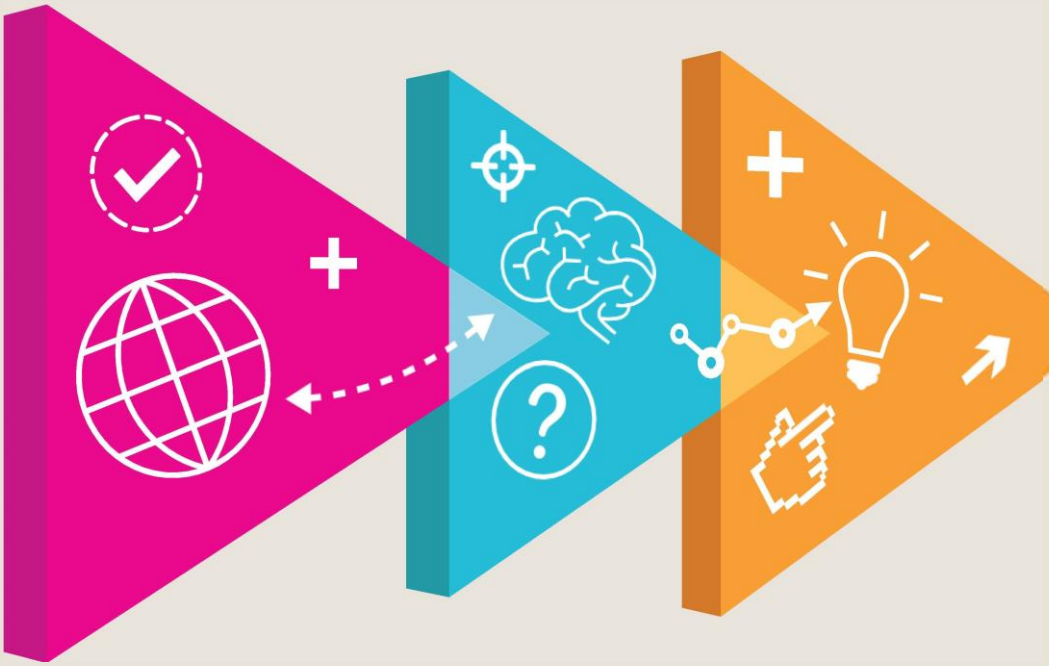
ITGT Consultancy

Yapay Zeka (AI) Denetim Çerçevesi Kılavuzu

İstişare için taslak kılavuz

Web linki için tıklayınız



ico.
Information Commissioner's Office

Çeviren: Gülsün YILDIZ

İçindekiler

20200214
Version 1.0

Kılavuz hakkında

Genel bir bakış

Yapay zeka (AI) uygulamaları hayatımızın birçok alanına girerek artan bir hızda yayılıyor. Yapay zekanın getirebileceği belirgin faydaları anlamanın yanı sıra kişilerin hak ve özgürlükleri için oluşturabileceği riskleri de anlıyoruz.

Bu nedenle, veri koruma uyumluluğu için en iyi uygulamalara odaklanarak yapay zekayı denetlemeye yönelik bir çerçeve geliştirdik – ister kendi AI sisteminizi tasarlayın, ister üçüncü parti bir sistem kullanın. Bu çerçeve AI uygulamalarını denetlemek ve kişisel verilerin adil bir şekilde işlenmesini sağlamak için sağlam bir metodoloji sağlar. Şunları içerir:

- Denetim ve incelemelerde kullanacağımız denetim araçları ve prosedürlerini içerir.
- Bu yapay zeka ve veri koruma hakkındaki ayrıntılı kılavuz, kişisel verileri işlemek için yapay zeka kullandığınızda uygulayabileceğiniz belirleyici risk ve kontrol ölçümlerini içerir.

Bu kılavuz, iki hedef kitleye yöneliktir:

- Veri koruma görevlileri (DPO-Data Protection Officers), genel danışmanlık, risk yöneticileri ve ICO'nun kendi denetçileri gibi uyumluluk odaklı kişiler.
- Makine öğrenimi uzmanları, veri bilimcileri, siber güvenlik yöneticileri, BT risk yöneticileri, yazılım geliştiricileri ve mühendisleri gibi teknoloji uzmanları.

Bu kılavuz, yapay zekanın oluşturabileceği hak ve özgürlüklere ilişkin riskleri ve bunları azaltmak için uygulayabileceğiniz uygun önlemleri nasıl değerlendirebileceğinizi açıklıyor.

Veri koruma ve "AI etiği" örtüşse de bu kılavuz, AI kullanımınız için genel etik veya tasarım ilkeleri sağlamaz. Farklı veri koruma ilkelerine benzer ve aşağıdaki gibi yapılandırılmıştır:

- Birinci bölüm, veri koruma etki değerlendirmeleri (DPIA-data protection impact assesments) dahil olmak üzere yapay zekada hesap verebilirlik ve yönetişime değinir.
- İkinci bölüm, yapay zeka sistemi performansının değerlendirilmesi, iyileştirilmesi ve potansiyel ayrımcılığın azaltılması da dahil olmak üzere yasal temelleri içeren adil, yasal ve şeffaf süreçleri kapsar.
- Üçüncü bölüm, veri minimizasyonunu ve güvenliğini ele alır.

- Dördüncü bölüm, otomatik karar vermeyle ilgili haklar da dahil olmak üzere, yapay zeka sistemlerinde kişisel hakların kullanılmasını nasıl kolaylaştırabileceğinizle ilgilidir.

Detaylı olarak

Bu kılavuzu neden hazırladınız?

Sağlıktan işe alıma, ticarete ve ötesine kadar her gün yapay zekanın (AI) yeni alanlarda kullanımlarını görüyoruz.

Yapay zekanın organizasyonlara ve kişilere getirebileceği faydaları anlıyoruz, ancak riskler de var. Bu nedenle yapay zeka en önemli üç stratejik önceliğimizden biridir ve bu nedenle yapay zekanın veri koruma yükümlülüklerine uygunluğunu denetlemek için bir çerçeve geliştirmeye karar verdik.

Çerçeve:

- AI uygulamalarını denetlemek ve kişisel verilerin adil, yasal ve şeffaf bir şekilde işlemelerini sağlamak için bize sağlam bir metodoloji sunar.
- AI'dan kaynaklanan hak ve özgürlüklere yönelik riskleri değerlendirmek ve yönetmek için gerekli tedbirlerin alınmasını sağlar.
- AI kullanan organizasyonların uyumluluğunu değerlendirirken araştırma ve güvence ekiplerimizin çalışmalarını destekler.

Çerçeveyi kendi faaliyetlerimize rehberlik etmesi için kullanmanın yanı sıra, bunun arkasındaki düşüncemizi de paylaşmak istedik. Bu nedenle çerçevenin iki farklı çıktısı vardır:

1. Yapay zeka kullanan organizasyonların uyumluluğunu değerlendirirken, araştırma ve güvence ekiplerimiz tarafından kullanılacak denetim araçları ve prosedürlerini içerir.

2. Yapay zeka ve veri korumasına ilişkin bu ayrıntılı kılavuz, organizasyonlar için düşüncelerimizi özetler ve ayrıca organizasyonların kendi yapay zeka sistemlerinin uyumluluğunu denetlemesine yardımcı olmak için her bölümün sonunda gösterge niteliğindeki risk ve kontrol tablolarını içerir.

Bu kılavuz, veri koruma prensipleri ile uyumlu yapay zeka için en iyi uygulamayı oluşturduğunu düşündüklerimiz hakkında sizi bilgilendirmeyi amaçlamaktadır.

Bu kılavuz yasal bir kod değildir. İlgili yasanın AI için geçerli olduğu şekliyle nasıl yorumlanacağına dair tavsiyeler ve AI'nın neden olabileceği veya kötüleştirebileceği kişilere yönelik riskleri azaltmak için organizasyonel ve teknik önlemler için iyi uygulama önerileri içerir. Yasaya uymanın başka bir yolunu bulduğunuz sürece, iyi uygulama önerilerini benimsememeniz durumunda herhangi bir ceza uygulanmaz.

İlave okuma – ICO kılavuzu

Technology Strategy 2018-2021

'AI' ile ne demek istiyorsunuz?

'AI- Yapay Zeka' teriminin çeşitli anlamları vardır. AI araştırma topluluğu içinde, "deneyimlerden öğrenmek ve insanın akıllı davranışını taklit etmek için insan olmayan bir sistemi kullanmak" gibi çeşitli yöntemlerle bahsedilirken, veri koruma kaynaklarında ise 'görsel algı, konuşma tanıma, karar verme ve diller arası çeviri gibi normalde insan zekası gerektiren görevleri yerine getirebilen bilgisayar sistemlerinin teorisi ve gelişimi' olarak anılır.

Bununla birlikte, kişiler hakkında tahminler veya sınıflandırmalar yapmak için büyük miktarda verinin kullanılması, 19. yüzyıldan bu yana sigorta gibi sektörlerde var olmuştur, 1950'lerde 'AI' teriminin ortaya çıkmasından çok önce. Bu geleneksel istatistiksel analiz biçimlerinde, istatistiksel modeller kalem ve kağıt kullanılarak (ve daha sonra bir hesap makinesiyle) hesaplanır.

Modern makine öğrenimi (ML) teknikleriyle; çok daha büyük, çok boyutlu veri setlerinde yoğun hesaplama teknikleri kullanarak istatistiksel modeller oluşturmak artık çok daha kolay. Bu tür modellerin karmaşıklığındaki bu artış; onları oluşturmanın azalan maliyetleriyle birleştiğinde, bu modellerin kişilerin hak ve özgürlüklerine yönelik risklere ilişkin endişeleri artırdı.

Yapay zekanın öne çıkan bir alanı; büyük miktarda veri (genellikle karmaşık) kullanarak istatistiksel modeller (tipik olarak) oluşturmak için hesaplama tekniklerinin kullanılması olan "makine öğrenimi"dir (ML). Bu modeller, yeni veri göstergeleri hakkında sınıflandırmalar veya tahminler yapmak için kullanılabilir.

AI'ın tamamı ML'i içermese de, yapay zekaya olan son ilginin çoğu, ister görüntü tanıma, ister konuşmadan metne veya isterse kredi riskini sınıflandırma durumlarında olsun, bir şekilde ML tarafından yönlendirilir. Bu nedenle bu

kılavuz, diğer yapay zeka türlerinin farklılık gösterebileceğini kabul ederken, makine öğrenimi tabanlı yapay zekanın sunabileceği veri koruma zorluklarına odaklanır.

Herhangi bir amaçla büyük miktarda kişisel veri işliyorsanız, veri koruma kanunu geçerlidir. Bu verileri istatistiksel modellerle işliyor ve bu modelleri insanlar hakkında tahminler yapmak için kullanıyorsanız eğer, bu faaliyetlerin ML (veya AI) olarak sınıflandırılmasına gerek yoktur ve bu kılavuz sizinle alakalı olacaktır.

"Yapay Zeka" şemsiye terimini kullanıyoruz çünkü bu; organizasyonların insan düşüncesini taklit eden bir dizi teknolojiye atıfta bulunmasının yaygın bir yolu haline geldi. Benzer risk kaynaklarına sahip benzer teknolojilerin, aynı risk ölçütleri setinden faydalanması muhtemeldir. Bu nedenle, buna ister yapay zeka, ister makine öğrenimi, isterse karmaşık bilgi işleme veya başka bir şey deyin, burada tanımlanan riskler ve kontroller yardımcı olacaktır. Farklı yapay zeka türleri arasında, örneğin basit regresyon modelleri ve derin sinir ağları arasında önemli farklılıklar olduğunda, bunlara açıkça değineceğiz.

Diğer kaynaklar

Telekomünikasyonda Veri Koruma Uluslararası Çalışma Grubu'nun (International Working Group on Data Protection in Telecommunications) 2018, gizlilik ve yapay zeka ile ilgili çalışma belgesine bakın (pdf, dış bağlantı)

Bu kılavuz, AI üzerindeki diğer ICO çalışmalarıyla nasıl ilişkilidir?

Bu kılavuz, aşağıdakiler dahil olmak üzere mevcut ICO kaynaklarını tamamlamak üzere tasarlanmıştır:

- 2014'de yayınlanan ve 2017'de güncellenen Büyük Veri, Yapay Zeka, Makine Öğrenimi ve Veri Koruma raporu
- Alan Turing Enstitüsü ile işbirliği içinde hazırlanan, AI ile alınan kararları açıklamaya yönelik kılavuzumuz (explAIn guidance)

Büyük Veri Raporu, bu teknolojilerin veri koruma içeriklerini anlamak için güçlü bir temel sağladı. Komisyonun 2017 baskısının önsözünde belirtildiği gibi, bu karmaşık ve hızlı gelişen bir alandır. Son üç yılda hem yapay zekanın kişilere getirdiği riskler hem de bu riskleri ele almak için alınabilecek organizasyonel ve teknik önlemler açısından yeni düşünceler ortaya çıktı. Paydaşlarla olan ilişkimiz sayesinde, organizasyonların AI'yı sahada nasıl kullandığına dair, 2017 raporunda sunulanların ötesinde ek bilgiler elde ettik.

AI tarafından gündeme getirilen bir diğer önemli zorluk da açıklanabilirliktir. Hükümetin AI Sektör Anlaşmasının bir parçası olarak, Turing Enstitüsü ile işbirliği içinde, organizasyonların AI kullanımlarını kişilere en iyi nasıl açıklayabilecekleri konusunda rehber hazırladık. Bu, geçen yıl istişare için taslak halinde yayınlanan Açıklama Kılavuzuyla (ExplAIn guidance) sonuçlandı. Şu anda paydaşlardan

gelen geri bildirimler ışığında açıklama kılavuzunu sonuçlandırma sürecindeyiz ve tamamlandığında bu kılavuzdaki bağlantıları güncelleyeceğiz.

Açıklama kılavuzu, kişiler için AI'nın açıklanabilirliğinin zorluğunu halihazırda önemli ayrıntılarla kapsıyor olsa da, bu kılavuz, kurum içinde AI açıklanabilirliği hakkında, örneğin dahili gözetim ve uyumluluk için bazı ek hususları içerir. İki kılavuz parçası birbirini tamamlayıcı niteliktedir ve her ikisini birlikte okumanızı öneririz.

İlave Okuma – ICO kılavuzu

ICO'nun Büyük Veri, Yapay Zeka, Makine Öğrenimi ve Veri Koruma raporu ve Turing danışmanlığında AI kararlarının açıklama kılavuzu

Bu rehber kimler içindir?

Bu kılavuz, veri koruma uyumlu yapay zeka için en iyi uygulamaları kapsar. İki geniş hedef kitleye yöneliktir.

İlk olarak, aşağıdakiler de dahil olmak üzere uyumluluk odaklı olanlar:

- Veri koruma görevlileri
- Genel danışman
- Risk yöneticileri
- ICO'nun kendi denetçileri – diğer bir deyişle, veri koruma mevzuatı kapsamında denetim fonksiyonlarımızı yerine getirirken bu kılavuzu kendimiz de kullanacağız.

İkincisi, aşağıdakiler de dahil olmak üzere teknoloji uzmanları:

- Makine öğrenimi geliştiricileri ve veri bilimcileri
- Yazılım geliştiricileri / mühendisleri
- Siber Güvenlik ve BT Risk yöneticileri.

Bu kılavuz, her iki hedef kitlenin de erişebileceği şekilde yazılmış olsa da, bazı bölümler öncelikle uyumluluk veya öncelikle teknoloji rollerinde bulunanlara yöneliktir ve buna göre işaretlenmiştir.

ICO neden risk temelli bir yaklaşımla yapay zekaya odaklanıyor?

Risk temelli bir yaklaşım benimsemek şu anlama gelir:

- AI kullandığınızda ortaya çıkabilecek kişilerin hak ve özgürlüklerine yönelik risklerin değerlendirilmesi.
- Bu riskleri azaltmak için uygun ve doğru teknik ve organizasyonel önlemlerin uygulanması.

Bunlar, veri koruma kanunundaki genel şartlardır. Riskler düşükse yasa görmezden gelinebilir anlamına gelmez ve bu riskleri yeterince azaltamıyorsanız, planlanmış AI projesini durdurmanız gerektiği anlamına gelebilirler.

Bu kılavuzu mevcut risk yönetimi sürecinize entegre etmenize yardımcı olmak için, onu birkaç ana risk alanında düzenledik. Her bir risk alanı için aşağıdakileri açıklıyoruz:

- İlgili riskler
- AI'nın, risk olasılık ve/veya etkilerini nasıl artırabileceği
- Bu riskleri belirlemek, değerlendirmek, en aza indirmek, izlemek ve kontrol etmek için kullanabileceğiniz bazı olası önlemler.

Dahil edilen teknik ve organizasyonel önlemler, çok çeşitli durumlarda iyi uygulama olarak kabul ettiğimiz önlemlerdir. Ancak, benimsemeniz gerekebilecek risk kontrollerinin çoğu duruma özel olduğundan, kapsamlı veya kesin bir liste ekleyemiyoruz.

