındaki kontrolünün kaybetmiş olması başlı başına insanlık onurunun çiğnenmesidir.

Siyasal nedenler: Mahremiyet, demokrasinin ön koşuludur. Örgütlenme özgürlüğü, bilimsel çalışmaların siyasi iktidarların müdahalesinden bağımsız olarak gerçekleştirilebilmesi, gizli oy hakkının güvence altında olması, medya organlarının bağımsız çalışması, polisin vatandaşların yaşamına gereksiz müdahalelerin önlenmesi için mahremiyetin sağlanması temeldir. Kısaca şöyle diyebiliriz, kişinin kamusal alandaki özgür faaliyeti, ancak ve ancak mahremiyetinin koruma altına alınması ile mümkündür. Muhalefetin telefonlarının, toplantılarının dinlendiği bir ortamda demokrasi söz konusu olamaz. Örneğin gizli oy ile kişilerin siyasi tercihlerinden dolayı baskı görmesinin önüne geçilebilir. Ancak kişiler dinleniyor/izleniyor ve fişleniyorsa gizli oyun da bir anlamı kalmaz. Farklı düşünen ve iktidar için potansiyel tehlike olarak görülen vatandaşlar, iktidarın hedefi haline gelir.

Araçsal nedenler: Burada stratejik bir bakış açısı söz konusudur. İtiraz, diğer iki maddede olduğu gibi kişisel bilginin toplanmasına değildir. İtiraz edilen nokta, bu bilginin farklı amaçlar için farklı kişiler tarafından kullanımıdır.

Ancak mahremiyet, 21. yüzyılda yeni bir boyut kazanmış ve mahremiyet ihlalleri artmıştır. Burada üç kritik gelişme etkili olmuştur.^[21] Birincisi, 11 Eylül 2001'de ABD'de gerçekleşen terörist saldırılar ve ardından ABD'den dünyaya yayılan kitlesel gözetimdir. 11 Eylül saldırıları sonrası ABD hükümeti, şüphelilere odaklanan bir güvenlik politikası yerine toplumun tamamını izleyen bir gözetim politikasına yöneldi. Bunun yanında, bilişim şirketlerinin veri politikalarını düzenleyen politikalar rafa kaldırıldı.

Edward Snowden'in ifşaatlarıyla öğrendiğimiz PRISM skandalı, ABD hükümeti ve şirketlerin işbirliğinin boyutu hakkında önemli ipuçları veriyordu. İkinci önemli gelişme ise başta "Gözetim Kapitalizmi"nin^[22] iki büyük aktörü Google ve Facebook olmak üzere teknoloji şirketlerinin mahremiyetin modası geçmiş bir değer olduğunu iddia etmesiydi. Şirketler, "gizleyecek bir şeyi olmayanın" mahremiyet ihtiyacı olmayacağını öne sürerek hem insanları daha fazla paylaşım yapmaya teşvik ettiler hem de mahremiyet ihlallerinin üzerini örtmeye çalıştılar. Üçüncü gelişme ise teknoloji şirketlerinin dijital yaşamlarımızdan kaynaklanan kişisel verilerin çok kârlı olabileceğini keşfetmeleriydi; özellikle de reklamcılıkta. Facebook'un eski çalışanlarından Jeff Hammerbacher'in belirttiği gibi bir neslin en iyi beyinleri zamanlarını reklam tıklarını artırmaya harcadılar. Bu üç gelişme birbirini besleyerek yapay öğrenmenin gelişimi ve yayılımı için elverişli bir ortam hazırladı.

Günümüzde YZ sistemleri, sosyal medya etkileşimlerinden sağlık kayıtlarına kadar geniş bir yelpazede veri topluyor. Bu veriler, bireylerin ilgi alanları, alışkanlıkları ve hatta sağlık durumları hakkında detaylı bir profil oluşturmaya olanak tanıyor. Ancak YZ algoritmaları, büyük miktarda kişisel veriyi işleyip öğrendikçe, mahremiyeti korumak ve gizliliği sağlamak daha da önem kazanıyor. Kişisel veriler, kötüye kullanıma son derece açık, güvende tutulmaları zor. İşin kötüsü, suçlulardan sigorta şirketlerine ve istihbarat kurumlarına kadar pek çok aktör kişisel verilerin peşinde. Çünkü veri söz konusu aktörlere iki önemli avantaj sağlıyor:

İnsanlar hakkında daha isabetli kestirimlerde bulunabiliyorlar ve bazen onların bile bilmedikleri şeyleri bilebiliyorlar. Örneğin sigorta şirketlerin genetik verilere ulaşabildiği durumlarda sigorta şirketi belirsizlikleri en aza indirerek önemli bir avantaj sağlayabiliyor ve sigortalananlarla sigortalayanlar arasındaki bilgi asimetrisi artıyor.

İnsanlar hakkındaki bilgileri arttıkça, insanların kararlarını etkileme olanakları da artıyor. Örneğin, bir ürünü satın alma veya bir filmi izleme olasılıklarını artırılabilir veya Facebook/Cambridge Analytica skandalında olduğu gibi seçim süreçlerine müdahale edilebilir. Sonuç olarak YZ, kişisel verilerin güvenliği açısından daha fazla risk oluşturmalarının temel nedeni, büyük miktarda veriyi işleme yeteneği, karmaşık algoritmaları ve sürekli öğrenme sürecidir.

2.6.1. Hukuksal Önlemler

YZ'nin sunduğu avantajlardan yararlanırken, bu risklerin farkında olmak ve gerekli önlemleri almak önemlidir. Bazı önlemler şunlardır:

Veri Gizliliği ile Yenilik Arasındaki Denge: YZ alanındaki hızlı gelişmeler, veri gizliliği ile çelişen durumlara neden olabilir. Bu iki değer arasında nasıl bir denge kurulacağı önemli bir sorudur.