Bu kılavuz, hem yapay zekaya hem de veri korumaya özgü riskleri ve bu risklerin yönetişim (governance) ve hesap verebilirlik (accountability) üzerindeki etkilerini kapsar. AI kullanıp kullanmadığınıza bakılmaksızın, hesap verebilirlik önlemlerine sahip olmalısınız.

Ancak yapay zeka uygulamalarını benimsemek, mevcut yönetişim ve risk yönetimi uygulamalarınızı yeniden değerlendirmenizi gerektirebilir. Yapay zeka uygulamaları mevcut riskleri artırabilir, yenilerini ortaya çıkarabilir veya genel olarak risklerin değerlendirilmesini veya yönetilmesini zorlaştırabilir. Bu nedenle organizasyonunuzdaki karar vericiler, mevcut veya önerilen AI uygulamaları ışığında organizasyonunuzun risk iştahını yeniden gözden geçirmelidir.

Bu kılavuzun bölümlerinin her biri, AI zorluk alanlarından birine derinlemesine dalar ve ilgili riskleri, süreçleri ve kontrolleri araştırır.

Bu kılavuz bir dizi yapay zeka ilkesi mi?

Bu kılavuz, yapay zeka kullanımı için genel etik ve/veya tasarım ilkeleri sağlamaz. "Yapay zeka etiği" ve veri koruma arasında (veri koruma yasasında halihazırda yansıtılan bazı önerilen etik ilkeleriyle) örtüşmeler olsa da, bu kılavuz veri koruma uyumluluğuna odaklanmıştır.

Veri koruma, AI tasarımcılarının işlerini nasıl yapmaları gerektiğini belirlememesine rağmen, kişisel verileri işlemek için AI kullanıyorsanız, tasarım ve default olarak veri koruma ilkelerine uymanız gerekir. Bunların, işlemde rol almayan geliştiriciler için doğrudan uygulaması olmayabilir - ancak şunu da unutmamalısınız:

- Veri koruma ilkelerini etkili bir şekilde uygulamak için tasarlanmış; uygun teknik ve organizasyonel önlemleri uygulamada kullanmak sizin sorumluluğunuzdadır ve

- Bir geliştiriciyseniz ve modellemede kişisel verileri kullanıyorsanız; o zaman bu işlem ve veri koruma yasasının bir denetleyicisi olacaksınız.

Belirli tasarım seçeneklerinin, şu veya bu şekilde veri korumasını ihlal eden yapay zeka sistemleriyle sonuçlanması daha olasıdır. Bu kılavuz, tasarımcıların ve mühendislerin bu seçenekleri daha iyi anlamalarına yardımcı olacak, böylece kişilerin hak ve özgürlüklerini korurken yüksek performanslı sistemler tasarlayabileceksiniz.

Çalışmamızın yalnızca yapay zeka ile gelen ve artan veri koruma zorluklarına odaklandığını belirtmekte fayda var. Bu nedenle, AI ile ilgili olan ve AI tarafından zorlanılan durumlar dışında kalan daha genel veri koruma hususları ele alınmaz.

Diğer kaynaklar

AI hususunda etik ve veri korumanın nasıl kesiştiği hakkında daha fazla bilgi için; Küresel Gizlilik Kurulu (Global Privacy Assembly)'nun Yapay zekada etik ve veri koruma bildirgesi (pdf, dış bağlantı) okuyabilirsiniz.

Hangi mevzuat geçerlidir?

Bu kılavuz, yapay zekanın veri koruması için ortaya çıkardığı zorluklarla ilgilenir. Bu nedenle, Birleşik Krallık mevzuatının en alakalı parçası 2018 Veri Koruma Yasasıdır (Data Protection Act).

DPA 2018, Genel Veri Koruma Yönetmeliği (GDPR) ile birlikte Birleşik Krallığın veri koruma çerçevesini belirler. Aşağıdaki veri koruma rejimlerini içerir:

- Bölüm 2 – Birleşik Krallık'ta GDPR'ı tamamlar ve uyarlar.
- Bölüm 3 – Emniyet yetkilileri için ayrı bir rejim belirler.
- Bölüm 4 – Üç istihbarat servisi için ayrı bir rejim belirler.

Bu kılavuzun çoğu; süreçlerinizde, DPA'nın hangi bölümünün uygulandığına bakılmaksızın geçerli olacaktır. Bununla birlikte, farklı rejimlerin gereksinimleri arasında ilgili farklılıklar olduğu durumlarda, bunlar metinde açıkça ele alınmaktadır.

Brexit'in veri koruma yasasını nasıl etkilediğine ilişkin kılavuzumuzu da gözden geçirmelisiniz.

AI'nın veri koruma dışındaki ICO yetkinliği alanları, özellikle Bilgi Özgürlüğü (Freedom of Information) üzerindeki etkileri burada dikkate alınmamıştır.

İlave okuma – ICO kılavuzu

Farklı rejimler hakkında daha fazla bilgi için Veri Koruma kılavuzu okuyun.

Veri koruma ve Brexit hakkında daha fazla ayrıntıya ihtiyacınız varsa, SSS'lerimize bakın

Bu kılavuz nasıl yapılandırılmıştır?

Bu kılavuz, farklı veri koruma ilkelerine ve haklarına karşılık gelen birkaç bölüme ayrılmıştır.

Birinci bölüm; öncelikle hesap verebilirlik ilkesiyle ilgili konuları ele almaktadır. Bu, veri koruma ilkelerine uymaktan ve bu uyumu göstermekten sorumlu olmanızı gerektirir. Bu bölümdeki kısımlar, veri koruma etki değerlendirmeleri (DPIA- data protection impact assesments), denetim / işleyiş sorumlulukları ve ödünleşmenin (trade-off) doğrulanması ve gerekçelendirilmesi dahil olmak üzere hesap verebilirliğin yapay zekaya özgü etkileriyle ilgilidir.

İkinci Bölüm; AI sistemlerindeki kişisel verilerin işlenmesi, AI sistemindeki kişisel verilerin değerlendirilmesi / iyileştirilmesi, yapay zeka sistemi performansının değerlendirilmesi / iyileştirilmesi ve adil işlemeyi sağlamak için potansiyel ayrımcılığın azaltılması ile ilgili kısımlarla, AI sistemlerindeki kişisel verilerin işlenmesine ilişkin yasalara uygun, adil ve şeffaf bilgileri kapsar.

Üçüncü bölüm; AI sistemlerinde güvenlik ve veri minimizasyonu ilkesini kapsar.

Dördüncü bölüm; AI sistemlerinde bireylerin kişisel verileriyle ilgili haklarını ve sadece otomatik kararlarla ilgili hakların kullanımını nasıl kolaylaştırabileceğinizi kapsar. Özellikle dördüncü bölümde; otomatik olmayan veya kısmen otomatik kararlarda anlamlı insan girdisini nasıl sağlayabileceğiniz ve tamamen otomatikleştirilmiş kararların anlamlı bir şekilde insan tarafından incelenmesini nasıl sağlayabileceğiniz konuları ele alınmaktadır.

Yapay zekanın hesap verebilirlik ve yönetişim etkileri nelerdir?

Genel bir bakış

Hesap verebilirlik ilkesi, sizi veri korumasına uygun olmaktan ve bu uyumluluğu herhangi bir AI sisteminde göstermekten sorumlu kılar. AI kapsamında, hesap verebilirlik şunları yapmanızı gerektirir:

- Sisteminizin uyumluluğundan sorumlu olma.
- Riskleri değerlendirme ve azaltma.
- Sistemin nasıl uyumlu olduğunu belgeleme ve gösterme ile yapılan seçimleri gerekçelendirme.

Kullanmayı düşündüğünüz herhangi bir sistem için bu konuları DPIA'nızın bir parçası olarak düşünmelisiniz. Kişisel verileri işleyen AI sistemlerini kullanıyorsanız, yasal olarak bir DPIA tamamlamanız gerektiğini unutmayın. DPIA'lar size kişisel verileri işlemek için AI sistemlerini nasıl ve neden kullandığınızı ve potansiyel risklerin neler olabileceğini düşünme fırsatı sunar.

Yapay zeka tedarik zincirlerinde tipik olarak yer alan çeşitli işlem türlerinin karmaşıklığı ve karşılıklı bağımlılığı nedeniyle, denetleyici / işleyici ilişkilerini anlamaya ve tanımlamaya da özen göstermeniz gerekir.

Ek olarak; nasıl tasarlandıklarına ve uygulandıklarına bağlı olarak, AI sistemleri kaçınılmaz olarak, gizlilik ile diğer artan haklar ve çıkarlar arasında ödünleşme yapmayı içerecektir. Bu ödünleşmelerin neler olabileceğini ve bunları nasıl yönetebileceğinizi bilmeniz gerekir, aksi takdirde bunları yeterince değerlendirememe ve doğru dengeyi sağlayamama riski vardır. Bununla birlikte, her zaman temel veri koruma ilkelerine uymanız gerektiğini ve bu gerekliliği "ödünleşme yapamadığınızı" da unutmamalısınız.

Detaylı olarak

- Yapay zeka yönetişimi ve risk yönetimine nasıl yaklaşmalıyız?
- Anlamlı bir risk iştahını nasıl oluşturmalıyız?
- Yapay zeka için veri koruma etki değerlendirmeleri yaparken nelere dikkat etmeliyiz?
- Yapay zekada denetleyici/işlem ilişkilerini nasıl anlamalıyız?
- Yapay zeka ile ilgili ödünleşmeler nelerdir ve bunları nasıl yönetmeliyiz?

Yapay zeka yönetişimi ve risk yönetimine nasıl yaklaşmalıyız?

Yapay zeka, iyi kullanılırsa organizasyonları daha verimli, etkili ve yenilikçi hale getirme potansiyeline sahiptir. Bunun yanında, AI, organizasyonlar için uyum zorluklarının yanında aynı zamanda kişilerin hak ve özgürlükleri için de önemli riskler doğurmaktadır.

Farklı teknolojik yaklaşımlar; bu sorunlardan bazılarını ya şiddetlendirecek ya da hafifletecektir, ancak birçokları belirgin teknolojilerden çok daha geniş kapsamlıdır. Bu kılavuzun geri kalanının da önerdiği gibi, yapay zekanın veri koruma etkileri; belirli kullanım durumlarına, bunların yaygınlaştırıldıkları popülasyona, örtüştüğü diğer düzenleyici gerekliliklere ve ayrıca sosyal, kültürel ve politik hususlara büyük ölçüde bağlıdır.

Yapay zeka; tasarım ve default olarak veri korumasını organizasyonun kültürüne ve süreçlerine oturtmanın önemini artırırken, AI sistemlerinin teknik karmaşıklıkları bunu zorlaştırabilir. Bu karmaşıklıkları nasıl ele aldığınızı göstermek, hesap verebilirliğin önemli bir unsurudur.

Bu sorunları veri bilimcilerine veya mühendislik ekiplerine devredemezsiniz. Veri Koruma Görevlileri (DPO'lar) dahil olmak üzere üst yönetiminiz, onları uygun ve hızlı bir şekilde anlamaktan ve ele almaktan sorumludur.

Bunu yapmak için; kendi becerilerini geliştirmelerine ek olarak, sorumluluklarını yerine getirirken kendilerini destekleyecek çeşitli, iyi kaynaklara sahip ekiplere ihtiyaçları vardır. Ayrıca dahili yapılarınızı, rol ve sorumluluk haritalarınızı, eğitim gereksinimlerinizi, politikalarınızı ve teşviklerinizi genel yapay zeka yönetişimi ve risk yönetimi stratejinizle uyumlu hale getirmeniz gerekir.

Yatırımının başlangıç ve devam eden aşamalarında; gerekli olan kaynak ve eforu hafife almamanız önemlidir. Yönetişim ve risk yönetimi yetenekleriniz, yapay zeka kullanımınız ile orantılı-uygun olmalıdır. Bu özellikle esastır. Çünkü AI'nın benimsenmesi hala ilk aşamalarında ve teknolojinin kendisinin yanı sıra ilgili yasalar, düzenlemeler, yönetişim ve risk yönetimi en iyi uygulamaları hala hızlı bir şekilde gelişmeye devam etmektedir.

Ayrıca şu anda daha genel bir hesap verebilirlik araç seti geliştiriyoruz. Bu, AI'ya özgü değildir, ancak GDPR kapsamında hesap verebilirliğinizi göstermek için bir temel sağlar ve AI hesap verebilirliğine yaklaşımınızı bunun üzerine inşa edebilirsiniz. Bu kılavuzun son halini, yayınlandığında hesap verebilirlik araç setinin son versiyonuna atıfta bulunacak şekilde güncelleyeceğiz.

Anlamlı bir risk iştahını nasıl oluşturmalıyız?

Veri koruma yasasının risk tabanlı yaklaşımı, yükümlülüklere uymanızı ve özel koşullar (yapmayı düşündüğünüz işlemlerin niteliği, kapsamı, durumu ve amaçları ile bunun kişilerin hak ve özgürlüklerine yönelik oluşturduğu riskler) konusunda uygun önlemleri uygulamanızı gerektirir.

Bu nedenle, uyumla ilgili değerlendirmeleriniz, kişilerin hak ve özgürlüklerine yönelik risklerin değerlendirilmesini ve bu koşullarda neyin uygun olduğuna ilişkin kararların alınmasını içerir. Her durumda, veri koruma gereksinimlerine uygunluğu sağlamanız gerekir.

Bu, kişisel verileri işleyen diğer teknolojilerde olduğu gibi AI kullanımı için de geçerlidir. Yapay Zeka ile ortaya çıkan risklerin özel niteliği ve işleme koşulları, veri koruma uyumluluğunu sağlamaya devam ederken, rekabet çıkarları arasında uygun bir denge kurmanızı gerektirir. Bu da işleminizin sonucunu etkileyebilir. Hak ve özgürlüklerle ilgili risklere "sıfır tolerans" yaklaşımını benimsemek gerçekçi değildir ve aslında yasa sizden bunu yapmanızı gerektirmez – mesele bu risklerin tanımlanmasını, yönetilmesini ve azaltılmasını sağlamakla ilgilidir. Aşağıdaki 'Ödünleşme nedir ve nasıl yönetmeliyiz?' konusuna bakın.

Yapay zeka sistemlerinizde kişisel verilerin işlenmesinden kaynaklanan kişilere yönelik riskleri yönetmek için, temel haklar, riskler ve bunların ve diğer çıkarların nasıl dengeleneceği konusunda olgun bir anlayış ve net ifade geliştirmeniz önemlidir. Sonuçta, yapmanız gerekenler:

- AI kullanımınızın oluşturduğu kişilerin haklarına yönelik riskleri değerlendirmek.
- Bunları nasıl ele almanız gerektiğini belirlemek.
- Bunun yapay zeka kullanımı üzerindeki etkisini belirlemek.

Yaklaşımınızın, hem organizasyonunuza hem de işlem koşullarınıza uygun olduğundan emin olmalısınız. Uygun olduğunda, risk değerlendirme çerçevelerini de kullanmalısınız.

Bu, doğru olması zaman alabilen karmaşık bir iştir. Ancak sonunda size ve ICO'ya risk pozisyonlarınız ile uyum ve risk yönetimi yaklaşımlarınızın yeterliliği hakkında daha eksiksiz ve daha anlamlı bir görünüm verecektir.

Aşağıdaki bölümler, hesap verebilirliğin yapay zekaya özgü sonuçlarıyla ilgilidir:

- AI sistemleri için veri koruma etki değerlendirmelerini nasıl üstlenmeniz gerektiği.
- AI sistemlerinin geliştirilmesi ve devreye alınmasında yer alan belirli süreç operasyonları için bir denetleyici veya işleyici olup olmadığınızı nasıl belirleyeceğiniz ve bunun sonucunda sorumluluklarınız üzerindeki etkileri nasıl belirleyebileceğiniz.
- Kişilerin hak ve özgürlüklerine yönelik riskleri nasıl değerlendirmeniz ve bir yapay zeka sistemi tasarlarken veya kullanmaya karar verirken bunları nasıl ele almanız gerektiği.
- Söz konusu işlem için AI kullanma kararınız da dahil olmak üzere, uyguladığınız yaklaşımı nasıl gerekçelendirmeniz, belgelemeniz ve göstermeniz gerektiği.

Yapay zeka için veri koruma etki değerlendirmeleri yaparken nelere dikkat etmeliyiz?

DPIA'lar; tasarım olarak veri koruma yasasının, sorumluluk ve veri koruma konularına odaklandığı önemli bir parçasıdır.

DPIA'ları yalnızca bir kutu işaretleme uyumluluk alıştırması olarak görmemelisiniz. Yapay zeka kullanımının yol açabileceği hak ve özgürlüklere yönelik riskleri belirlemeniz ve kontrol etmeniz için etkili bir şekilde yol haritaları olarak kullanılabilirler. Ayrıca, AI sistemlerinin tasarımında veya tedariğinde verdiğiniz kararlar için hesap verebilirliği dikkate almanız ve göstermeniz için mükemmel bir fırsattır.

Veri koruma yasası kapsamında neden DPIA'ları uygulamamız gerekiyor?

Kişisel verileri işlemek için yapay zekanın kullanılması, büyük ihtimal kişilerin hak ve özgürlükleri için yüksek riskle sonuçlanabilir ve dolayısı ile DPIA'yı üstlenmek yasal gerekliliği tetikler.

Değerlendirmenin sonucu, kişiler için yeterince azaltılamayacak yüksek artık risk gösteriyorsa, işleme başlamadan önce ICO'ya danışmalısınız.

DPIA'yı gerçekleştirmenin yanı sıra, diğer etki değerlendirme türlerini gerçekleştirmeniz veya bunu gönüllü olarak yapmanız da gerekebilir. Örneğin, kamu sektörü organizasyonlarının eşitlik etki değerlendirmeleri yapması gerekirken, diğer organizasyonların gönüllü olarak `algoritma etki değerlendirmeleri' yapması gerekir. Değerlendirme bir DPIA'nın tüm gerekliliklerini kapsadığı sürece bu alıştırmaları birleştirmemeniz için hiçbir neden yoktur.

ICO, DPIA'lar hakkında, ne zaman gerekli olduklarını ve bunların nasıl tamamlanacağını açıklayan DPI'lar hakkında ayrıntılı bir kılavuz oluşturmuştur. Bu bölüm, AI sistemlerinde kişisel verilerin işlenmesi için bir DPIA gerçekleştirirken düşünmeniz gereken bazı şeyleri ortaya koymaktadır.

Mevzuattaki ilgili hükümler

GDPR'ın 35 ve 36 maddeleri ve 74-77, 84, 89-92, 94 ve 95 gerekçeleri bakın (dış bağlantı)

DPA 2018'in- 64 and 65 bölümleri (dış bağlantı)

DPIA yapıp yapmamaya nasıl karar veririz?

AI, yüksek riskle sonuçlanması muhtemel işlem türleri listesinde yer aldığından, eğer AI'yi kişisel verileri işlemek için kullanırsanız, bir DPIA gerçekleştirmelisiniz. Her durumda, kişisel verilerin kullanımını içeren büyük bir projeniz varsa, bir DPIA yapmak iyi bir uygulamadır.

DPIA gerektiren operasyon örnekleri için "yüksek riskle sonuçlanması muhtemel" işlem operasyonları listesini ve hangi kriterlerin diğerleriyle birlikte yüksek riskli olduğu hakkında daha fazla ayrıntı okuyabilirsiniz.

İlave okuma – ICO kılavuzu

DPIA'lara ilişkin kılavuzumuzda 'DPIA yapı yapmamaya nasıl karar veririz? konusuna bakın.

DPIA'mızda neleri değerlendirmeliyiz?

DPIA'nızın; kişisel verilerin herhangi bir şekilde işlenmesinin niteliğini, kapsamını, durumunu ve amaçlarını tanımlaması gerekir - verileri işlemek için AI'yı nasıl ve neden kullanacağınızı netleştirmesi gerekir. Aşağıdakileri detaylandırmanız gerekir:

- Verileri nasıl toplayacağınız, saklayacağınız ve kullanacağınız.
- Verilerin hacmi, çeşitliliği ve hassasiyeti.
- Kişilerle olan ilişkinizin doğası.
- Sizin için olduğu kadar kişiler veya daha geniş toplum için amaçlanan sonuçlar.

AI yaşam döngüsü kapsamında, proje geliştirmenin en erken aşamalarında bir DPIA üstlenirseniz, DPIA amacına en iyi şekilde hizmet edecektir. DPIA asgari olarak aşağıdaki temel bileşenleri içermelidir.

İşlemi nasıl tanımlarız?

DPIA'nız şunları içermelidir:

- Yapay zeka süreçlerinin ve otomatik kararların kişiler üzerinde etkisi olabileceği aşamalar ve veri akışları dahil olmak üzere işlem faaliyetlerinin sistematik bir açıklaması.
- Kişisel veri işlemenin tarafsızlığını etkileyebilecek sistemin performansındaki herhangi ilgili bir varyasyon veya hata payının açıklaması (bkz. 'İstatistiksel Doğruluk')
- Aşağıdakiler dahil olmak üzere, işlemenin kapsamının ve durumunun bir açıklaması:
    - Hangi verileri işleyeceksiniz
    - Dahil olan veri konularının sayısı

- Veri kaynağı
- Kişilerin işlemeyi ne kadar bekleyebilecekleri.

Bir DPIA, karar verme sürecine herhangi bir insan katılımının derecesini ve bunun hangi aşamada gerçekleştiğini belirlemeli ve kaydetmelidir. Otomatik kararların insan müdahalesine veya incelemesine sebep olduğu durumlarda, bunun anlamlı olmasını sağlamak için süreçler uygulamalı ve ayrıca kararların bozulabileceği gerçeğini detaylandırmalısınız.

Karmaşık bir yapay zeka sisteminin işleme faaliyetini tanımlamak zor olabilir. Aşağıdakilerle birlikte bir değerlendirmenin iki versiyonunu sağlamanız uygun olabilir:

- Birincisi, uzman izleyiciler için kapsamlı bir teknik açıklama sunma.
- İkincisi, işlemenin daha üst düzey bir tanımını içerir ve kişisel veri girdilerinin kişileri etkileyen çıktılarla ilişkisinin mantığını açıklama.

Bir DPIA, bir denetleyici olarak rollerinizi ve yükümlülüklerinizi belirlemeli ve ilgili işleyicileri içermelidir. AI sistemlerinin kısmen veya tamamen dış sağlayıcılara yaptırıldığı durumlarda, hem siz hem de ilgili diğer organizasyonlar, GDPR'nin 26. Maddesi kapsamında ortak denetçiliğin mevcut olup olmadığını da değerlendirmelisiniz ve eğer varsa, DPIA sürecinde uygun şekilde işbirliği yapmalısınız.

Bir işleyici kullandığınızda, o işleyiciden gelen bilgileri yeniden üreterek bir DPIA'daki işlem faaliyetinin daha teknik öğelerinden bazılarını tanımlayabilsiniz. Örneğin bir işleyicinin kılavuzundan bir akış şeması oluşturmak. Ancak, bir işleyicinin kendi literatüründeki büyük bölümleri kendi değerlendirmenize kopyalamaktan genellikle kaçınmalısınız.

Mevzuattaki ilgili hükümler

Bkz. GDPR - 35(7)(a) madde ve 84, 90 ve 94 gerekçeleri (Dış bağlantı)

Kimseye danışmamıza gerek var mı?

Aşağıdakileri yapmalısınız:

- Kişilerin veya onların temsilcilerinin görüşlerini araştırın ve belgeleyin. (yapmamak için iyi bir sebep olmadıkça).
- İlgili tüm iç paydaşlara danışın.
- İşleyici kullanıyorsanız, işleyicinizi dikkate alın.
- Uygun olduğunda yasal tavsiye veya başka bir uzmanlık almayı göz önünde bulundurun.

Bunu yapmamak için iyi bir neden olmadıkça, bir DPIA sırasında amaçlanan işlem operasyonu hakkında kişilerin veya onların temsilcilerinin görüşlerini almalı ve belgelemelisiniz. Bu nedenle, süreci danışanların anlayabileceği bir şekilde tanımlayabilmeniz önemlidir.

Mevzuattaki ilgili hükümler

Bknz. GDPR - 28(3)(f) madde ve 35 (9) madde (dış bağlantı)

Gerekliliği ve orantılılığı nasıl değerlendiririz?

Kişisel verileri işlemek için bir yapay zeka sisteminin devreye alınması, teknolojinin uygunluğu ile değil, bu sistemin belirli ve meşru bir amacı yerine getirme konusundaki kanıtlanmış yeteneği ile yönlendirilmelidir.

Bir DPIA, AI sistemi tarafından kişisel verilerinizin işlenmesinin orantılı bir faaliyet olduğunu göstermenize de olanak tanır. Orantılılığı değerlendirirken; yapay zekayı, kişilerin hak ve özgürlüklerine karşı oluşturabileceği risklere karşı kullanma konusundaki çıkarlarınızı tartmanız gerekir. AI sistemleri için, kullanılan algoritmalarda ve veri setlerinde yanlılık (önyargı) veya yanlışlıktan kaynaklanabilecek kişilere yönelik her türlü zararı düşünmeniz gerekir..

Bir DPIA'nın orantılılık unsuru dahilinde, kişilerin bir AI sisteminin işlemi gerçekleştirmesini makul bir şekilde bekleyip bekleyemeyeceğini değerlendirmeniz gerekir. Yapay zeka sistemleri, insan karar verme sürecini tamamlıyor veya yerini alıyorsa, bizi daha iyi savunabilmesi için projenin insan ve algoritmik doğruluğu yan yana nasıl karşılaştırabileceğini DPIA'da belgelemelisiniz.

Ayrıca istatistiksel doğruluk ve verilerin en aza indirilmesi gibi yapılan tüm ödünleşmeleri açıklamalı ve bunların metodolojisini ve gerekçesini belgelemelisiniz.

Riskleri nasıl tanımlar ve değerlendiririz?

DPIA süreci, ilgili riskleri objektif bir şekilde tanımlamanıza yardımcı olacaktır. Her bir riske, kişilerin üzerindeki etkinin şiddetine ve olasılığına göre ölçülen bir puan veya seviye atamanız gerekir.

Yapay zeka sistemlerinin geliştirilmesi ve dağıtımında kişisel verilerin kullanılması, yalnızca kişilerin bilgi hakları için risk oluşturmayabilir. Bir DPIA, risk kaynaklarını değerlendirirken, maddi ve maddi olmayan zararın veya hasarın kişiler üzerindeki potansiyel etkisini dikkate alması gerekir.

Örneğin; makine öğrenimi sistemleri, eşitlik mevzuatına aykırı olabilecek şekilde, verilerdeki tarihi kalıplardan ayrımcılığı yeniden üretebilir. Benzer şekilde, içerik oluşturucunun kişisel verilerinin analizine dayalı olarak içeriğin yayınlanmasını

durduran yapay zeka sistemleri, onların ifade özgürlüklerini etkileyebilir. Bu tür durumlarda, veri korumanın ötesinde ilgili yasal çerçeveleri göz önünde bulundurmalısınız.

Mevzuattaki ilgili hükümler

Bknz. GDPR - 35(7)(c) madde ve 76 ve 90 gerekçeler (dış bağlantı)

Azaltıcı önlemleri nasıl belirleriz?

Tanımlanan her riske karşı, değerlendirilen risk seviyesini daha da azaltmak için seçenekleri göz önünde bulundurmalısınız. Buna örnek olarak; veri minimizasyonu veya kişilere işlemeden vazgeçme olanakları sağlamak olabilir.

Riski azaltmanın veya önlemenin yollarını düşünürken DPO'nuza danışmalısınız ve seçtiğiniz önlemin söz konusu riski azaltıp azaltmadığını veya ortadan kaldırıp kaldırmadığını DPIA'nıza kaydetmelisiniz.

DPO'ların ve diğer bilgi yönetişimi profesyonellerinin en erken aşamalardan itibaren AI projelerine dahil olmaları önemlidir. Onlar ve proje ekipleri arasında açık ve net iletişim kanalları olmalıdır. Bu, AI yaşam döngüsünün başlarında riskleri tanımlayabilmelerini ve ele alabilmelerini sağlayacaktır.

Veri koruma sonradan düşünülmemeli ve bir DPO'nun profesyonel görüşü on birinci saatte sürpriz olmamalıdır.

Yapay Zeka sistemlerinin geliştirilmesi, test edilmesi, doğrulanması, uygulamaya alınması ve izlenmesinden sorumlu kişilerin yeterince eğitimli olmaları ve işlemenin veri koruma etkileri konusunda anlam sahibi olmalarını sağlamak amacıyla DPIA'yı kullanarak uygulamaya koymuş olabileceğiniz önlemleri belgeleyebilirsiniz.

DPIA'nız, uygun eğitim gibi insan hatasıyla ilişkili riskleri azaltmak için uyguladığınız kurumsal önlemleri de ispatlayabilir. Ayrıca yapay zeka sisteminizde işlenen kişisel verilerin güvenliği ve doğruluğu ile ilgili riskleri azaltmak için tasarlanmış tüm teknik önlemleri de belgelemelisiniz.

Belirlenen riskleri azaltmak için önlemler alındıktan sonra DPIA, işlemden kaynaklanan artık risk seviyelerini belgelemelidir.

Tanımlanan her riski ortadan kaldırmanız gerekmez. Ancak değerlendirmeniz yeterince azaltamamanız gereken yüksek bir risk olduğunu gösteriyorsa işleme devam etmeden önce ICO'ya danışmanız gerekir.

DPIA'mızı nasıl sonlandırırız?

Şunları kaydetmelisiniz:

- Almayı planladığınız ek önlemler.
- Her bir riskin ortadan kaldırılıp kaldırılmadığı, azaltılıp azaltılmadığı veya kabul edilip edilmediği.
- Ek önlemler alındıktan sonra genel 'artık risk' seviyesi.
- Varsa DPO'nuzun görüşü.
- ICO'ya danışmanız gerekip gerekmediği.

Sonra ne olur?

Kişisel verilerin işlenmesi başlamadan önce DPIA'nızı gerçekleştirmeniz gerekse de, bunu 'canlı' bir belge olarak da düşünmelisiniz. Bu, DPIA'nın düzenli olarak gözden geçirilmesi ve uygun olduğunda yeniden değerlendirme yapılması anlamına gelir (örneğin, işlemenin niteliği, kapsamı, durumu veya amacı ile kişilere yönelik riskler vb etkenler herhangi bir nedenle değişirse).

Örneğin, uygulamaya bağlı olarak, hedef popülasyonun demografik bilgileri değişebilir veya insanlar işlemenin kendisine yanıt olarak davranışlarını zaman içinde ayarlayabilir.

Mevzuattaki ilgili hükümler

Bknz. GDPR - 35(11), 36(1) and 39(1)(c) maddeleri ve 84 gerekçe (dış bağlantı)

İlave okuma – ICO kılavuzu

DPIA'ların yasal olarak gerekli olduğu yüksek riskle sonuçlanması muhtemel işleme operasyonları listesi de dahil olmak üzere, GDPR Kılavuzu'ndaki DPIA'lara ilişikin kılavuzumuzuokuyun.

Ayrıca, yukarıda açıklanan her adım da dahil olmak üzere, bir DPIA'nın nasıl yapılacağına ilişkin ayrıntılı kılavuzumuzu da okumalısınız.

Ayrıca Kılavuzun aşağıdakilerle ilgili bölümlerini de okumak isteyebilirsiniz:

- Yasallık, adalet ve şeffaflık
- İşleme için yasal esaslar
- Veri minimizasyonu
- Doğruluk

İlave okuma – Avrupa Veri Koruma Kurulu

29. Madde Çalışma Grubunun (WP29) yerini alan Avrupa Veri Koruma Kurulu (EDPB - The European Data Protection Board), her AB üye devletinin veri koruma yetkililerinden temsilciler içerir.

GDPR gerekliliklerine uymak için yönergeleri benimser.WP29, veri koruma etki değerlendirmeleri kılavuzu (WP248 rev.01) üretti ve EDPB tarafından onaylandı.

Diğer ilgili yönergeler şunları içerir:

- Veri Koruma Görevlileri ('DPO'lar') hakkında yönergeler (WP243 rev.01) (Guidelines on Data Protection Officers ('DPOs') (WP243 rev.01))
- Otomatik kişisel karar verme ve profil oluşturmaya ilişkin yönergeler (Guidelines on automated individual decision-making and profiling (WP251 rev.01))

Yapay zekada denetleyici/işlemci ilişkilerini nasıl anlamalıyız?

Yapay zeka sistemleri için denetcilik neden önemlidir?

Çoğu durumda, yapay zekaya dahil olan çeşitli işleme operasyonları bir dizi farklı organizasyon tarafından gerçekleştirilebilir. Bu nedenle, AI kullanımınız birden fazla organizasyonu içeriyorsa kimin denetleyici, kimin ortak denetleyici veya kimin işleyici olduğunu belirlemeniz çok önemlidir.

Bu ilişkileri anlamanın ilk adımı, farklı işlem operasyon gruplarını ve bunların amaçlarını belirlemektir (bkz. (bknz. 'DPIA'mızda neler değerlendirmeliyiz?'). Bunların her biri için siz ve birlikte çalıştığınız organizasyonlar; bir denetleyici, işleyici veya ortak denetleyici olup olmadığınızı değerlendirmelisiniz. Bu değerlendirmede yardım için denetçi/işleyici hakkındaki rehberimize başvurmalısınız, ancak özünde:

- İşleme amaç ve yöntemlerine karar verirseniz, bir denetçi olursunuz.
- Kişisel verileri başka bir organizasyonun talimatı ile işliyorsanız, işleyicisiniz.
- Başka bir organizasyonla işleme amaç ve yöntemlerini birlikte belirlerseniz, ortak denetçi olursunuz.