Uluslararası İş birliği: Mahremiyet, küresel bir sorun olduğu için ülkeler arasında işbirliği gereklidir.

Düzenlemeler: Mahremiyeti koruyan ulusal ve uluslararası yasal düzenlemeler yapılmalı ve mevcut düzenlemeler güncellenmelidir.

YZ Etik Kurulları: YZ teknolojilerinin geliştirilmesi ve kullanımı sürecinde ortaya çıkabilecek etik sorunları ele almak ve çözüm önerileri sunmak üzere oluşturulan, farklı disiplinlerden uzmanların bir araya geldiği yapılar oluşturulmalıdır.

2.6.2. Teknik Çözümler

Anonimleştirme ve Pseudonymizasyon: Kişisel verilerin doğrudan tanımlanmasını engelleyen teknikler kullanılabilir.

Anonimleştirme: Kişisel verilerdeki doğrudan tanımlayıcılar (isim, TC kimlik numarası gibi) tamamen kaldırılır. Veri, hiçbir bireye doğrudan bağlanamayacak hale gelir. Ancak, bazı durumlarda ileri düzey istatistiksel analizler veya diğer verilerle birleştirildiğinde yeniden tanımlanabilirlik riski olabilir.

Pseudonymizasyon: Kişisel veriler, doğrudan tanımlayıcılar yerine benzersiz bir kod (pseudonim) ile değiştirilir. Bu kod, yalnızca belirli bir anahtar kullanılarak orijinal veriye geri çevrilebilir. Bu yöntem, anonimleştirmeye göre daha esnektir ancak anahtarın güvenliği kritik öneme sahiptir.

Şifreleme: Verilerin güvenli bir şekilde saklanması için şifreleme yöntemleri kullanılmalıdır.

Erişim Kontrolü: Verilere sadece yetkili kişilerin erişebilmesi sağlanmalıdır.

Hesap Verebilirlik Mekanizmaları: YZ sistemlerinin kararlarının nasıl alındığını anlamak ve açıklamak için teknik ve standart mekanizmalar geliştirilmelidir.

Veri Azaltma: Kişisel verilerin yalnızca belirli bir amaç için gerekli olanlarla sınırlı tutulması.

Rıza ve Şeffaflık: Bireylerin, verilerinin nasıl kullanılacağı konusunda bilgilendirilmesi ve veri toplama öncesinde açık rıza alınması.

Kullanıcı Hakları: Bireylerin kişisel verilerine erişim, düzeltme veya silme talebinde bulunabilmesi ve belirli durumlarda veri işleme işlemlerinden vazgeçme hakkına sahip olmaları

Amaç Sınırlaması: Kişisel verilerin, sadece toplandığı anda belirtilen belirli amaçlar için kullanılmasını sağlayacak kontrollerin olması. Kişisel verilerin işlendikleri amaçlarla bağlantılı olarak yalnızca yeterli, ilgili ve gerekli olanlarla sınırlı bir şekilde işlenmesi sağlanmalıdır.

Veri Güvenliği: Kişisel verilerin yetkisiz erişim, ihlal veya kötüye kullanımına karşı korunması için teknik önlemler uygulamalara eklenmelidir.

2.6.3. Özel Sorun Alanları

Biyometrik Bilgi Kullanımı

YZ sistemlerinde yüz tanıma teknolojileri gibi biyometrik bilgilerin kullanımı yaşamsal mahremiyet sorunlarına neden olabilir. Biyometrik bilgiler, doğası gereği kişisel ve çoğu durumda değiştirilemez olduğundan özellikle hassastır. Bu tür bilgilerin yetkisiz toplanması, depolanması veya kullanılması, ciddi bir ihlaldir. Kötü kullanım potansiyeli vardır.

Çocukların Kişisel Verilerinin Gizliliklerinin Korunması

Çocuklar, özellikle YZ söz konusu olduğunda, gizlilik risklerine karşı daha savunmasızdır. Bunun nedeni, çocukların genellikle çevrimiçi kişisel bilgi paylaşmanın risklerinin daha az farkında olması ve yetişkinlere ve teknolojiye daha çok güvenmeleri olabilir. Ayrıca, çocukların beyinleri gelişmekte olduğu için, bu onları manipülasyon ve sömürüye karşı daha savunmasız hale getirebilir.

YZ sistemleri, çocukların konumu, çevrimiçi etkinlikleri ve hatta biyometrik verileri de dahil olmak üzere, çocuklar hakkında geniş bir yelpazede kişisel bilgi toplayabilir ve onları takip edebilir. Örneğin eğitimin kişiselleştirilmesi amacıyla veya bireysel sağlık hizmetlerinde toplanan veriler daha sonra çocuklara reklam göstermek veya toplanan verileri üçüncü taraflara satmak için, hatta bundan çok daha kötü amaçlar için kullanılabilir.

Bu riskleri ele almak için, çocukların verilerini koruyan güçlü gizlilik yasalarına ve düzenlemelerine sahip olmak önemlidir. Ayrıca, çocukları çevrimiçi kişisel bilgi paylaşmanın riskleri konusunda eğitmek ve gizliliklerini korumak için ihtiyaç duydukları becerileri geliştirmelerine yardımcı olmak da önemlidir. Bu noktada ailelerin ve öğretmenlerin de özel eğitimi gereklidir. Çocukların aileleri ve okul tarafından çevrimiçi etkinliğini izleyecek bilgi ve donanımları edinmeleri sağlanmalı ve gizliliğin önemi konusunda bilinçlendirilmelidir.

2.7. HESAP VEREBİLİRLİK

Algoritmalar ve onları yönlendiren veriler insanlar tarafından tasarlanır ve oluşturulur. Sonuçlar; insanların tasarım hakkındaki kararları, neyin optimize edileceğiyle ilgili belirledikleri kurallar ve hangi eğitim verilerinin kullanılacağı hakkındaki seçimleriyle şekillenir.