AI genellikle birkaç farklı aşamada kişisel verilerin işlenmesini içerdiğinden, bazı aşamalar için bir denetçi veya ortak denetçi, diğerleri için bir işleyici olmanız mümkündür.

İşleme araçlarına denetçiler tarafından karar verilirken; işleyiciler, bu araçların *temel olmayan* bazı ayrıntılarına karar verme konusunda biraz takdir yetkisine sahip olabilir. Her durum hakkında genelleme yapmak mümkün olmasa da; aşağıdaki örnekler sizin bir denetçi olduğunuzu gösterebilecek AI ile ilgili işleme yöntemleri hakkındaki karar türlerini ve bir işleyici tarafından alınabilecek temel olmayan detaylar hakkındaki kararları göstermektedir.

Ne tür kararlar bizi denetçi yapabilir?

Sizi denetçi yapma olasılığı bulunan karar türleri arasında aşağıdakileri bulunmaktadır:

- İlk sırada kişisel verileri toplamak.
- Bu işlemenin amacı.
- Hangi kişiler hakkında veri toplanacak ve onlara işlem hakkında ne söylenecek.
- Verileri ne kadar süreyle saklayacağınız.
- Kişi haklarına uygun olarak yapılan taleplere nasıl cevap verileceği.

Sizi denetleyici yapan durumlarla ilgili daha fazla ayrıntı için denetçiler ve işleyiciler hakkındaki kılavuzumuzu okuyun. Yapay zeka bağlamında, denetçiler tarafından verilen karar türleri şunları içerebilir:

- Bir AI modelini eğitmek için kullanılan verilerin kaynağı ve niteliği.
- Modelin hedef çıktısı (yani tahmin edilen veya sınıflandırılan şey).
- Verilerden modeller oluşturmak için kullanılacak çok çeşitli ML algoritmaları (örn. regresyon modelleri – regression models, karar ağaçları – decision tree, rastgele ormanlar – random forest, sinir ağları – neural networks).
- Özellik seçimi – her modelde kullanılabilecek özellikler.
- Anahtar model parametreleri (örneğin, bir karar ağacının ne kadar karmaşık olabileceği veya bir grubun içinde kaç model olacağı).
- Yanlış pozitifler ve yanlış negatifler arasındaki ödünleşme gibi temel değerlendirme ölçütleri ve kayıp fonksiyonları.
- Herhangi bir modelin nasıl sürekli olarak test edilip güncelleneceği: ne sıklıkta, ne tür veriler kullanılarak ve devam eden performansın nasıl değerlendirileceği.

Yukarıdaki liste, kapsamlı olmamakla birlikte, amaç ve yöntemlerle ilgili kararları oluşturmaktadır. Bu kararlardan herhangi birini verirseniz, muhtemelen bir denetçi olursunuz; siz ve başka bir organizasyon bunları ortaklaşa belirlerseniz, ortak denetçisiniz.

İşleyiciler hangi kararları alabilir?

Yukarıdaki kararlardan herhangi birini almazsanız ve verileri yalnızca üzerinde anlaşmaya varılan şartlar temelinde işlerseniz, bir işleyici olmanız daha olasıdır.

İşleyicilerin farklı yükümlülükleri vardır ve denetçinizle aranızdaki herhangi bir sözleşmenin koşullarına bağlı olarak aşağıdakiler gibi belirli kararlar alabilirsiniz:

- Kişisel verileri işlemek için kullandığınız BT sistemleri ve yöntemleri
- Verileri nasıl sakladığınız
- Onu koruyacak güvenlik önlemleri
- Bu verileri nasıl aldığınız, aktardığınız, sildiğiniz veya imha ettiğiniz.

Bir AI bağlamında, işleyiciler (sözleşmelerinin koşullarına bağlı olarak) aşağıdakilere karar verebilir:

- Yazıldıkları programlama dili ve kod kütüphaneleri gibi genel ML algoritmalarının özel uygulaması.
- Veri ve modellerin nasıl depolandığı, örneğin serileştirildikleri ve depolandıkları formatlar ve lokalde önbelleğe alma.
- Bilgi işlem kaynaklarının tüketimini en aza indirmek için öğrenme algoritmalarını ve modellerini optimize etmeye yönelik önlemler. (örn. onları paralel süreçler olarak uygulayarak)
- Sanal makinelerin, mikro hizmetlerin, API'lerin seçimi gibi modellerin nasıl uygulamaya alınacağına ilişkin mimari ayrıntılar.

Sadece bu tür kararlar alıyorsanız, bir denetçiden çok bir işleyici olabilirsiniz.

Pratikte bu, bir denetçi olarak görülmeden çeşitli AI hizmetleri sağlayabileceğiniz anlamına gelir. Ancak önemli olan, hangi verilerin hangi amaçlarla işlendiğine ilişkin kapsayıcı kararların yalnızca bir denetçi tarafından alınabileceğini hatırlamaktır.

AI geliştirme araçları sağlama

Örneğin, işlem ve depolamaya sahip özel bir bulut bilişim ortamından ve makine öğrenimi için bir dizi ortak araçtan oluşan bulut tabanlı bir hizmet sağlayabilirsiniz. Bu hizmetler, müşterilerinizin işleme için seçtikleri verilerle; sadece onlara bulutta sağladığınız araçları ve altyapıyı kullanarak kendi modellerini oluşturmalarını ve çalıştırmalarını sağlar.

Bu örnekte, siz bir işleyici iken müşterilerin denetçi olma olasılığı daha yüksektir.

Bu nedenle, bir işleyici olarak, bu araçların hangi programlama dillerinde ve kod kütüphanelerinde yazılacağına, depolama çözümlerinin konfigürasyonuna, grafiksel kullanıcı arayüzüne ve bulut mimarisine karar verebilirsiniz.

Müşterileriniz, kullanmak istedikleri veri ve modelleri, temel model parametreleri ve bu modelleri değerlendirme, test etme ve güncelleme süreçleri hakkında kapsamlı kararlar aldıkları sürece denetçidir.

Müşterilerinizin verilerini onlar tarafından kararlaştırılan amaçlar dışında kullanırsanız, bu tür işlemler için bir denetçi olursunuz.

Bir hizmet olarak AI tahmini sağlama

Bazı şirketler, müşterilere canlı AI tahmini ve sınıflandırma hizmetleri sunar. Örneğin, kendi AI modellerinizi geliştirebilir ve müşterilerin bir API aracılığıyla onlara sorgu göndermesine (örn: 'bu görüntüde hangi nesneler var?') ve yanıtlar almasına (örn: görüntüdeki nesnelerin sınıflandırılması) izin verebilirsiniz..

Bu durumda, bir AI tahmin hizmet sağlayıcısı olarak, işlemelerin en azından bir kısmı için bir denetçi olmanız muhtemeldir.

İlk olarak, hizmetlerinizi güçlendiren modelleri oluşturmak ve geliştirmek için gerekli işlemler vardır; bu işlemlerin araçları ve amaçlarına öncelikle hizmet sağlayıcı olarak sizin tarafınızdan karar verilirse, bu tür işlemler için denetçi olmanız muhtemeldir. İkincisi, müşterileriniz adına belirli örneklerle ilgili tahmin ve sınıflandırmaları yapmak için gerekli olan işlemler; bu tür bir işlemler için, işleyici olarak müşterinin sizinle birlikte denetçi olması daha olasıdır.

Bununla birlikte, hizmetlerinizi, müşterilerin tahminde yer alan işlemenin temel unsurları ve amaçları üzerinde yeterli etkiye sahip olamayacak şekilde tasarlarsanız ikinci tür işleme için bir denetçi veya ortak denetçi olarak bile kabul edilebilirsiniz. Örneğin, müşterilerinizin yanlış pozitifler ve negatifler arasındaki dengeyi kendi gereksinimlerine göre ayarlamasına veya bir modele belirli özellikler eklemesine veya bir modele belirli özellikleri kaldırmasına izin vermiyorsanız, pratikte ortak bir denetlçi olabilirsiniz. Bunun nedeni, temel işleme araçları üzerinde yeterince güçlü bir etki uygulamanızdır.

Ayrıca, başlangıçta bir müşteri adına bir hizmet sağlamanın bir parçası olarak verileri işliyorsanız, ancak daha sonra müşterilerinizden gelen aynı verileri kendi modellerinizi geliştirmek için işliyorsanız, o zaman bu işleme için bir denetçi olursunuz. Bunun nedeni, bunu kendi amaçlarınız için üstlenmeye karar vermiş olmanızdır. Örneğin, bir işe alım danışmanı, her başvuru sahibi için puanlar sağlayan bir ML sistemi kullanarak iş başvurusunda bulunanların verilerini müşteriler adına işleyebilir ve aynı zamanda bu verileri ML sistemlerini daha da geliştirmek için saklayabilir.

Bu gibi durumlarda, bu amaçları en baştan açıkça belirtmeli ve kendi yasal dayanağınızı çıkarmalısınız. Daha fazla bilgi için amaçlar ve yasal dayanaklar bölümüne bakın. (AI'yı kullanırken amaçlarımızı ve yasal dayanağımızı nasıl belirleriz?)

Ek olarak, veri koruma mevzuatı; denetçinin veya işleyicinin yetkisi altında hareket eden hiç kimsenin, kanunen gerekli olmadıkça, denetçinin talimatları dışında kişisel verileri işleyemeyeceğini belirtir. Bunun sizin özel durumunuzda nasıl geçerli olduğunu düşünmeniz gerekir. Örneğin, bir işlemcinin yalnızca denetçinin talimatlarına göre işlem yapmasına izin veriliyor – bu yüzden, yalnızca işleyici rolü nedeniyle kişisel verilerin daha fazla işlenmesi durumunda, bu işlem,

adil ve uyumlu amaçlara sahip olsa bile GDPR'yi yine de ihlal edebilir. Asıl denetçinin, bir üçüncü taraf denetlçi olarak, yani bir veri paylaşım işleminde (kendisinin veri koruma yasasına uyması gereken) bu verileri size açmayı kabul etmesi gerekir.

Mevzuattaki ilgili hükümler

Bknz. GDPR –madde 29 (dış bağlantı)

Bknz. DPA 2018 bölüm 60 ve 106 (dış bağlantı).

Lokal kullanım için yapay zeka modellerini tedarik ederken sorumluluklarımız nelerdir?

Bazı şirketler, önceden eğitilmiş AI modellerini müşterilerinin kurup çalıştırabileceği bağımsız yazılım parçaları olarak satar (veya ücretsiz olarak sağlar).

Sizinki gibi organizasyonların satın alacağı sağlam bir AI sistemi geliştirmek ve pazarlamak için, bu üçüncü tarafların kişisel verileri bir şekilde (ör. sistemi eğitmek) işlemesi olasıdır. Ancak, sistemi işlem ortamınıza entegre ettiğinizde; geliştirici, işlemin kendisinde başka bir rol oynamayabilir.

Dolayısıyla, bu sağlayıcılar sattıkları modelleri eğitmek amacıyla kişisel verilerin işlenmesine ilişkin kararlar aldıklarında, bu tür işlemler için bir denetçi konumundadırlar.

Bununla birlikte, yapay zeka sistemlerinin üçüncü taraf geliştiricileri, modellerini satın aldıktan sonra kişisel verileri kendi haklarına (denetlçiler olarak) veya sizin adınıza (işleyiciler olarak) işleme niyetinde olmayabilir. Modeli devreye aldığınızda, sağlayıcı kişisel verileri işlemezse, o modeli kullanarak üstlendiğiniz herhangi bir işlemin denetçisi siz olursunuz.

Bu tür yapay zeka hizmetlerinin sağlayıcısıysanız, hangi işleme operasyonlarının denetçisi veya işleyicisi olduğunuzu belirlemeli ve bunun müşterilerinize açık olduğundan emin olmalısınız. Bu hizmetleri alıyorsanız, farklı sorumluluklarınız vardır. Kişisel verileri işlemek için bir yapay zeka sistemi seçerken dikkat etmeniz gereken hususlar:

- İster bir sistem satın alın, ister kendi sisteminizi kurun, veri koruma yasasına uymanın sizin sorumluluğunuzda olduğunu unutmayın.
- Veri koruma ilkelerine uygun olarak yapıp yapmadıkları da dahil olmak üzere geliştiricinin yapay zeka sistemini nasıl tasarladığını kontrol edin.
- Sonuç olarak, bu hizmeti kullanmanın hem işleme hedeflerinizi hem de veri koruma yükümlülüklerinizi yerine getirmenize yardımcı olup olmayacağını değerlendirin.

Bir denetleyici olarak, sözleşmeler ve hizmet seviyesi anlaşmaları yazarken, sözleşme sizin denetçi veya işleyici olarak statünüzü şart koşuyor olsa da, veri koruma perspektifinden önemli olan; **pratikte,** işleme amaçlarına ve temel araçlara kimin karar verdiğidir.

Benzer şekilde, bu tür AI hizmetlerinin bir sağlayıcısıysanız, hangi işlem operasyonlarının denetçsi veya işlemcisi olduğunuzu belirlemeli ve bunun müşterilerinize açık olduğundan emin olmalısınız.

Mevzuattaki ilgili hükümler

Bknz. GDPR'ın 4(7), 4(8), 5(1), 5(2), 25, 28 maddeleri and 78 gerekçesi (dış bağlantı)

İlave okuma – ICO kılavuzu

GDPR Kılavuzundaki denetleyici/işleyici ve sözleşmeler/yükümlülükler hakkındaki kılavuzumuzu okuyun.

Diğer kaynaklar

Bknz. Mahkeme kararı - Court of Justice of the European Union's (CJEU) judgment in the case of Unabhängiges Landeszentrum für Datenschutz (ULD) Schleswig-Holstein against Wirtschaftsakademie Schleswig-Holstein GmbH.

Ayrıca bkz. Davaya ilişkin karar Fashion ID GmbH & Co. KG against Verbraucherzentrale NRW eV.

Yapay zeka ile ilgili ödünleşmeler nelerdir ve bunları nasıl yönetmeliyiz?

AI kullanımınız, veri koruma yasasının gerekliliklerine uygun olmalıdır. Bununla birlikte, zaman zaman farklı yönlere çekilebilecek bir dizi farklı değerler ve çıkarlar olabilir. Veri koruma yasasının risk temelli yaklaşımı, bir yanda mahremiyet ile diğer yanda yarışan değerler ve çıkarlar arasındaki potansiyel 'ödünleşmeleri' yönlendirmenize yardımcı olabilir.

Bu nedenle, eğer AI kullanıyorsanız, bu çıkarları belirlemeli ve değerlendirmeli ve yasa kapsamındaki yükümlülüklerinizi yerine getirmeye devam ederken durumunuzu göz önüne alarak bunlar arasında uygun bir denge kurmalısınız.

Herhangi bir özel ödünleşmede doğru denge, faaliyet gösterdiğiniz belirli sektörel ve sosyal durum ile kişiler üzerindeki etkiye bağlıdır. Ancak, birçok kullanım durumuyla ilgili olan ödünleşmeleri değerlendirmek ve azaltmak için kullanabileceğiniz yöntemler vardır.

Aşağıdaki bölümler, AI sistemlerini tasarlarken veya tedarik ederken karşılaşabileceğiniz en dikkate değer ödünleşmelerden bazılarına kısa bir genel bakış sunmaktadır.

Gizlilik ve istatistiksel doğruluk

Bir veri koruma hususunda adalet; genel olarak insanların makul olarak bekleyeceği şekillerde kişisel verileri işlemeniz ve bu verileri onlar üzerinde haksız olumsuz etkileri olacak şekilde kullanmamanız anlamına gelir. AI sisteminizin çıktılarının "istatistiksel doğruluğunu" iyileştirmek, adalet ilkesine uygunluğu sağlamak için göz önünde bulundurmanız gereken hususlardan biridir.

'Doğruluk' kelimesinin veri koruma ve yapay zeka bağlamında farklı bir anlamı olduğunu belirtmek önemlidir. Veri korumada doğruluk, kişisel verilerin doğru olduğundan ve gerektiğinde güncel tutulduğundan emin olmanızı gerektiren temel ilkelerden biridir. AI'da (ve daha genel olarak istatistiksel modellemede) doğruluk, bir AI sisteminin doğru cevabı ne sıklıkla tahmin ettiğini ifade eder ve çoğu durumda bu cevaplar kişisel veriler olacaktır.

"**Doğruluk ilkesi**" işlediğiniz kişisel veriler için geçerli olsa da, bir AI sisteminin bu ilkeye uyması için %100 "**istatistiksel olarak doğru**" olması gerekmez. Bununla birlikte, bir AI sistemi istatistiksel olarak ne kadar doğruysa, işleminizin adalet ilkesine uygun olması o kadar olasıdır.

Bu nedenle, net olması için bu kılavuzda:

- 'Doğruluk' terimini veri koruma yasasının doğruluk ilkesine atıfta bulunmak için kullanıyoruz ve
- Bir yapay zeka sisteminin doğruluğuna atıfta bulunmak için "istatistiksel doğruluk" terimini kullanıyoruz.

"Veri koruma" doğruluğu ile istatistiksel doğruluk arasındaki farklar hakkında daha fazla bilgi için "İstatistiksel doğruluk hakkında ne yapmamız gerekiyor?" bölümüne bakın.

Genel olarak, bir AI sistemi verilerden öğrendiğinde (ML modellerinde olduğu gibi), ne kadar çok veri üzerinde eğitilirse, istatistiksel olarak o kadar doğru olur. Diğer bir deyişle, veri setlerindeki özellikler arasındaki temel ve istatistiksel olarak yararlı ilişkileri yakalama olasılığı o kadar yüksektir.

Örneğin, müşterilerin satın alma geçmişine dayalı olarak gelecekteki satın almaları tahmin etmeye yönelik bir model, eğitim verilerine ne kadar çok müşteri dahil edilirse istatistiksel olarak daha doğru olma eğiliminde olacaktır. Ve mevcut

bir veri setine eklenen herhangi bir yeni özellik, modelin tahmin etmeye çalıştığı şeyle ilgili olabilir; örneğin, ek nufüs istatistikleri verileri ile zenginleştirilmiş satın alma geçmişleri, modelin istatistiksel doğruluğunu daha da iyileştirebilir.

Bununla birlikte, genel olarak konuşursak, her bir kişi hakkında ne kadar çok veri noktası toplanırsa ve verileri veri setine dahil edilen kişi sayısı ne kadar çok olursa, bu kişilere yönelik riskler o kadar büyük olur.

İlave okuma – ICO kılavuzu

GDPR Kılavuzundaki veri minimizasyonu hakkındaki kılavuzumuzu okuyun.

İstatistiksel doğruluk ve ayrımcılık

'Yanlılık (önyargı) ve ayrımcılık risklerini nasıl ele almalıyız?' bölümünde tartışıldığı gibi, yapay zeka sistemlerinin kullanımı yanlı veya ayrımcı sonuçlara yol açabilir. Bunlar da adalet ilkesi açısından uyum riskleri oluşturabilir. Bu riski azaltmak için uygun teknik önlemleri uygulamanız gerekir.

Değerlendirmeleriniz; bu tekniklerin yapay zeka sisteminin performansının istatistiksel doğruluğu üzerindeki etkisini de içermelidir. Örneğin, ayrımcılık potansiyelini azaltmak için, bir kredi riski modelini, farklı korunan özelliklere (örneğin erkekler ve kadınlar) sahip kişiler arasındaki olumlu tahminlerin oranı eşitlenecek şekilde değiştirebilirsiniz. Bu, ayrımcı sonuçların önlenmesine yardımcı olabilir, ancak genel olarak yönetmeniz gereken daha fazla sayıda istatistiksel hatayla da sonuçlanabilir.

Uygulamada, istatistiksel doğruluk ile ayrımcılığı önlemek arasında her zaman bir sıkıntı olmayabilir. Örneğin, modeldeki ayrımcı sonuçlar bir azınlık nüfusu hakkında ilgili bir veri eksikliğinden kaynaklanıyorsa, o zaman modelin istatistiksel doğruluğu, onlar hakkında daha fazla veri toplanarak ve aynı zamanda doğru tahminlerin oranları eşitlenerek artırılabilir.

Ancak, bu durumda, farklı bir seçenekle karşı karşıya kalırsınız - Azınlık popülasyon hakkında, karşılaştıkları orantısız sayıdaki istatistiksel hataları azaltmak adına daha fazla veri toplamak veya bu kişilerin diğer hak ve özgürlüklerine yönelik riskler nedeniyle bu tür verileri toplamamak arasında.

Açıklanabilirlik ve istatistiksel doğruluk

Adaletle ilgili değerlendirmelerinizin bir kısmı, yapay zeka sistemlerinin açıklanabilirliği ve istatistiksel doğruluğu arasında bir dengeyi de içerir.

Derin öğrenmeye dayalı olanlar gibi çok karmaşık sistemler için sistemin mantığını takip etmek zor olabilir ve bu nedenle nasıl çalıştıklarını yeterince açıklamanız zor olabilir. Bu bazen 'kara kutu sorunu' olarak nitelendirilir.

Koşullara bağlı olarak, özellikle görüntü tanıma gibi sorunlar söz konusu olduğunda, bu karmaşık sistemleri istatistiksel olarak en doğru ve etkili sistemler olarak görebilirsiniz. Bu nedenle, açıklanabilirlik ve istatistiksel doğruluk arasında bir denge ile karşı karşıya olduğunuzu düşünebilirsiniz.

Bununla birlikte, daha basit modellerin iyi performans gösterdiği birçok uygulama için, açıklanabilirlik ve istatistiksel doğruluk arasındaki dengeler aslında nispeten küçük olabilir. Bu sorunlar ExplAIn proje kılavuzunda daha derinlemesine ele alınmaktadır, ancak genel bir nokta olarak, "kara kutu- black box" modellerini yalnızca şu durumlarda kullanmalısınız:

- Olası riskleri ve etkilerini önceden iyice değerlendirdiniz, ekip üyelerinizin; kullanım gerekliliklerinizi ve organizasyonel kapasite/kaynaklarınızın bu sistemlerin tasarım ve uygulamasını desteklediğini belirlediğinde.
- Sisteminiz, etki alanına uygun seviyede açıklanabilirlik sağlayan ek yorumlanabilirlik araçları içerdiğinde.

Açıklanabilirlik, kişisel verilerin ifşa edilmesi ve ticari güvenlik

Kişilere yapay zekaya dayalı bir kararın mantığı hakkında anlamlı bilgiler sağlamak süreç içinde kişisel veriler ve sistemin özel mantığı gibi diğer bilgiler de dahil olmak üzere gizli tutmak için ihtiyaç duyduğunuz bilgilerin farkında olmadan ifşa edilmesi riskini potansiyel olarak artırabilir.

Son araştırmalar; makine öğrenimi modellerini açıklanabilir kılmak için önerilen bazı yöntemlerin, modeli eğitmek için verilerini kullandığınız kişiler hakkında kişisel veriler bulmadan nasıl kolaylaştırabileceğini göstermiştir. ('Model inversiyon-ters çevirme saldırıları (inversion attack) nedir?' ve 'Üyelik çıkarım saldırıları (inference attacks) nelerdir?' bölümlerine bakın).

Bazı araştırmalar, kişilere bir açıklama sağlarken yanlışlıkla bir AI modelinin nasıl çalıştığına dair özel bilgileri ifşa etme riskini de vurgulamaktadır. Ancak, ticari çıkarları veri koruma gerekliliklerileyle (örneğin ticari güvenlik ve veri koruma güvenliği) karıştırmamaya dikkat etmeniz ve bunun yerine böyle bir ödünleşmenin gerçekten ne ölçüde var olduğunu düşünmelisiniz.

Şimdiye kadar yaptığımız araştırma ve paydaş katılımı bu riskin oldukça düşük olduğunu gösteriyor. Ancak en azından teoride, kişilerin açıklama alma hakkı ve işletmelerin ticari sırları koruma önemlerini göz önünde bulundurmanız gereken durumlar olabilir, ayrıca veri koruma uyumluluğunun 'ödünleşme yapılamadığına' de dikkat edin.

Bu risklerin her ikisi de aktif araştırma alanlarıdır, olasılıkları ve şiddeti tartışma ve inceleme konusudur. Bu riskleri izlemeye ve incelemeye devam edeceğiz ve bu kılavuzu buna göre güncelleyebiliriz.

Bu ödünleşmeleri (trade-off) nasıl yönetebiliriz?

Çoğu durumda, birden fazla ödünleşme arasında doğru dengeyi sağlamak; kullanım durumuna ve AI sisteminin devreye alınma şartlarına özel bir karar meselesidir.

Yaptığınız seçimler ne olursa olsun, onlardan sorumlu hesap verebilir olmanız gerekir. Çabalarınız, kişilere yaymayı düşündüğünüz AI sisteminin riskleriyle orantılı olmalıdır. Yapmanız gerekenler:

- Bir yapay zeka sistemi tasarlarken veya tedarik ederken mevcut veya potansiyel ödünleşmeleri belirleyip değerlendirin ve bunun kişiler üzerindeki etkisini değerlendirin.
- Herhangi bir ödünleşme ihtiyacını en aza indirmek için mevcut teknik yaklaşımları göz önünde bulundurun.
- Makul bir yatırım ve çaba ile uygulayabileceğiniz teknikleri göz önünde bulundurun.
- Nihai ödünleşme kararları hakkında net kriterlere ve hesap verebilirlik sınırlarına sahip olun. Bu, sağlam, riske dayalı ve bağımsız bir onay sürecini içermelidir.
- Uygun olduğunda, kişilere veya AI çıktılarını gözden geçirmekle görevli herhangi bir kişiye herhangi bir ödünleşmeyi açıklamak için tedbirler alın.
- Diğer şeylerin yanı sıra, bunları azaltmak için, kişilerin (veya temsilcilerinin) görüşlerini ve gelişen teknikleri veya en iyi uygulamaları dikkate alarak ödünleşmeleri düzenli olarak gözden geçirin.

Bu süreçleri ve bunların sonuçlarını denetlenebilir bir standartta belgelemelisiniz ve bunları uygun düzeyde ayrıntılarıyla DPIA'nıza kaydetmelisiniz.

Ayrıca şunları da belgelemelisiniz:

- Kişisel verileri işlenen bireylere yönelik riskleri nasıl hesaba kattığınız.
- Kapsamdaki ödünleşmeleri belirleme ve değerlendirme metodolojisi, belirli teknik yaklaşımları benimseme veya reddetme nedenleri (eğer uygunsa).
- Nihai kararınızın önceliklendirme kriterleri ve gerekçesi.
- Nihai kararın genel risk iştahınıza nasıl uyduğu.

İki veya daha fazla veri koruma gereksinimi arasında uygun bir ödünleşme sağlamak mümkün değilse, herhangi bir AI sisteminin devreye alımını durdurmaya da hazır olmalısınız.

Dış kaynak kullanımı ve üçüncü taraf yapay zeka sistemleri

Üçüncü taraftan bir yapay zeka çözümü satın aldığınızda veya tamamen dışarıdan tedarik ettiğinizde, durum tespiti sürecinizin bir parçası olarak herhangi bir ödünleşmenin bağımsız bir değerlendirmesini yapmanız gerekir. Ayrıca, sonradan ödünleşmeleri ele almak yerine, tedarik aşamasında gereksinimlerinizi belirtmeniz gerekir.

GDPR'nin 78. gerekçesi, yapay zeka çözüm üreticilerinin aşağıdakileri yapması için teşvik edilmesi gerektiğini söylüyor:

- Sistemlerini geliştirirken ve tasarlarken veri koruma hakkını dikkate almak; ve
- Denetçilerin ve işleyicilerin veri koruma yükümlülüklerini yerine getirebildiğinden emin olun.

Doğrudan veya üçüncü taraf sağlayıcı aracılığıyla dağıtım öncesinde ve esnasında sistemleri değiştirme yetkiniz olduğundan emin olmalısınız, böylece başlangıçta ve yeni riskler ve düşünceler ortaya çıktıkça sizin uygun ödünleşmelar olarak düşündüğünüz şeylerle uyumlu olacak şekilde değiştirebilirsiniz.

Örneğin, bir satıcı, gelecek vaat eden iş adaylarını etkin bir şekilde puanlayan ancak değerlendirmeyi yapmak için her aday hakkında görünüşte çok fazla bilgi gerektirebilecek bir CV tarama aracı sunabilir. Böyle bir sistem satın alıyorsanız, adaylardan bu kadar çok kişisel veri toplamayı haklı gösterip gösteremeyeceğinizi düşünmeniz ve yoksa sağlayıcıdan sistemlerini değiştirmesini talep etmeniz veya başka bir sağlayıcı aramanız gerekebilir. Veri minimizasyonu bölümüne bakın.

Kültür, çeşitlilik ve paydaşlarla etkileşim

Uygun ödünleşmeleri belirlerken önemli değerlendirme çağrıları yapmanız gerekir. Etkili risk yönetimi süreçleri gerekli olmakla birlikte, organizasyonunuzun kültürü de temel bir rol oynar.

Bu tür bir alıştırmayı üstlenmek, organizasyon içindeki farklı ekipler arasında işbirliğini gerektirecektir. Çeşitliliğin olduğu, işbirliği içinde çalışmak için teşviklerin olduğu ve ayrıca personelin endişelerini dile getirebildiği ve alternatif yaklaşımlar önermek için cesaretlendirildiği bir ortam önemlidir.

AI'nın farklı durumlarda sosyal olarak kabul edilebilirliği ve ödünleşmelerle ilgili en iyi uygulamalar, devam eden toplumsal tartışmaların konusudur. Ödünleşmeden etkilenenler de dahil olmak üzere organizasyonunuz dışındaki paydaşlarla istişare, farklı kriterlere vermeniz gereken değeri anlamanıza yardımcı olabilir.

Ödünleşmelerin değerlendirilmesi: çalışılmış bir örnek

Çoğu durumda ödünleşmeler kesin olarak ölçülebilir değildir, ancak bu keyfi kararlara yol açmamalıdır. Belirli yapay zeka kullanım durumları için farklı gereksinimlerin nispi değeri hakkındaki varsayımlarınızı belgeleyerek ve gerekçelendirerek içeriksel değerlendirmeler yapmalısınız.

Mümkün olan en iyi ödünleşmeyi belirlemenize yardımcı olacak olası bir yaklaşım, farklı tasarımlar arasındaki seçimleri bir grafik üzerinde görsel olarak temsil etmektir.

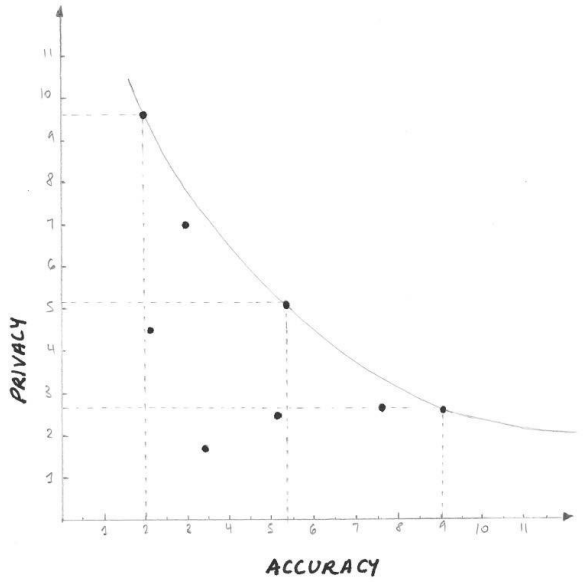
X ve Y ekseninde iki kriter dengelenerek (örneğin istatistiksel doğruluk ve gizlilik) bir sistemin nasıl tasarlanabileceğine ilişkin olası seçenekleri grafik üzerinde çizebilirsiniz. İstatistiksel doğruluk, 'İstatistiksel doğrulukla ilgili ne yapmamız gerekiyor?' bölümünde açıklanan şekillerde kesin bir ölçüm verilebilir.

Gizlilik önlemlerinin doğası gereği daha az net ve daha belirleyici nitelikte olması muhtemeldir, ancak şunları içerebilir:

- Gerekli kişisel veri miktarı
- Bu verilerin hassasiyeti
- Kişiyi ne ölçüde benzersiz (eşsiz) bir şekilde tanımlayabileceği
- İşlemin niteliği, kapsamı, durumu ve amaçları
- Veri işlemenin kişilere sunabileceği hak ve özgürlüklere yönelik riskler
- AI sistemlerinin uygulanacağı kişi sayısı.

Bir grafik, 'üretim-olasılık sınırı' olarak bilinen şeyi ortaya çıkarabilir ve karar vericilerin; sistem tasarım kararlarının, farklı değerler arasındaki dengeyi nasıl etkileyebileceğini anlamalarına yardımcı olmanın bir yoludur.

Gizlilik ve istatistiksel doğruluk arasındaki bir ödünleşmenin nasıl görünebileceğini görselleştirmek için Şekil 1'de bu yöntemi kullandık. Bu, ödünleşmeleri anlamasına yardımcı olmak için belirli bir sistemin seçimini onaylamaktan sorumlu kıdemli bir karar vericiye sunulabilir.