Yaşam döngüleri boyunca YZ'lerin düzgün çalışmasının sağlanması için tasarımcıların, geliştiricilerin ve bu sistemleri kuranların uyacağı standartlar ve mevzuat geliştirilmeli; sorumluluklar açık seçik olarak belirlenmeli; kişiler veya kurumlar, gerçekleştirdikleri eylemlerden sorumlu tutulmalıdır.

YZ sistemlerinin insan yaşamı üzerindeki etkileri her geçen gün daha da artmaktadır. Bu sistemler, iş başvuruları, performans değerlendirmeleri, yargı süreçleri, kredi notu belirleme, hastalık teşhisi gibi çok çeşitli alanlarda kullanılmakta; doğrudan veya dolaylı olarak karar alma süreçlerinde yer almaktadır. Dolayısıyla YZ sistemlerinin şeffaf, etik ve adil şekilde kullanıldığından emin olunması büyük önem taşımaktadır.

Çünkü işler her zaman yolunda gitmeyebilir. Tüm şartları sağlamanıza karşın iş başvurunuz gerekçesiz bir şekilde reddedilebilir; para transferleriniz şüpheli işlemler kategorisine girebilir ve hesabınız bloke edilebilir; yanlış teşhis ve onu takip eden yanlış tedavi nedeniyle sağlığınıza kaybedebilirsiniz. Bu gibi durumlar yaşandığında ne olacaktır? Ya da bu gibi sorunların önüne nasıl geçilebilir? Kimler, neye göre sorumluluk alacaktır?

YZ sistemleri kusursuz değildir. Hata yapabilir ve verdikleri kararlarla insanları mağdur edebilir. Algoritmalar ve onları yönlendiren veriler insanlar tarafından tasarlanır ve oluşturulur. Sonuçlar; insanların tasarım hakkındaki kararları, neyin optimize edileceğiyle ilgili belirledikleri kurallar ve hangi eğitim verilerinin kullanılacağı hakkındaki seçimleriyle şekillenir.

YZ'de hesap verebilirlik (*accountability*), bu gibi sorunlar bağlamında ve farklı öznelerin sorumluluklarını tanımlayabilmek amacıyla gündeme gelmiştir. YZ sistemlerinin tasarımı, geliştirilmesi, dağıtılması ve kullanımından sorumlu olan kişilerin veya kurumların bu süreçler sırasında yaptıkları eylemlerden sorumlu tutulabilmesini ifade eder. Hesap verebilirliği belirli bir cevap verebilirlik ilişkisi olarak tanımlayabiliriz. Bu ilişkinin gerçekleşebilmesi için birbiriyle iç içe geçen üç temel koşul vardır: otoritenin tanınması, sorgulama ve gücün sınırlandırılması.

2.7.1. Hesap Verebilirlik İçin Gerekli Koşullar

Otoritenin Tanınması

Görevlerin devredilmesiyle ilgili ve karşılıklı bir ilişkidir: A (örneğin vatandaş) X'e (örneğin kamu görevlisi) kendi çıkarlarına hizmet etmesi için yetki verdiğinde X de A'nın bu çıkarlara gerektiği gibi hizmet edilip edilmediği konusunda kendisini sorumlu tutacağını kabul eder. Bu ilişki gerçekleşmeden, hesap verebilirlik sadece "lütuf ve iyilikle raporlama veya bilgilendirme" halini alır.

Sorgulama

A'nın X'in hareketlerini incelediği gerçeğini ifade eder. Böylece 'hesap verebilirlik' ile 'ahlaki sorumluluk' arasındaki geleneksel ayırım ortadan kalkar. Hesap verebilirlik dışsal bir bakış açısından incelemeyi varsayar. Ahlaki sorumluluk ise içsel bir bakış açısına atıfta bulunur. Yalnızca kendine karşı sorumlu olmak, hiç sorumlu olmamak anlamına gelir.

Gücün Sınırlandırılması

A'nın X'in performansını izlemesine ve sonuçlarını değerlendirmesine olanak tanıyarak devredilen yetkinin keyfi kullanımını kısıtlar. Demokratik toplumlarda, devredilen yetkilerin kısıtlanabilmesi, bu yetkilerin meşruiyeti için bir ön koşuldur.

2.7.2. Hesap Verebilirliğin Başlıca Özellikleri

YZ'de hesap verebilirlik hedefi; tasarımcıların, geliştiricilerin ve YZ sistemlerini kuranların, yaşam döngüleri boyunca YZ'lerin düzgün çalışmasını sağlaması için standartlara ve mevzuata uymaları beklentisiyle ilgilidir. Ancak YZ sistemleri, geleneksel teknolojilerden birçok açıdan farklıdır. Örneğin derin öğrenme sistemlerinin opaklığı ve öngörülemezliği, istenmeyen sonuçların nedenlerinin tespit edilmesini zorlaştırır. Ayrıca çeşitli etkenler, YZ'lerin yanlış (veya tartışmalı) işler yapmasına yol açabilmektedir. Örneğin YZ sistemlerinin neden olduğu (ya da artırdığı) ayrımcılık ön yargılı eğitim verilerinden, sistem hatalarından, programcı hatalarından, yanlış kullanımlardan vs. kaynaklanabilir veya bazen bu etkenlerin bir kombinasyonu olabilir. Birden fazla aktör ve kaynak içeren sistemlerde bireysel sorumlulukları tam olarak belirlemek zor bir iştir. Ayrıca hiç kimsenin olumsuz sonuçları önlemek zorunda hissetmeyeceği bir durum ortaya çıkabilir. Tam tersi bir durum ise her biri farklı beklenti ve değerlendirme kriterlerine sahip tarafların çokluğunda ortaya çıkar. Bu sefer de taraflar arasında bir kafa karışıklığı ortaya çıkar.

Kısaca kimsenin sorumlu tutulmadığı hesap verebilirlik boşlukları ile prosedürlerin verimsiz bir şekilde biriktirildiği hesap verebilirlik fazlalıklarına karşı YZ'de hesap verebilirliği tartışırken aşağıdaki konuların netleştirilmesi gerekir:

❖ **Hesap verebilirlik hangi bağlam için tartışılmaktadır? (Niçin?)**

Bu soruya alan (finans, sağlık, adalet, kamu yönetimi vs.), işlev (doğal dil işleme, filtreleme, sınıflandırma, robot kontrolü, bilgisayarla görü vb.) ve özerklik düzeyi (elle kontrol, eylem desteği, paylaşımlı kontrol, tam otomasyon vb.) bağlamında yanıt verilebilir.

❖ **Hesap verebilirlik ne hakkındadır? (Ne?)**

YZ'nin yaşam döngüsü içinde üç görev kümesi etrafında bir tanımlama yapılabilir: Tasarım (planlama, odaklanılan hedef kitle, veri stratejisi, geliştirme stratejisi vb.), geliştirme (modelin eğitimi, kodlama, test, güvenlik mekanizmaları vb.), kurulum (izleme, işletme, kullanım).

❖ **Yukarıda belirtilen görevler kimler tarafından yerine getirilmektedir? (Kim?)**

Bu soruya bireyler (YZ geliştiricileri, YZ tasarımcıları, uzmanlar vb.), hiyerarşideki rolleri (patent sahipleri, yöneticiler, denetçiler, test uzmanları), kurumsal durumları (politika yapıcılar, geliştirme şirketleri, veri denetleyicileri ve hissedarlar) göz önünde bulundurularak yanıt verilebilir.

❖ **Hesap verebilirlik kime karşı yerine getirilmektedir? (Kime?)**

Örneğin, verileri YZ'leri eğitmek için kullanılan özneler; hissedarlar, alan uygulayıcıları (klinik uygulamalarından sorumlu uzmanlar).

❖ **Hesap verebilirlik hangi standarda göre tanımlanmaktadır? (Neye göre?)**

Bu standart yasal (haksız fiiller, suçlar, adil olmayan ticari uygulamalar, mahremiyet vb.), etik (adalet, şeffaflık, özerklik, kapsayıcılık, kırılabilirlik, güvenilirlik, sürdürülebilirlik vb.), teknik (sağlamlık, uyarlanabilirlik, doğruluk, verimlilik, bakımı yapılabilirlik, güvenlik vb.) olabilir.

❖ **Hesap verebilirlik süreçleri nasıldır? (Nasıl?)**

İçsel (dahili denetimler, simülasyonlar, öz değerlendirmeler, piyasa sonrası izleme vb.), insan-makine etkileşimi sonucu (geribildirim döngüleri, denetleyici kontroller ve etkileşimli yapay öğrenme), dışsal (sistem doğrulamaları, dış denetimler, dış uygunluk ve etki değerlendirmeleri, kamu denetimi ve adli incelemeler) olabilir.

❖ **Sonuçları ne olacak?**

Bir kararla (öneri, onay, ret, yasak vb.) sonuçlanabilir; yasalar çerçevesinde bir durumda şirket için itibar veya pazar kaybı olabilir; yasa dışı bir durumdaysa tazminat, tutuklama, para cezası, disiplin tedbiri, uzaklaştırma, revizyon talebi ve iptal gibi durumlar olabilir.

2.7.3. YZ'de Hesap Verebilirliğin Hedefleri

Hesap verilebilirliğin dört temel hedefi vardır.

- ❖ **Uyumluluk:** YZ'nin etik ve yasal standartlara uyumu sağlanmalıdır. Bu hedef, bir yapay zekânın tüm yaşam döngüsü boyunca karşılanması gereken tasarım, geliştirme ve dağıtım standartlarını tanımlar, ancak özenli biçimde hazırlanmazsa oldukça genel geçer bir anlam taşır.
- ❖ **Raporlama:** Hesap verebilirliğin diyalog boyutunu temsil eder. YZ'lerin davranışlarının açıklanmasını ve gerekçelendirilmesini sağlayan uygulamalardır.
- ❖ **Gözetim:** YZ'lerin yaşam döngüsü performansını değerlendirmek için ilgili gerçekleri veya bilgileri bulmayı ve kanıt oluşturmayı amaçlar. Organizasyonun içinden veya dışından farklı organlar tarafından veya insan-makine arayüzleri aracılığıyla gerçekleştirilebilir.
- ❖ **Yaptırım:** YZ'lerin performansının izlenmesini ve değerlendirilmesini resmi veya gayri resmî sonuçlara bağlar. Bu aynı zamanda istenmeyen davranışları caydırma amacını da taşır.

2.7.4. Ön alıcı ve Tepkisel Hesap Verebilirlik

Ön alıcı hesap verebilirlik, amaçlanan sonuçların üretilmesine ve amaçlanmayan sonuçların önlenmesine veya azaltılmasına odaklanır. Suistimal ve hatalara neden olan kurumsal hususları bunlar meydana gelmeden önce tespit etmeyi ve düzeltmeyi amaçlar. Ön alıcı hesap verebilirlikte, hedeflerin, rollerin, sorumlulukların ve komuta hatlarının doğru ve açık seçik bir şekilde belirlenmesi gerekir. Sistemik çarpıklıkları ön alıcı bir duruşla düzeltmek için, standart belirlemeye (uyum) ve YZ'lerin uygunluğu ve etkisine ilişkin ön kontrollere (gözetim) daha fazla önem verilir.