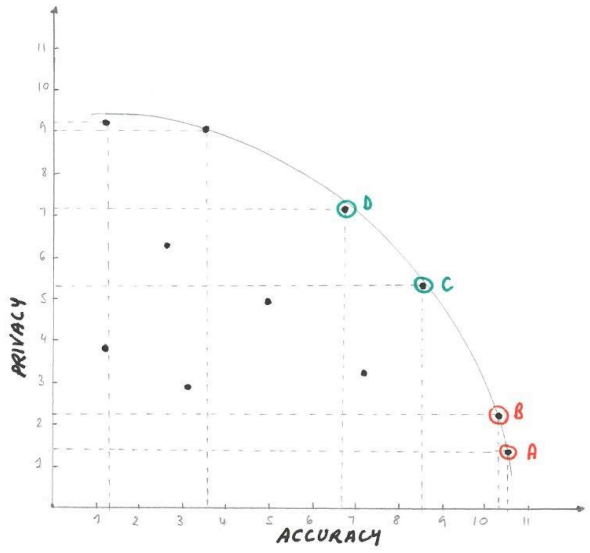


Şekil 1

Grafikteki veri noktaları, farklı tasarım seçeneklerinden, ML modellerinden ve kullanılan veri miktarı ve türlerinden kaynaklanan bir AI sisteminin farklı olası teknik konfigürasyonlarını temsil eder. Her nokta, istatistiksel doğruluk ve gizlilik arasındaki belirli bir ödünleşmenin olası bir sonucunu temsil eder..

Şekil 1'deki senaryoda, önerilen sistemlerden hiçbiri hem yüksek istatistiksel doğruluk hem de yüksek gizlilik elde edemez ve gizlilik ile istatistiksel doğruluk arasındaki ödünleşme önemlidir.

Farklı bir kullanım durumu; Şekil 2'de görselleştirildiği gibi, ödünleşmeler çok farklı görünebilir.

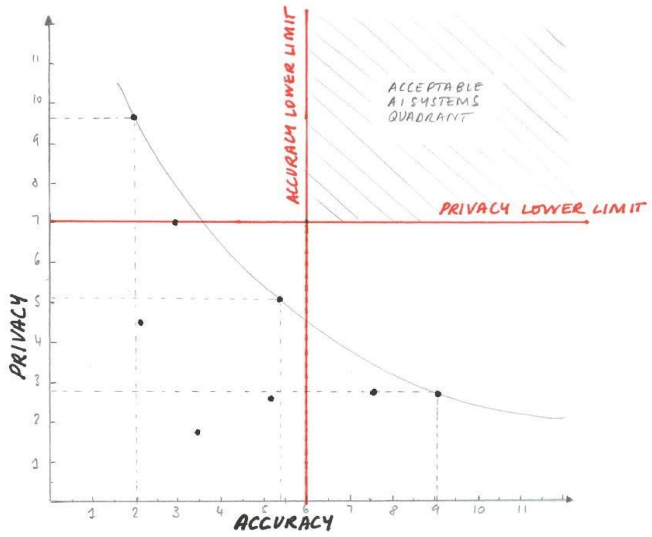


Şekil 2

Bu senaryoda, istatistiksel doğruluk ve gizlilik arasında makul bir denge elde etmek daha kolay olabilir. Grafik ayrıca, eğrinin ortasındaki (C ve D) AI sistemleri için gizlilikten veya istatistiksel doğruluktan ödün vermenin maliyetinin, kenardakilere (A ve B) göre daha düşük olduğunu göstermektedir.

Bu örnekte, sağ altta olası sistemler için istatistiksel doğrulukta azalan getiriler vardır. A sistemini B yerine seçmeyi düşünüyorsanız, bu koşullar altında istatistiksel doğruluğa daha yüksek bir değer vermenin neden garanti edildiğini gerekçelendirmeniz gerekir. Kişilerin hak ve özgürlükleri üzerindeki etkisini hesaba katmanız ve işlemin genel olarak hala adil ve orantılı olduğunu gösterebilmeniz gerekir.

Ödünleşmelerin görsel temsili, altına gitmek istemediğiniz her iki değişken için de alt limitler içerebilir (Şekil 3).



Şekil 3

Şekil 3'teki senaryoda, hem istatistiksel doğruluk hem de gizlilik için alt sınırları karşılayan olası bir sistem yoktur, bu da bu sistemlerden herhangi birinin dağıtımını takip etmemenizi önerir. Bu, diğer yöntemlere ve veri kaynaklarına bakmak, sorunu yeniden formüle etmek veya sorunu çözmek için yapay zekayı kullanma girişiminden vazgeçmek anlamına gelebilir.

Ödünleşmeleri en aza indirmek için matematiksel yaklaşımlara ne dersiniz?

Bazı durumlarda, ödünleşmelerin unsurlarını tam olarak ölçebilirsiniz. 'Sınırlı optimizasyon' olarak bilinen bir dizi matematik ve bilgisayar bilimi tekniği, bu tür ödünleşmeleri en aza indirmek için en uygun çözümleri bulmayı amaçlar. Makine öğrenimi mühendisi gibi bir teknik uzman, bu tekniklerin kendi özel durumunuzda uygulanabilirliğini değerlendirmelidir.

Örneğin, diferansiyel gizlilik teorisi, bir veri setinden veya istatistiksel modelden elde edilebilecek bilgi ile içindeki kişilerin mahremiyeti arasındaki ödünleşmeleri ölçmek ve en aza indirmek için bir çerçeve sağlar. Benzer şekilde, istatistiksel doğruluğu optimize ederken aynı zamanda matematiksel olarak tanımlanmış ayrım ölçütlerini en aza indiren ML modelleri oluşturmak için çeşitli yöntemler mevcuttur.

Bu yaklaşımlar teorik garantiler sağlarken, bunları anlamlı bir şekilde uygulamaya koymak zor olabilir. Çoğu durumda, gizlilik ve adalet gibi değerleri anlamlı bir şekilde ölçmek zordur. Örneğin, diferansiyel gizlilik, bir kişiin belirli bir veri setinden benzersiz olarak tanımlanma olasılığını ölçebilir, ancak bu tanımlamanın hassasiyetini ölçemez. Bu nedenle, bu yöntemleri her zaman daha kalitatif (niteliksel) ve bütüncül bir yaklaşımla desteklemelisiniz. Ancak, söz konusu değerleri tam olarak ölçememek, ödünleşmeyi değerlendirmekten ve haklı çıkarmaktan tamamen kaçınabileceğiniz anlamına gelmez; yine de seçimlerinizi gerekçelendirmeniz gerekiyor.

Kontrol örneği

Risk bildirisi

Yetersiz veya uygun olmayan ödünleşme analiz/kararları; bir kriteri diğer önemli kriterlere göre yanlış bir şekilde önceliklendiren yapay zeka sistemlerine yol açar.

Önleyici

- Modelin amacını ve model sözleşmenizdeki en önemli kriterleri açıkça belgeleyin.
- Bu sözleşmenin ilgili yönetim tarafından imzalandığından emin olun.
- Üst yönetim çeşitli modelleri inceler (ödünleşme analizi) ve belirli bir modeli kullanım için onaylar.
- Ödünleşme seçeneklerini sistematik olarak gözden geçirin ve belirli modelin neden seçildiğine dair gerekçe sağlayın.
- İnceleme tamamlandığından ve sonucunda aksiyon alındığından emin olun
- AI sistem tasarımcılarının en son teknikler ile güncel kalmasını sağlamak için eğitimleri tamamlayın.

Tespit edici

- Dağıtım tarihinden itibaren mevcut olan yeni veriler ile ödünleşmenin periyodik gözden geçirilmesini sağlayın.
- Ödünleşmeleri periyodik olarak yeniden analiz yapın.

Düzeltici

- Daha uygun bir ödünleşme seçin ve değişiklik için kapsamlı gerekçe ekleyin.
- AI sistem geliştiricilerini yeniden eğitin.

Yapay zeka sistemlerinde yasallık, adalet ve şeffaflığı sağlamak için ne yapmalıyız?

Genel bir bakış

Kişisel verileri işlemek için yapay zekayı kullandığınızda, bunun yasal, adil ve şeffaf olduğundan emin olmalısınız. Bu ilkelere uyum, bir yapay zeka bağlamında zor olabilir.

AI sistemleri, kişisel verileri çeşitli amaçlarla çeşitli aşamalarda işlediğinden, her bir farklı işlemi uygun bir şekilde ayırt edememe ve bunun için uygun bir yasal dayanak belirleyememe durumlarında risk bulunmaktadır. Bu durum veri koruma kanuna uygunluk ilkesine uyulmamasına neden olabilir.

Bu bölüm, AI oluştururken veya kullanırken, kullanılan çeşitli kişisel veri işleme türleri için uygun yasal bir temel bulmanıza ve bu tür işlemlerin adil olmasını sağlamanıza yardımcı olacak hususları sunar.

Detaylı olarak

- Yasallık, adalet ve şeffaflık ilkeleri yapay zekaya nasıl uygulanır?
- AI'yı kullanırken amaçlarımızı ve yasal dayanağımızı nasıl belirleriz?
- İstatistiksel doğruluk konusunda ne yapmamız gerekiyor?
- Önyargı ve ayrımcılık risklerini nasıl ele almalıyız?

Yasallık, adalet ve şeffaflık ilkeleri yapay zekaya nasıl uygulanır?

İlk olarak, yapay zeka sistemlerinin geliştirilmesi ve devreye alınması, kişisel verilerin farklı amaçlar için farklı şekillerde işlenmesini içerir. Yasallık ilkesine uymak için bu amaçları belirlemeli ve uygun bir yasal dayanağa sahip olmalısınız.

İkinci olarak, insanlar hakkında veri çıkarmaya yönelik bir yapay zeka sistemi kullanıyorsanız, bu işlemin adil olması için aşağıdakileri sağlamanız gerekir:

- sistemin istatistiksel olarak yeterince doğrudur ve ayrımcılığı önler.
- kişilerin makul beklentilerinin etkisini göz önünde bulundurulur.

Son olarak, şeffaflık ilkesine uymak için bir yapay zeka sisteminde kişisel verileri nasıl işlediğiniz konusunda şeffaf olmanız gerekir. Yapay zeka ve şeffaflık ilkesiyle ilgili temel konular explAIn kılavuzunda ele alındığı için burada ayrıntılı olarak tartışılmamaktadır.

AI'yı kullanırken amaçlarımızı ve yasal dayanağımızı nasıl belirleriz?

Yasal dayanaklara karar verirken neleri dikkate almalıyız?

İster yeni bir yapay zeka sistemi eğitin, ister mevcut bir sistemi kullanarak tahminlerde bulunun, kişisel verileri ne zaman işlerseniz, bunu yapmak için uygun bir yasal dayanağa sahip olmanız gerekir.

Özel koşullarınıza bağlı olarak farklı yasal dayanaklar geçerli olabilir. Bununla birlikte, bazı yasal dayanakların; AI'nın eğitimi ve/veya devreye alımı için diğerlerinden daha uygun olma olasılığı daha yüksek olabilir.

Aynı zamanda aşağıdakileri de unutmamalısınız:

- İşleminiz için hangi yasal dayanağın geçerli olduğuna karar vermek **sizin sorumluluğunuzda**dır.
- Her zaman kişiyle olan **ilişkinizin gerçek niteliğini** ve işlemenin amacını **en yakından yansıtan** yasal temeli seçmelisiniz.
- İşlemeye başlamadan **önce** bu tespiti yapmalısınız.
- Kararınızı **belgelemeli**siniz.
- Yasal dayanakları iyi bir sebep olmadan daha sonraki bir tarihte **değiştiremezsiniz.**
- Gizlilik bildiriminize (amaçlarla birlikte) **yasal dayanağınızı eklemelisiniz.**
- **Özel kategorilerdeki verileri** işliyorsanız, işleme için **hem** yasal bir dayanağa **hem de** ek bir koşula ihtiyacınız vardır.

İlave okuma – ICO kılavuzu

GDPR Kılavuzu'ndaki işleme için yasal dayanaklar kılavuzumuzu okuyun

Yapay zeka geliştirme ve dağıtımı arasındaki amaçları nasıl ayırt etmeliyiz?

Çoğu durumda, amaç(lar)ınızı ve yasal dayanağı belirlerken, AI sistemlerinin **geliştirilmesini** veya **eğitimini, devreye alınmalarından** ayırmanız sizin için mantıklı olacaktır. Bunun nedeni, bunların farklı koşullar ve risklerle, farklı ve ayrı amaçlar olmasıdır.

Bu nedenle, yapay zeka geliştirme ve devreye almanız için farklı yasal temellerin geçerli olup olmadığını düşünmelisiniz. Örneğin, bunu aşağıdaki durumlarda yapmanız gerekir.

- AI sistemi genel amaçlı bir görev için **eğitildi** ve ardından onu farklı amaçlar için farklı durumlarda **devreye alıyorsunuz**. Örneğin, bir yüz tanıma sistemi yüzleri tanımak için eğitilebilir, ancak bu işlevsellik, suçu önleme, kimlik doğrulama ve bir sosyal ağdaki arkadaşları etiketleme gibi birçok amaç için kullanılabilir. Bu ilave uygulamaların her biri farklı bir yasal dayanak gerektirebilir.
- Bir üçüncü taraf AI sistemi uyguladığınız durumlarda; geliştirici tarafından üstlenilen kişisel verilerin herhangi bir şekilde işlenmesi, sistemi kullanmayı düşündüğünüzden farklı bir amaç için olacaktır, bu nedenle farklı bir yasal dayanak belirlemeniz gerekebilir.
- Kişisel verilerin bir modelin eğitilmesi amacıyla işlenmesi kişileri doğrudan etkilemeyebilir, ancak model devreye alındıktan sonra otomatik olarak yasal veya önemli etkileri olan kararlar alabilir. Bu, otomatik karar vermeyle ilgili hükümlerin geçerli olduğu anlamına gelir; sonuç olarak, **eğitim** ve **devreye alma** aşamalarında farklı bir dizi mevcut yasal dayanak geçerli olabilir.

Aşağıdaki bölümler, GDPR'nin yasal dayanaklarının her biri için AI ile ilgili bazı hususları özetlemektedir. Bu aşamada DPA'nın 3. Bölümü dikkate alınmıyor.

Rızaya güvenebilir miyiz?

Modelinizi eğitmek ve dağıtmak için verilerini işlemek istediğiniz kişilerle doğrudan bir ilişkiniz olduğu durumlarda, rıza uygun bir yasal dayanak olabilir.

Bununla birlikte, rızanın özgürce verildiğinden, belirli, bilgilendirilmiş ve net olduğundan ve kişiler adına açık bir olumlu eylem içerdiğinden emin olmalısınız.

Rızanın avantajı, hizmetinizi kullanan kişilerden daha fazla güven oluşturması ve daha fazla katılıma yol açabilmesidir. Kişilere kontrol sağlamak da DPIA'larınızda bir faktör olabilir.

Ancak, rızanın geçerli olması için kişilerin, verilerini kullanıp kullanamayacağınıza dair gerçek bir seçeneğe sahip olmaları gerekir. Bunun, verilerle ne yapmayı amaçladığınıza bağlı olarak etkileri olabilir - daha karmaşık işleme operasyonları için geçerli onay toplamanızı sağlamak zor olabilir. Örneğin, verilerle ne kadar çok şey yapmak isterseniz, rızanın gerçekten spesifik ve bilgilendirici olmasını sağlamak o kadar zor olur.

Buradaki kilit nokta, bireylerin kişisel verilerini nasıl kullandığınızı anlamaları ve bu kullanıma rıza göstermeleridir. Örneğin, çeşitli sonuçları tahmin etmek için farklı modelleri keşfetmek üzere geniş bir özellik yelpazesi toplamak istiyorsanız, kişileri bu faaliyetler hakkında bilgilendirmeniz ve geçerli bir rıza almanız şartıyla, rıza uygun bir yasal dayanak olabilir.

Rıza ayrıca, bir AI sisteminin devreye alınması sırasında bir kişinin verilerinin kullanımı için uygun bir yasal dayanak olabilir (örneğin, hizmeti kişiselleştirmek veya bir tahmin veya öneride bulunmak gibi amaçlar için).

Ancak, rızanın geçerli olması için kişilerin de rıza verdikleri kadar kolay bir şekilde rızalarını geri alabilmeleri gerektiğini bilmelisiniz. Dağıtım sırasında bir AI sistemiyle veri işlemenin temeli olarak onaya güveniyorsanız (örneğin, kişiselleştirilmiş içerik sunma için), bu işlem için onayın geri çekilmesine uyum sağlamaya hazır olmalısınız.

GDPR'deki ilgili hükümler

Bknz. 4(11), 6(1)(a) 7, 8, 9(2)(a) maddeler ve 32, 38, 40, 42, 43, 171 gerekçeleri (dış bağlantı)

İlave okuma – ICO kılavuzu

Daha fazla bilgi için GDPR Kılavuzu'ndaki rızayla ilgili kılavuzumuzu okuyun.