Tepkisel hesap verebilirlik, olay gerçekleşikten sonra gerekli adımları tetikler. İstenmeyen bir durum olduğunda, başarısızlıkların etkilerini telafi etmeyi amaçlar. Bu görüşe göre, hesap verebilirlik bir erdemden çok bir mekanizmadır. Olaydan sonra, sunulacak açıklama ve gerekçelere (raporlama) ve hesap verebilirlik değerlendirmesinin sonuçlarına (yaptırım) önem verilir. Gözetim, burada bir ön kontrolden çok geriye dönük bir inceleme (örneğin, yargı denetimi) olarak karşımıza çıkar.

Ön alıcı ve tepkisel hesap verebilirlik genellikle bir arada bulunsa da zaman zaman biri daha ön planda olabilmektedir. Ancak her iki hesap verebilirlik mekanizması da YZ'lerin yaşam döngüsü boyunca gözetim ve şeffaflık mekanizmaları sağlar ve böylece YZ'ye olan (kamusal) güvenin artırılmasında merkezi bir rol oynarlar.

2.8. ZARAR VERMEMEK

YZ sistemleri, kullanım öncesinde veya sonrasında kamu çıkarlarını hedefleyen, bağımsız gözetim kuruluşları tarafında izlenmeli; olumsuzluklara karşı ulusal veya uluslararası kuruluşlar tarafından (gerekli durumlarda söz konusu sistem veya ürünün tamamen yasaklanmasını içeren) yaptırımlar uygulanmalıdır.

Bireysel, toplumsal ve çevresel zararların araştırılması, tespiti ve çözümünü hakkındaki sorumluluk ulusal ve uluslararası düzenleyici kurumlara yüklenmemeli, teknoloji şirketlerine zarar vermediklerini kanıtlama yükümlülüğü getirilmelidir.

YZ sistemlerinin gündelik hayatta kullanımının öneminin artmasıyla birlikte, bu sistemlerden kaynaklı sorunlar daha görünür oldu. Bu nedenle, YZ sistemlerinin bireylere, toplumlara ve çevreye zarar vermeyecek şekilde tasarlanması ve kullanılmasının üzerinde durulmaya başlandı.

Zarar Vermeme ilkesinin hayata geçirilebilmesi için birinci olarak, YZ sistemlerinden kaynaklı doğrudan zararların önüne geçilmesi gerekmektedir. YZ sistemleri, insanlara fiziksel, duygusal veya psikolojik zarar verebilecek şekilde tasarlanmamalı veya kullanılmamalıdır. Otonom araçlar veya robotik sistemler, yaralanma veya ölüme yol açabilecek kazaları veya hataları en aza indirerek, kullanıcıların ve çevredekilerin güvenliğine öncelik vermelidir. Ancak güvenlik, sadece tasarım ve geliştirme aşamasıyla ilgili değildir. Şirketlerin ürünleri hakkındaki abartılı iddiaları, kamuoyunun ilgili sistem veya ürünler hakkında yanlış bir kanıya kapılmasına neden olabilir. Örneğin, Tesla ve otonom sürüş teknolojisi hakkındaki abartılı reklam, haber ve bizzat Musk'ın yazdıkları; tüketicileri aldatmış ve kazalara neden olmuştur. Tesla araçlarını satın alanlar; Musk'ın abartılı iddialarına nedeniyle, kullanım kılavuzunda öyle yazmamasına karşı aracın kontrolünü tamamen Tesla'ya bırakmışlardır.^{[23],[24]}

İkincisi, YZ sistemlerinin mevcut sistemler ve sosyal bağlamlarla nasıl etkileşime girebileceği mutlaka göz önünde bulundurulmalıdır. Belirli bir alandaki etkileyici sonuçların farklı alanlarda istenmeyen sonuçlara neden olabileceği unutulmamalıdır.

Üçüncüsü, verilerin korunması ve mahremiyettir. Mevcut YZ yaklaşımlarının temelinde veri vardır. Verinin toplanması, eğitim amacıyla kullanımı, başka veri setleri ile birleştirilmesi, paylaşılması vb. hassas süreçlerdir. Bunun yanında YZ sistemleri, insanların itibarına, finansal güvenliğine veya kişisel güvenliğine zarar verebilecek yetkisiz erişim, sızıntı veya kötüye kullanıma karşı sıkı önlemler almalı, kişisel ve hassas verileri sorumlu bir şekilde ele almalıdır.

Dördüncüsü, sosyal medya algoritmaları, reklam sistemleri, sohbet robotları gibi YZ sistemlerinin kullanıcıların kararlarını etkilemeyi hedefleyebildiğinin unutulmaması gerektiğidir. Bu sistemler, çoğu zaman kullanıcıların zaaflarından yararlanırlar. YZ sistemleri, yanlış bilginin yayılmasını önlemek veya bireylerin zihinsel refahına zarar verebilecek bağımlılık davranışlarını teşvik etmeyecek biçimde tasarlanmalıdır.

Beşincisi, YZ sistemlerinin şeffaf olması gerekliliği, böylece kullanıcıların söz konusu sistemlerin işleyişini ve potansiyel risklerini anlayabilmeleridir. Bu sayede, hesap verebilirlik için gerekli koşullar sağlanabilir ve zararlı etkiler, hızlı bir şekilde tespit edilip ele alınabilir.

Altıncısı, mevcut YZ sistemlerinin çıktılarının nedensel ifadeler içerse de, öğrenme mekanizmalarının veri içindeki korelasyonlara dayanması ve nedensellik içermemesidir. Nedensel çıktılar, özellikle büyük dil modellerini daha ikna edici kılmakta ve kararlarına olan kullanıcı güvenini artırmaktadır. Kullanıcıların mağduriyet yaşamaması için, bu sistemlerin halihazırda sadece verilerin istatistiksel ilişkileri üzerinden eğitildiği konusunda kullanıcılar bilgilendirilmeli ve aldıkları kararların sorgulanmadan kabul edilmesinin önüne geçilmelidir.

YZ sistemleri, kullanım öncesinde veya sonrasında kamu çıkarlarını hedefleyen, bağımsız gözetim kuruluşları tarafında izlenmeli; olumsuzluklara karşı ulusal veya uluslararası kuruluşlar tarafından (gerekli durumlarda söz konusu sistem veya ürünün tamamen yasaklanmasını içeren) yaptırımlar uygulanmalıdır. Musk, 2022'deki bir konuşmasında ABD ve Avrupa arasındaki farkı şöyle açıklıyordu: “ABD’de bazı şeyler varsayılan olarak yasaldır, Avrupa’da ise varsayılan olarak yasa dışıdır.” Ama tam da bu nedenle, düzenlemenin daha sıkı olduğu AB’de, ABD’deki gibi sorunlar yaşanmadı. AB, otonom araçları en başta sıkı tuttu ve Tesla’nın otomatik pilotlu araçlarının Avrupa’daki satışına hemen izin vermedi.^[25]

Ancak bireysel, toplumsal ve çevresel zararların araştırılması, tespiti ve çözümü hakkındaki sorumluluk ulusal ve uluslararası düzenleyici kurumlara yüklenmemeli; teknoloji şirketlerine zarar vermediklerini kanıtlama yükümlülüğü getirilmelidir.

Araştırmacı gazetecilik ve bağımsız araştırmalar, teknolojinin hesap verebilirliği açısından kritik öneme sahiptir. Şeffaf olmayan sistemler hakkında yapılan araştırmalar, reel ve potansiyel zararlara ilişkin kanıtlar ortaya koyar, çok önemli riskleri ve sorunları görünür hale getirir. Fakat, bu tarz denetimler ve benzeri hesap verebilirlik çerçeveleri, zararların belirlenmesi ve ele alınması sorumluluğunu şirketin dışında konumlandırarak doğrudan teknoloji şirketlerinin avantajına bir durum yaratabilir ve yeterli değildir.

Ekonomik süreçlerde, iletişimde ve iş gücü piyasasında yaygın ve öngörülemez etkilere sahip olan teknolojik sistemlerin açıklarının ve olası zararlarının, sadece araştırmacı gazetecilik ve bağımsız araştırmalar yoluyla bulunması, sonrasında bu teknolojiyi üreten şirketlerin ‘pardon’ diyerek sorumluluktan sıyrılabileceği kısır bir döngü yaratmaktadır.

Olası zararların önüne geçebilmek için (gıda ve ilaç piyasasının düzenlenmesinde olduğu gibi) teknoloji üreten şirketler, yasalara uyduklarını göstermekle yükümlü olmalıdır. Bunun için, her durumda kullanımının yasaklanması gereken türler ve uygulama alanlarını belirleyen kesin ayrımlara ve yasak bölgeler gibi yapısal sınırlamalara ihtiyaç vardır. Örneğin, birçok şehirde yüz tanımaya ilişkin yasaklar getirilmiştir.

Yüz tanıma gibi duygu tanıma teknolojilerinin de hiçbir zaman ama özellikle de iş yerlerinde kullanılmaması gerekir. Çünkü, bilimselliğinin tartışmalı olması bir yana, bu teknolojilere eşlik eden ayrımcı etkilerle ilgili endişeler de henüz giderilememiştir.

2.9. ÖZERKLİK

YZ, çalışanların yerini almayı hedefleyen tarzda değil, çalışanlara destek olacak ve onları güçlendirebilecek şekilde geliştirilmeli ve uygulanmalıdır.

Üretken YZ ile beraber (GPT) YZ uygulamaları sanat (!) eserleri üretmeye başladılar. Harika resimler, verilen herhangi bir konu hakkında yazılan şiir ve romanlar kamuoyunu etkiledi. Ancak fantezi ve bilim kurgu yazarı Joanna Maciejewska'nın bir Twitter mesajında belirttiği gibi YZ yanlış yönde ilerliyor.^[26]

“Sanat ve yazı yazabilmem için yapay zekânın çamaşırlarımı ve bulaşıklarımı yıkamasını istiyorum, ben çamaşırlarımı ve bulaşıklarımı yıkayabileyim diye yapay zekânın sanatımı ve yazımı yapmasını değil.”



Joanna Maciejewska

@AuthorJMac

You know what the biggest problem with pushing all-things-AI is? Wrong direction.

I want AI to do my laundry and dishes so that I can do art and writing, not for AI to do my art and writing so that I can do my laundry and dishes.

Teknoloji şirketleri de dahil olmak üzere birçok aktör YZ'nin hedefinin insanların yerini almak değil, insanları güçlendirmek olması gerektiğini savunuyor. Scneiderman'ın İnsan Merkezli YZ adlı çalışmasında savunduğu gibi insan yeteneklerini güçlendirmek değerli bir hedeftir.^[27]

Teleskoplar ve mikroskoplar, insan yeteneklerini artıran insan gözünün uzantılarıdır. Hesap makineleri, dijital kütüphaneler ve e-posta, kullanıcıların hiçbir insanın yardım almadan yapamayacağı şeyleri yapmasını sağlar. İnsanları güçlendiren daha güçlü artırma ve güçlendirme araçlarına ihtiyacımız var. Bu yaklaşımlardan biri, sanatçılara, müzisyenlere, şairlere, oyun yazarlarına, fotoğrafçılara ve kameramanlara alternatifleri keşfetmeleri ve yaratıcı bir şekilde yeni, ilginç ve anlamlı bir şeyler üretmeleri için daha fazla esneklik sağlayan yaratıcılık destek araçlarının geliştirilmesidir. Kameralar ve müzik aletleri insanların yapabileceklerinin sınırlarını genişletmiştir, ancak insan hala yaratıcı kaynağın itici gücüdür. Yeni cihazların bu geleneği devam ettirmesi muhtemeldir.