İlave okuma - Avrupa Veri Koruma Kurulu

29. Madde Çalışma Grubunun (WP29) yerini alan Avrupa Veri Koruma Kurulu (EDPB - European Data Protection Board), her AB üye devletinin veri koruma yetkililerinden temsilciler içerir. GDPR gerekliliklerine uymak için yönergeleri kabul eder.

WP29, EDPB'nin Mayıs 2018'de onayladığı rızaya ilişkin kılavuz ilkeleri kabul etti.

Bir sözleşmenin yerine getirilmesine güvenebilir miyiz?

Bu yasal esas, ilgili kişi için sözleşmeye dayalı bir hizmet sunmak veya kişinin talebi üzerine sözleşmeye girmeden önce atılan adımlar için yapay zeka kullanılarak yapılan işlemin nesnel olarak gerekli olduğu durumlarda geçerlidir. (örn. Bir hizmete yönelik yapay zeka kaynaklı bir fiyat teklifi sağlamak için.)

Aynı hizmeti sağlamak için verileri işlemenin daha az müdahaleci bir yolu varsa veya verileri işleme, sözleşmenin yerine getirilmesi için uygulamada nesnel olarak gerekli değilse, verilerin AI ile işlenmesi için bu yasal esasa güvenemezsiniz.

Ayrıca, sistemin **kullanımı** için uygun bir zemin olsa bile, bu, bir AI sistemini **eğitmek** üzere kişisel verilerin işlenmesi için uygun bir zemin olmayabilir. Bir AI sistemi; bireyin kişisel verileri konusunda eğitim **almadan** yeterince iyi performans gösterebiliyorsa, sözleşmenin yerine getirlmesi bu tür işleme bağlı değildir.

Benzer şekilde, bir sözleşmenin yerine getirilmesini, sözleşmeden önce fiyat teklifi vermek üzere, AI ile kişisel verileri işlemek için yasal bir dayanak olarak

kullanabilseniz bile, bu, AI sistemini eğitmek için bu verileri kullanmayı haklı çıkarmak için de kullanabileceğiniz anlamına gelmez.

Ayrıca, AI sisteminizin 'hizmet iyileştirme' gibi amaçlarla kişisel verileri işlemek için bu temele güvenemeyeceğinizi de unutmamalısınız. Bunun nedeni çoğu durumda; bir hizmetin kullanımı hakkında kişisel verilerin toplanmasının, kullanıcıların bu hizmetle nasıl etkileşime girdiğinin ayrıntılarının veya bu hizmet içinde yeni işlevlerin geliştirilmesinin bir sözleşmenin sağlanması için nesnel olarak gerekli olmamasıdır. Bunun nedeni, hizmetin böyle bir işlem yapılmadan teslim edilebilmesidir.

Buna karşılık, içeriği kişiselleştirme amacıyla kişisel verileri işlemek için yapay zekanın kullanılması, bir sözleşmenin ifası için gerekli olarak kabul edilebilir - ancak sadece bazı durumlarda. Bu işlemin hizmetiniz için "esas" olarak kabul edilip edilemeyeceği şunlara bağlıdır:

- Hizmetin niteliği
- Kişilerin beklentileri
- Hizmetinizi bu işleme olmadan sağlayıp sağlayamayacağınız (yani, bir AI sistemi aracılığıyla içeriğin kişiselleştirilmesi hizmetin ayrılmaz bir parçası değilse, alternatif bir yasal dayanak düşünmelisiniz).

Mevzuattaki ilgili hükümler

Bknz. GDPR'ın 6(1)(b) maddesi and 44 gerekçesi (dış bağlantı)

İlave okuma – ICO kılavuzu

Daha fazla bilgi için GDPR Kılavuzu'ndaki sözleşmeyle ilgili kılavuzumuzu okuyun.

İlave okuma - Avrupa Veri Koruma Kurulu

29. Madde Çalışma Grubunun (WP29) yerini alan Avrupa Veri Koruma Kurulu (EDPB - European Data Protection Board), her AB üye devletinin veri koruma yetkililerinden temsilciler içerir. GDPR gerekliliklerine uymak için yönergeleri kabul eder.

EDPB, Kasım 2019'da çevrimiçi hizmetler bağlamında Madde 6(1)(b) kapsamında kişisel verilerin işlenmesine ilişkin Kılavuz İlkeler 2/2019'u yayınladı.

Yasal zorunluluğa, kamu yararına veya hayati çıkarlara güvenebilir miyiz?

Kişisel verileri işlemek için bir AI sisteminin kullanılmasının **yasal zorunluluk** olabileceği bazı örnekler vardır (örneğin, 2002 Suç Hareketleri Yasası'nın 7. Bölümü kapsamında kara para aklamanın önlenmesi amacıyla).

Benzer şekilde, bir organizasyon yapay zekayı resmi yetkisinin bir parçası olarak kullanması veya kanunla belirlenen **kamu yararına** bir görevi gerçekleştirmesi durumunda, ilgili kişisel verilerin gereği gibi işlenmesi bu gerekçelere dayalı olabilir.

Sınırlı durumda, bir AI sistemi tarafından kişisel verilerin işlenmesi, kişilerin **hayati çıkarlarının** korunmasına dayalı olabilir. Örneğin, başka bir şekilde rıza gösteremeyecek durumda olan hastaların acil tıbbi teşhisi için (örneğin, bilinci kapalı bir hastanın FMRI taramasının bir AI teşhis sistemi tarafından işlenmesi).

Bununla birlikte, hayati çıkarların bir AI sistemini eğitmek için bir temel oluşturması pek olası değildir, çünkü bu, nadiren doğrudan ve hemen kişilerin hayati çıkarlarının korunmasına yol açacaktır, hatta inşa edilen modeller daha sonra hayat kurtarmak için kullanılsa bile. Potansiyel olarak hayat kurtaran AI sistemlerinin eğitimi için diğer yasal temellere güvenmek daha iyi olacaktır.

Mevzuattaki ilgili hükümler

Bknz. Yasal yükümlülüğün kullanılmasına ilişkin hükümler için GDPR'ın 6(1)(c) madde ve 41, 45 gerekçesi (dış bağlantı)

Bknz. Kamu Çıkarlarının kullanılmasına ilişkin hükümler için GDPR'ın 6 (1)(e) ve 6(3) maddeleri, ve 41, 45 and 50 gerekçeleri

Bknz. Hayati menfaatlerin kullanılmasına ilişkin hükümler için GDPR Article 6(1)(d) , 9(2)(c) maddeleri ve 46 gerekçesi (dış bağlantı)

Bknz. Veri Koruma Yasasının Bölüm 7 ve 8, ve Program 1'in 6 ve 7 paragrafları(external link)

İlave okuma – ICO kılavuzu

Daha fazla bilgi için GDPR Kılavuzundaki yasal zorunluluk, hayati çıkarlar ve kamu görevi hakkındaki kılavuzumuzu okuyun.

Meşru çıkarlara güvenebilir miyiz?

Koşullarınıza bağlı olarak, kişisel verilerinizi hem eğitim hem de devam eden AI kullanımı için yasalara uygun olarak meşru çıkarlara dayandırabilirsiniz.

Meşru çıkarların, işleme için en esnek yasal dayanak olmasına rağmen, her zaman en uygun olanın bu olmadığını belirtmete fayda vardır. Örneğin, insanların verilerini kullanma şekliniz beklenmedik bir durum veya gereksiz zarara neden

olursa. Ayrıca, insanların haklarını ve çıkarlarını göz önünde bulundurmak ve korumak için ek sorumluluk aldığınız anlamına gelir.

Ayrıca, bir kamu idaresi iseniz, bir kamu makamı olarak görevlerinizi yerine getirmek dışında meşru bir nedenle işlem yapıyorsanız yalnızca meşru çıkarlara güvenebilirsiniz.

Meşru çıkarların yasal dayanağının üç unsuru vardır ve bunları 'üç parçalı test' olarak düşünmek yardımcı olabilir. Yapmanız gerekenler:

- Meşru bir çıkarın belirlenmesi (‘amaç testi’)
- İşlemenin bunu başarmak için gerekli olduğunun gösterilmesi ('gereklilik testi')
- Kişinin çıkarları, hakları ve özgürlüklerine göre dengelenmesi (“dengeleme testi”).

Veri koruma hukukunda 'meşru çıkarları' oluşturan çok çeşitli çıkarlar olabilir. Bunlar sizin veya üçüncü şahısların yanı sıra ticari veya toplumsal çıkarlar olabilir. Bununla birlikte, kilit nokta, meşru çıkarlar daha esnek olsa da bunun ek sorumluluklarla geldiğini ve işlemenizin kişiler üzerindeki etkisini değerlendirmenizi ve işlemenin ikna edici bir yararı olduğunu kanıtlayabilmenizi gerektirdiğini anlamaktır.

Bu hususları meşru çıkar değerlendirmenizin (LIA – legitimate interests assesment) bir parçası olarak ele almalı ve belgelemelisiniz.

Örnek

Bir organizasyon, bir makine öğrenme modelini eğitmek amacıyla kişisel verileri işlemek için meşru çıkarlara güvenmeye çalışır.

Meşru çıkarlar, organizasyonun modeli için farklı değişkenler denemek üzere en fazla alana izin verebilir.

Ancak, meşru çıkar değerlendirmesinin bir parçası olarak, organizasyon, kullanmayı amaçladığı değişkenler ve modeller yelpazesinin, sonuca ulaşmak için makul bir yaklaşım olduğunu göstermelidir.

Bunu, tüm amaçlarını doğru bir şekilde tanımlayarak ve toplanan her tür verinin kullanımını gerekçelendirerek en iyi şekilde başarabilir - bu, organizasyonun LIA'sının gerekliliği ve dengeleyici yönleri üzerinde çalışmasına izin verecektir.

Örneğin, bazı verilerin bir tahmin için faydalı olma olasılığı, organizasyonun bu verileri işlemenin modeli oluşturmak için gerekli olduğunu göstermesi için tek başına yeterli değildir.

Mevzuattaki ilgili hükümler

Bknz. GDPR'ın 6(1)(f) madde ve 47-49 gerekçeleri (external link)

İlave okuma – ICO kılavuzu

GCPR Kılavuz'undaki meşru çıkarlar hakkındaki kılavuzumuzu okuyun.

Ayrıca, sizin için hangi temelin uygun olduğuna karar vermenize yardımcı olması için kullanabileceğiniz yasal bir temel değerlendirme aracının yasal bir temel değerlendirme aracının yanı sıra meşru çıkar şablonu (Word) yayınladık.

Özel kategori verileri ve cezai suçlarla ilgili veriler ne olacak?

Özel kategori verilerini veya cezai suçlarla ilgili verileri işlemek için AI kullanmayı planlıyorsanız, DPA 2018'in yanı sıra GDPR'nin 9. ve 10. Maddelerinin gereksinimlerine uyduğunuzdan emin olmanız gerekir.

Özel kategori veriler, hassas olduğu için daha fazla korunması gereken kişisel verilerdir. Bunu işlemek için 6. Madde kapsamında yasal bir dayanağa ve ayrıca 9. Madde kapsamında ayrı bir koşula ihtiyacınız vardır, ancak bunların bağlantılı olması gerekmez.

Bu koşulların bir kısmı, DPA 2018'in 1. Çizelgesi'nde belirtilen ek gereksinimleri ve önlemleri karşılamanızı da gerektirecektir.

Yapmanız gerekenler:

- Başlamadan önce işleme koşulunuzu belirleyin ve belgeleyin.
- Gerektiğinde uygun bir politika belgesine sahip olduğunuzdan emin olun.
- Yüksek riskli olması muhtemel herhangi bir işleme için bir DPIA tamamlayın.

Cezai suçlarla ilgili veriler için GDPR'nin 6. Maddesi uyarınca yasal bir dayanağa ve 10. Maddeye göre yasal veya resmi otoriteye ihtiyacınız vardır.

DPA 2018, yasal yetki sağlayan belirli koşulları ortaya koymaktadır. Resmi yetkiniz varsa (yani verileri resmi bir sıfatla işliyorsanız) bu tür verileri de işleyebilirsiniz.

Özel kategori verilerde olduğu gibi, işlemeye başlamadan önce durumunuzu kararlaştırmalı (veya resmi yetkilinizi belirlemelisiniz) ve bunu da belgelemelisiniz.

Mevzuattaki ilgili hükümler

Bknz. GDPR'ın 9 ve 10 maddeleri (dış bağlantı)

İlave okuma – ICO kılavuzu

GDPR Kılavuzu'ndaki özel kategori verileri ve cezai suç verileri hakkındaki kılavuzumuzu okuyun.

GDPR'nin 22. Maddesinin etkisi nedir?

Veri koruma kanunu, tüm otomatik kişisel karar verme ve profil oluşturma işlemleri için geçerlidir. GDPR'nin 22. Maddesi; üzerlerinde yasal veya benzer şekilde önemli etkileri olan yalnızca otomatik karar verme işlemi yürütüyorsanız, kişileri korumaya yönelik ek kurallara sahiptir.

Bunun AI bağlamında uygulaması olabilir. Örneğin: bu tür kararları vermek için AI sistemi kullandığınız yer.

Ancak, bu tür bir karar verme işlemini yalnızca kararın verildiği aşağıdaki durumlarda gerçekleştirebilirsiniz.

- Karar, sözleşmeye girmek veya sözleşmenin uygulanması için gereklidir.
- Karar, sizin için geçerli olan yasa tarafından yetkilendirilmiştir.
- Karar, kişinin açık rızasına dayanmaktadır.

Bu nedenle, işleminizin Madde 22'nin kapsamına girip girmediğini belirlemelisiniz ve bu kapsama giriyorsa, aşağıdakileri yaptığınızdan emin olmalısınız:

- Kişilere işlem hakkında bilgi verildiğinden.
- İnsan müdahalesi talep etmeleri veya bir karara itiraz etmeleri için basit yollar tanıtıldığından.
- Sistemlerinizin amaçlandığı gibi çalıştığından emin olmak için düzenli kontroller yapıldığından.

İlave okuma – ICO kılavuzu

GDPR Kılavuzu'nda profiloluşturma dahil otomatik karar verme ile ilgili haklara ilişkin kılavuzumuzu okuyun.

Örnek kontroller

Risk Bildirimi

İşlem için uygun olmayan bir yasal temele güvenmek; lazım olan gereksinimlerin yerine getirilmemesi ve DP mevzuatına uyulmaması ile sonuçlanabilir.

Önleyici

- Yapay zeka sistem geliştiricilerinin eğitimi ve ilgili yetkinlik değerlendirmelerini tamamladıklarından emin olun.
- Kilit paydaşlar için eğitim ve ilgili personelin nasıl belirlendiğini belgeleyin (örn. üst yönetim, risk yöneticileri, denetim).
- DPIA'nıza işlemek için yasal dayanağınızı iyice değerlendirin ve gerekçelendirin.
- Model tasarım iş gücünüzdeki DP uzmanlarına danışın.
- Bir DPIA gereksiniminin belgelendiğinden ve AI geliştiricilerine değerlendirme kriterleri hakkında net rehberlik sağlandığından emin olun.
- Meşru çıkarlara yasal bir dayanak olarak güvenilmesi durumunda meşru bir çıkar değerlendirmesini tamamlayın.

Tespit edici

- Sonuç olarak alınan önlemler de dahil olmak üzere kişilerden gelen kişisel hak taleplerini ve şikayetlerini izleyin (hem kişisel düzeyde hem üst düzey analizlerinde)
- Doğru ve güncel kalmasını sağlamak için periyodik bir DPIA incelemesi yapın.
- Amacın aynı kaldığından, gereklilik ve meşru menfaatlerin (LI) hala geçerli olduğundan emin olmak için model kullanımını periyodik olarak değerlendirin.
- Yasal dayanağın geçerliliğini sağlamak için işlem kayıtlarını periyodik olarak gözden geçirin.

Düzeltici

- Orijinal yasal temeli karşılamak için AI sistemine düzeltici önlemler uygulayın.
- Yeni bir yasal dayanak ve ilgili eylemleri seçin. Örneğin, meşru bir çıkar değerlendirmesi yapmak veya onay almak.
- Yasal temellerin değerlendirilmesine dahil olan yapay zeka sistem geliştiricilerini / kişileri yeniden eğitin.

İstatistiksel doğruluk hakkında ne yapmamız gerekiyor?

İstatistiksel doğruluk, bir AI sisteminin doğru veya yanlış aldığı cevapların oranını ifade eder.

Bu bölüm, AI sistemlerinizin işledikleri kişisel verilerin adalet ilkesine uygun olmasını sağlamaya yönelik yeterince istatistiksel olarak doğru olması için uygulayabileceğiniz kontrolleri açıklar.

Veri koruma yasasındaki "doğruluk" ile yapay zekadaki "istatistiksel doğruluk" arasındaki fark nedir?