Fakat, özellikle GPT teknolojisiyle beraber tam tersi yönde ilerlediğimiz görülüyor. Daha birkaç yıl önce, YZ tarafından üretilen görüntü ve müzik için fikri mülkiyet hakkı tartışması gündemdeyken şimdi sanatçılar ürünlerinin kendilerinden izinsiz olarak YZ sistemlerinin eğitiminde kullanıldığını ve haklarının ihlal edildiğini savunuyorlar.

Bunun yanında sık sık dile getirilen insan performansını yükselten, artıran, güçlendiren ve geliştiren ürün ve hizmetler yaratma hedefi gerçek uygulamalarla çalışmaktadır. Bazı görevleri insanlardan daha iyi yapan makineler veya kendi başına çalışan akıllı otonom sistemlerin yer aldığı YZ atılımları coşkuyla karşılanmaktadır. Aslında robotların insanların işini ellerinden almasından insan türüne son vermesine kadar uzanan felaket senaryoları YZ'nin insanı güçlendiren değil onun yerini alan bir vizyon etrafında şekillendirilmesinin bir sonucudur.

Bu bağlamda, YZ'nin günümüzde izlediği yolu gücü elinde tutan büyük teknoloji şirketlerinin vizyonundan bağımsız ele alamayız. Çalışanların yeteneklerini tamamlayıcı niteliğe sahip teknolojiler yerine onların yerini almayı hedefleyen teknolojilere yönelmek bir zorunluluk değil, politik bir tercihtir. Son yıllarda teknolojik uygulamaların eşitsizliği artıran bir rol üstlenmiş olması daha çok büyük şirketlerin ve iktidar sahiplerinin bir tercihidir. Şu anki eğilim tamamen işsiz bir gelecekte çok insanların daha düşük ücretlerle çalıştığı ve yoksullaştığı bir gelecek yönündedir.

Ancak farklı bir vizyonla teknoloji çalışanların mevcut işlerindeki üretkenliğini artırmak, yeni işler yaratmak veya insanların karar verme süreçlerini iyileştirmek için de kullanılabilir. Bu da ancak insanoğlunu tamamlayan ve yetkinleştiren teknolojilere daha çok eğilmeyi teşvik eden teknoloji politikalarıyla gerçekleştirilebilecek bir hedeftir.

Ayrıca YZ araçlarının insanın yerini alma hedefiyle geliştirilmesi, sunulması ve haklarındaki abartılı iddialar insanların yeteneklerini güçlendirmek bir yana tam tersi sonuçlar ortaya çıkabilmektedir. Belirsizliği azaltmak ve kararları basitleştirmek için tasarlanan YZ araçları, bilişsel bir kısayol olan otomasyon ön yargısına neden olabilmektedir. Çünkü YZ sistemleri çalışırken insanlar dikkatlerini ve gözetimlerini azaltma eğilimindedir.

İnsanlar YZ sistemlerin insanlardan daha doğru, tutarlı ve güvenilir olduğunu varsayabilmekte ve bu önerilere gereğinden fazla itibar edebilmektedir. Ancak YZ kusursuz değildir. Irk, cinsiyet vb. ön yargılar içerir. Bu sistemlerin kararlarına karşı gözü kapalı güven ayrımcı, yanlış ve potansiyel olarak zararlı kararlara neden olabildiği gibi bir süre sonra insanların karar verme yetilerini de zayıflatır. İnsanlar, kendileri zaman zaman yanılabilirler bile bilgisayarların hep doğruyu söylediği düşüncesine kapılabilirler.

Örneğin radyologlar arasında yapılan bir çalışmanın gösterdiği gibi YZ hakkındaki abartılı iddialar insanların kendi kararlarına kuşkuyla yaklaşmasına neden olabilmektedir. Söz konusu çalışmada araştırmacılar hangi mamogramlarda kanser olduğunu önceden bilmektedir ve YZ'yi taramaların bir alt kümesi için yanlış cevap verecek şekilde ayarlarlar. YZ yanlış bir sonuç verdiğinde, deneyimsiz ve orta derecede deneyimli radyologların kanser tespit doğruluklarının yaklaşık %80'den yaklaşık %22'ye düştüğü görülmüştür. Çok deneyimli radyologların doğruluk oranı ise yaklaşık %80'den %45'e düşmüştür.^[28]

Yurttaşların YZ uygulamalarının her zaman gerçekten çok doğru ve tutarlı yanıtlar üretmediğinin, çeşitli ön yargılar içerebildiğinin ve kimi zaman şirketlerin tam anlamıyla test edilmemiş teknolojileri piyasaya sürdüğünün farkında olabilmesi için YZ okuryazarlıkları güçlendirilmelidir.

2.10. SONUÇLAR

Mary Shelley'in 19. yüzyılda yayımlanan Frankenstein adlı romanı, bilim ve teknolojinin büyüüne kapılan insanlık için önemli bir uyarı niteliğindedir. Yaşamın sınırlarını çözmek isteyen genç ve hırslı bilim insanı Victor Frankenstein, ceset parçalarını birleştirip bilimsel yöntemlerle yarattığı yapay varlık yüzünden sevdiklerini kaybeder ve hayatı cehenneme döner. Bilim kurgu yazarı Isaac Asimov, insanların yarattığı teknolojilerle yaşadığı çelişkilere bir çözüm olarak, robotlarla insanlar arasındaki ilişkiyi düzenleyen Üç Robot Yasası'nı öne sürmüştür. Asimov'un eserlerindeki robotlar, bir insana zarar vermemek ya da hareketsiz kalarak bir insanın zarar görmesine seyirci kalmamak üzere programlanmıştır. YZ etiğinin birinci dalgasındaki tartışmalar, robotların dünyayı ele geçirmesi gibi spekülâtif konulara odaklanırken, Asimov'un üç robot yasaı gibi YZ'ye entegre edilebilecek kuralların peşine düşmüştür.

YZ tartışmalarının ikinci dalgası, birinci dalğanın spekülâtif yaklaşımından farklı olarak, somut ve uygulamalı sorunları merkezine alır. Tartışmalar, özellikle veri odaklı YZ sistemlerinden kaynaklanan algoritmik önyargı, veri gizliliği ihlalleri, açıklanabilirlik gibi konulara kayar. İkinci dalgada, daha çok teknik çözümlerin uygulanabileceği sorunlar öne çıkar. Örneğin, algoritmik önyargıların giderilmesi için daha çeşitli ve dengeli veri kümeleri oluşturulmasına odaklanılırken, veri gizliliğini korumak adına anonimleştirme gibi teknikler devreye sokulmaktadır. Öte yandan, YZ sistemlerinin kararlarını daha anlaşılır kılmak amacıyla, açıklanabilirlik alanında yöntemsel geliştirmeler ve yeni araçlar üzerine yoğun araştırmalar yürütölmektedir.

Üçüncü dalga YZ etiğı, ikinci dalgada öne çıkan teknik olarak ayrıştırılmış ve yönetilebilir sorunlara odaklanan yaklaşımlardan farklı olarak, bu meseleleri daha bütöncöl ve toplumsal bir çerçevede ele almayı hedefler. Bu yeniden çerçeveleme, etik tartışmalarını yalnızca şirketlerin etik kurullarıyla sınırlı teknik çözümlerden çıkararak, daha geniş bir politik ve kamusal alana taşır. Böylece, YZ sistemlerinden doğrudan etkilenen bireylerin özneleşmesini mümkün kılarak, etik meseleleri aynı zamanda bir politik mücadele alanı hâline getirir.

Etik ilkelerin kendi başlarına bir yaptırımı yoktur. Fakat etik ilkeler, önümüzde duran somut sorunlar karşısında ve kamusal düzenlemelerde hangi değerleri öne çıkaracağımız hakkında yol gösterici olacaktır. YZ okur yazarlığı yüksek kamuoyunun düzenleyici politikalarda ısrarcı olması, teknoloji şirketlerini tamamen durduramasa da yıkıcılıklarını sınırlandırabilir. Teknolojinin toplum için kullanımını tartışmaya açabilir.

KAYNAKÇA

[1] <https://futureoflife.org/open-letter/pause-giant-ai-experiments>, son erişim 29.12.2025

[2] <https://safe.ai/work/statement-on-ai-risk>, son erişim 29.12.2025

[3] Van Wynsberghe, A. (2021). Sustainable AI: AI for sustainability and the sustainability of AI. *AI and Ethics*, 1(3), 213-218.

[4] Dignum, V. (2019). *Responsible artificial intelligence: how to develop and use AI in a responsible way* (Vol. 2156). Cham: Springer.

[5] AWS'nin bazı müşterileri için bkz. <https://spacelift.io/blog/who-is-using-aws>

[6] Luitse, D. (2024). Platform power in AI: The evolution of cloud infrastructures in the political economy of artificial intelligence. *Internet Policy Review*, 13(2), 1-44.

[7] van der Vlist, F., Helmond, A., & Ferrari, F. (2024). Big AI: Cloud infrastructure dependence and the industrialisation of artificial intelligence. *Big Data & Society*, 11(1), 20539517241232630.

[8] Hu, W., & Singh, R. (2024). Enrolling Citizens: A Primer on Archetypes of Democratic Engagement with AI. Available at SSRN 4863148.

[9] Xu, Feiyu, ve diğerleri. "Explainable AI: A brief survey on history, research areas, approaches and challenges."

[10] Gunning, David, ve diğerleri. "DARPA's explainable AI (XAI) program: A retrospective."

[11] Nannini, Luca, Agathe Balayn, ve Adam Leon Smith. "Explainability in AI policies: a critical review"

[12] European Commission, "Proposal for a Regulation on Artificial Intelligence," ec.europa.eu

[13] Artificial Intelligence Act Initiative, "AI Act Explained," artificialintelligenceact.eu

[14] <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/>

- [15] Speith, Timo. "A review of taxonomies of explainable artificial intelligence (XAI) methods."
- [16] <https://www.theguardian.com/technology/2024/sep/15/data-center-gas-emissions-tech>, son erişim 29.12.2025
- [17] https://hbr.org/2024/07/the-uneven-distribution-of-ais-environmental-impacts?tpcc=orgsocial_edit&s=31, son erişim 29.12.2025
- [18] <https://www.context.news/ai/video/forget-jobs-ai-is-coming-for-your-water>, son erişim 29.12.2025
- [19] Bennett, C. J. (2010). The privacy advocates: Resisting the spread of surveillance. MIT Press.
- [20] Fried, C. (1968). "Privacy." Yale Law Journal 77, no. 3, s. 475–493
- [21] Véliz, C. (2021). Privacy is power. Brooklyn: Melville House.
- [22] Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. Profile Books.
- [23] <https://www.nytimes.com/2023/01/17/magazine/tesla-autopilot-self-driving-elon-musk.html>, son erişim 29.12.2025
- [24] <https://www.trthaber.com/haber/guncel/tesla-otopilot-arizasi-nedeniyle-iki-milyon-araci-geri-cagiriyor-820554.html>, son erişim 29.12.2025
- [25] <https://slate.com/technology/2023/02/tesla-recall-full-self-driving-nhtsa-musk-regulation.html>, son erişim 29.12.2025
- [26] <https://x.com/AuthorJMac/status/1773679197631701238>, son erişim 29.12.2025
- [27] Shneiderman, B. (2022). Human-centered AI. Oxford University Press.
- [28] <https://www.brainfacts.org/neuroscience-in-society/tech-and-the-brain/2024/why-ai-can-push-you-to-make-the-wrong-decision-at-work-090324>, son erişim 29.12.2025