"Ödünleşmeler nelerdir ve bunları nasıl yönetmeliyiz?" bölümünde belirtildiği gibi, doğruluk, veri koruma ve yapay zeka bağlamlarında biraz farklı anlamlara sahiptir.

Veri korumada doğruluk temel ilkelerden biridir. İşlediğiniz kişisel verilerin 'herhangi bir konuda yanlış veya yanıltıcı' olmadığından ve gerektiğinde gereksiz gecikme olmaksızın düzeltildiğinden veya silindiğinden emin olmak için tüm makul adımları atmanızı gerektirir.

AI'da doğruluk, bir AI sisteminin doğru cevabı ne sıklıkta tahmin ettiğini ifade eder. Birçok durumda, AI sisteminin sağladığı cevaplar kişisel veriler olacaktır. Örneğin, bir AI sistemi, birinin demografik bilgilerini veya ilgi alanlarını bir sosyal ağdaki davranışlarından çıkarabilir.

Veri korumanın **doğruluk ilkesi**, bir kişi hakkındaki bilgiler ister bir yapay zeka sistemine girdi olarak kullanılsın isterse sistemin bir çıktısı olsun, tüm kişisel veriler için geçerlidir. Ancak bu, doğruluk ilkesine uymak için bir AI sisteminin %100 **istatistiksel doğru** olması gerektiği anlamına gelmez.

Çoğu durumda, bir AI sisteminin çıktılarının, kişi hakkında gerçek bilgiler olarak ele alınması amaçlanmamıştır. Bunun yerine, kişi hakkında şimdi veya gelecekte doğru olabilecek bir şey hakkında istatistiksel olarak bilgilendirilmiş bir tahmini temsil etmeleri amaçlanır. Bu tür kişisel verilerin gerçeğe dayalı olarak yanlış yorumlanmasını önlemek için, kayıtlarınızın gerçeklerden ziyade istatistiksel olarak bilgilendirilmiş tahminler belirttiğinden emin olmalısınız. Kayıtlarınız ayrıca, verilerin kaynağı ve çıkarım oluşturmak için kullanılan AI sistemi hakkında bilgi içermelidir.

Ayrıca, çıkarımın yanlış verilere dayandığı veya bunu oluşturmak için kullanılan AI sisteminin, çıkarımın kalitesini etkilemiş olabilecek şekilde istatistiksel olarak kusurlu olduğu anlaşılırsa da kaydetmelisiniz.

Benzer şekilde, yanlış çıkarımın işlenmesinin kişiler üzerinde bir etkisi olabilirse, bir kişi yanlış çıkarsamaya karşı kayıtlarına ek bilgilerin dahil edilmesini talep edebilir. Bu, potansiyel olarak yanlış çıkarım temelinde alınan herhangi bir kararın, yanlış olabileceğine dair herhangi bir kanıt tarafından bilgilendirilmesine yardımcı olur.

GDPR, gerekçe 71'de profil oluşturma ve otomatik karar verme bağlamında istatistiksel doğruluktan bahseder. Organizasyonların teknik önlemlerinin bir parçası olarak kişilerin profilini çıkarmak için 'uygun matematiksel ve istatistiksel prosedürleri' devreye sokması gerektiğini belirtir. Kişisel verilerde yanlışlıklara

neden olabilecek faktörlerin düzeltilmesini ve hata riskinin en aza indirilmesini sağlamalısınız.

İnsanlar hakkında çıkarımlarda bulunmak için bir yapay zeka sistemi kullanıyorsanız, sistemin amaçlarınız için yeterince istatistiksel olarak doğru olduğundan emin olmanız gerekir. Bu, her çıkarımın doğru olması gerektiği anlamına gelmez, ancak bunların yanlış olma olasılığını hesaba katmanızı ve bunlara dayanarak alabileceğiniz herhangi bir karar üzerindeki etkisini hesaba katmanızı gerektirir. Bunun yapılmaması, işleminizin adalet ilkesine uygun olmadığı anlamına gelebilir. Çıkarımlar da dahil olmak üzere kişisel verilerin amacınıza uygun ve yeterli olması gerektiğinden, veri minimizasyonu ilkesine uymanızı da etkileyebilir.

Bu nedenle, AI sisteminizin, ürettiği herhangi bir kişisel verinin yasal ve adil bir şekilde işlenmesini sağlamak için istatistiksel olarak yeterince doğru olması gerekir.

Bununla birlikte, genel istatistiksel doğruluk özellikle yararlı bir ölçü değildir ve genellikle farklı ölçülere bölünmesi gerekir. Doğru olanları ölçmek ve önceliklendirmek önemlidir (bir sonraki bölüme bakınız).

Mevzuattaki ilgili hükümler

Bknz. GDPR'ın 5(1)(d), 22 maddeleri ve 71 gerekçesi (dış bağlantı)

İlave okuma – ICO kılavuzu

GDPR Kılavuzumuzdaki doğruluk kılavuzumuz ile düzeltme ve silme haklarına ilişkin kılavuzumuzu okuyun.

İlave okuma - Avrupa Veri Koruma Kurulu

29. Madde Çalışma Grubunun (WP29) yerini alan Avrupa Veri Koruma Kurulu (EDPB - European Data Protection Board), her AB üye devletinin veri koruma yetkililerinden temsilciler içerir. GDPR gerekliliklerine uymak için yönergeleri kabul eder.

WP29, 2017'de otomatik karar verme ve profil oluşturma kılavuz ilkeleri yayınladı. EDPB, Mayıs 2018'de bu kılavuz ilkeleri onayladı.

Farklı istatistiksel doğruluk ölçütlerini nasıl tanımlamalı ve önceliklendirmeliyiz?

Genel bir ölçü olarak istatistiksel doğruluk, bir yapay zeka sisteminin tahminlerinin, test verilerinde tanımlandığı gibi doğru etiketlerle ne kadar yakından eşleştiği ile ilgilidir.

Örneğin, e-postaları spam veya spam olmayan olarak sınıflandırmak için bir AI sistemi kullanılıyorsa, istatistiksel doğruluğun basit bir ölçüsü, analiz edilen tüm e-postaların bir oranı olarak, spam veya spam olmayan olarak doğru şekilde sınıflandırılan e-postaların sayısıdır.

Ancak böyle bir önlem yanıltıcı olabilir. Örneğin, bir gelen kutusuna alınan tüm e-postaların %90'ı spam ise, her şeyi spam olarak etiketleyerek %90 doğru bir sınıflandırıcı oluşturabilirsiniz. Ancak bu, gerçek bir e-posta'yı geçirmeyeceği için sınıflandırıcının amacını ortadan kaldırır.

Bu nedenle, bir sistemin ne kadar iyi olduğunu değerlendirmek için alternatif ölçüler kullanmalısınız. Bu önlemler, iki farklı hata türü arasındaki dengeyi yansıtmalıdır:

- **yanlış pozitif** veya 'tip I' hata: bunlar, AI sisteminin yanlış olarak pozitif olarak etiketlediği durumlardır (örneğin, gerçek olduklarında spam olarak sınıflandırılan e-postalar).
- **yanlış negatif** veya 'tip II' hata: bunlar, gerçekte pozitif olduklarında AI sisteminin yanlış olarak negatif olarak etiketlediği durumlardır (ör. gerçek olarak sınıflandırılan e-postalar, aslında spam olduklarında).

Bu iki tür hata arasındaki dengeyi sağlamak önemlidir. Aşağıdakiler dahil olmak üzere bu iki tür hatayı yansıtan daha yararlı önlemler vardır:

- **kesinlik-precision**: pozitif olarak tanımlanan ve aslında pozitif olan vakaların yüzdesi ("pozitif tahmin değeri" olarak da adlandırılır). Örneğin, spam olarak sınıflandırılan 10 e-postadan dokuzu gerçekten spam ise, AI sisteminin kesinliği %90'dır.
- **hatırlama-recall** (veya hassaslık-sensitivity): aslında pozitif olan ve bu şekilde tanımlanan tüm vakaların yüzdesi. Örneğin, 100 e-postadan 10'u gerçekten spam ise, ancak AI sistemi yalnızca yedi tanesini tanımlıyorsa, hatırlama oranı %70'tir.

Kesinlik ve hatırlama arasında ödünleşme vardır ("F-1" istatistiği gibi önlemler kullanılarak değerlendirilebilir - aşağıdaki bağlantıya bakın). Mümkün olduğunca çok sayıda pozitif vakalar bulmaya daha fazla önem verirseniz (hatırlamayı en üst düzeye çıkarırsanız), bu, bazı yanlış pozitiflerin (kesinliği düşürme) maliyetine neden olabilir.

Ayrıca, yanlış pozitiflerin ve yanlış negatiflerin kişiler üzerindeki sonuçları arasında önemli farklılıklar olabilir.

Örnek

Bir görüşme için nitelikli adayları seçen bir özgeçmiş filtreleme sistemi yanlış bir pozitif sonuç verirse, o zaman niteliksiz bir aday görüşmeye davet edilerek işvereni ve başvuranın zamanını gereksiz yere boşa harcar.

Yanlış bir negatif üretirse, nitelikli bir aday bir istihdam fırsatını kaçıracak ve organizasyon iyi bir adayı kaçıracaktır.

Risklerin şiddetine ve niteliğine göre belirli hata türlerinden kaçınmaya öncelik vermelisiniz.

Genelde, bir ölçü olarak istatistiksel doğruluk, bir sistemin çıktılarının performansını bazı "temel gerçek" ile karşılaştırmanın ne kadar mümkün olduğuna bağlıdır (yani, AI sisteminin sonuçlarını gerçek dünyayla kontrol etmek). Örneğin, kötü huylu tümörleri tespit etmek için tasarlanmış bir tıbbi teşhis aracı, bilinen hasta sonuçlarını içeren yüksek kaliteli test verilerine karşı değerlendirilebilir.

Diğer bazı alanlarda, bir temel gerçeğe ulaşılamaz olabilir. Bunun nedeni, yüksek kaliteli test verilerinin olmaması veya tahmin etmeye veya sınıflandırmaya çalıştığınız şeyin öznel olması (örneğin, bir sosyal medya gönderisinin saldırgan olup olmadığı) olabilir. Bu durumlarda istatistiksel doğruluğun yanlış yorumlanması riski vardır, bu nedenle yapay zeka sistemleri, nesnel gerçeklerden ziyade bir dizi insan etiketleyicinin düşündüklerinin ortalamasını yansıtsalar bile istatistiksel olarak oldukça doğru olarak görülür.

Bundan kaçınmak için, kayıtlarınız AI çıktılarının nesnel gerçekleri yansıtmasının amaçlanmadığı yerleri belirtmelidir ve bu tür kişisel veriler temelinde alınan herhangi bir karar bu sınırlamaları yansıtmalıdır. Bu aynı zamanda doğruluk ilkesini dikkate almanız gereken yerlere bir örnektir – daha fazla bilgi için, görüşlerin doğruluğuna atıfta bulunan **doğruluk ilkesi**ne ilişkin kılavuzumuza bakın.

Son olarak, istatistiksel doğruluk statik bir ölçü değildir. Genellikle statik test verileri ile ölçülse de, gerçek hayatta AI sistemleri yeni ve değişen popülasyonlara uygulanır. Bir sistem, mevcut bir popülasyonun verileri hakkında istatistiksel olarak doğru olduğu için (örneğin, son bir yıldaki müşteriler), o popülasyonun veya sistemin gelecekte uygulandığı başka bir popülasyonun özelliklerinde bir değişiklik olursa, iyi performans göstermeye devam etmeyebilir. Davranışlar ya kendi istekleriyle ya da sisteme tepki olarak adapte oldukları için değişebilir ve yapay zeka sistemi zamanla istatistiksel olarak daha az doğru hale gelebilir.

Bu olgu, makine öğreniminde 'kavram / model kayması' (‘concept/model drift’) olarak adlandırılır ve bunu tespit etmek için çeşitli yöntemler mevcuttur. Örneğin, zaman içinde sınıflandırma hataları arasındaki açıklığı ölçebilirsiniz; giderek artan sıklıkta oluşan hatalar kaymaya sebeb olabilir.

Kaymayıı düzenli olarak değerlendirmeli ve gerektiğinde modeli yeni veriler üzerinde yeniden eğitmelisiniz. Hesap verebilirliğinizin bir parçası olarak, modelinizin yeniden eğitilmesi gerekip gerekmediğine, işlemin doğasına, kapsamına, durumuna, amaçlarına ve oluşturduğu risklere dayanarak belirlemeli, uygun eşiklere karar vermeli ve belgelemelisiniz. Örneğin, modeliniz bir işe alım alıştırmasının parçası olarak CV'leri puanlıyorsa ve adayların belirli bir işte ihtiyaç duyduğu beceri türleri her iki yılda bir değişirse, en az bu sıklıkla yeni verilerinizi yeniden eğitme ihtiyacını değerlendirmeyi öngörmelisiniz.

Ana özelliklerin çok sık değişmediği diğer uygulama alanlarında (örneğin, el yazısı rakamları tanıma), daha az kayma olmasını bekleyebilirsiniz. Bunu kendi koşullarınıza göre değerlendirmeniz gerekecek.

İlave okuma – ICO kılavuzu

GDPR kılavuzundaki doğruluk ilkesine ilişkin kılavuzumuza bakın

Diğer kaynaklar

Bknz ‘Farklı istatistiksel doğruluk ölçütlerini nasıl tanımlamalı ve önceliklendirmeliyiz?’

Kavram kaymasının daha fazla açıklaması için ‘Learning under concept drift: an overview’ bakın

Ne yapmalıyız?

Herhangi bir tahmin veya karar verme sürecini otomatikleştirmenin uygun olup olmadığını her zaman baştan dikkatlice düşünmelisiniz. Bu, AI sisteminin, kişisel verileri işlenen kişiler hakkında istatistiksel olarak doğru tahminler yapmadaki etkinliğini değerlendirmeyi içermelidir.

Belirli bir yapay zeka sistemini kullanmanın yararlarını, doğru ve dolayısıyla değerli tahminler yapmadaki etkinliğini göz önünde bulundurarak değerlendirmelisiniz. Tüm AI sistemleri, kullanımlarını haklı göstermek için yeterli düzeyde istatistiksel doğruluk göstermez.

Bir yapay zeka sistemini benimsemeye karar verirseniz, veri koruma ilkelerine uymak için şunları yapmalısınız:

- Geliştirme, test etme, doğrulama, devreye alma ve izlemeden sorumlu tüm fonksiyonların ve kişilerin, ilgili istatistiksel doğruluk gerekliliklerini ve önlemlerini anlamak için yeterince eğitilmiş olduğundan emin olun.

- Verilerin çıkarımlar ve tahminler olarak açıkça etiketlendiğinden ve gerçeğe dayalı oldukları iddia edilmediğinden emin olun.
- Ödünleşmeleri ve makul beklentileri yönettiğinizden emin olun.
- Personelin; sınırlamalar ve kişiler üzerindeki olumsuz etkileri de dahil olmak üzere, istatistiksel doğruluk performans ölçütlerini tartışmak için kullanabileceği ortak bir terminoloji benimseyin.

Başka ne yapmalıyız?

Tasarım gereği ve default olarak veri koruma uygulama yükümlülüğünüzün bir parçası olarak, tasarım aşamasından itibaren değerlendirmek için istatistiksel doğruluğu ve uygun önlemleri dikkate almalı ve bu önlemleri AI yaşam döngüsü boyunca test etmelisiniz.

Dağıtımdan sonra, yanlış bir çıktının kişiler üzerindeki etkisiyle orantılı olması gereken izlemeyi uygulamalısınız. Etki ne kadar yüksek olursa, o kadar sık izlemeli ve rapor etmelisiniz. Kavram kayması riskini azaltmak için istatistiksel doğruluk ölçümlerinizi de düzenli olarak gözden geçirmelisiniz. Değişiklik politikası prosedürleriniz bunu en baştan dikkate almalıdır.

Bir AI sisteminin geliştirilmesini üçüncü bir tarafa (tamamen veya kısmen) dış kaynaktan sağlıyorsanız veya harici bir satıcıdan bir AI çözümü satın alıyorsanız, istatistiksel doğruluk da önemli bir husustur. Bu durumlarda, tedarik sürecinin bir parçası olarak üçüncü şahıslar tarafından yapılan iddiaları incelemeli ve test etmelisiniz.

Benzer şekilde, değişen veri popülasyonuna ve kavram / model kaymasına karşı korunmak için düzenli güncellemeleri ve istatistiksel doğruluk incelemelerini kabul etmelisiniz. Yapay zeka hizmetleri sağlayıcısıysanız, bunların organizasyonların veri koruma yükümlülüklerini yerine getirmelerine olanak tanıyacak şekilde tasarlandığından emin olmalısınız.

Son olarak, AI sistemlerinizin bir parçası olarak tutabileceğiniz ve işleyebileceğiniz çok miktarda kişisel veri, tanımlamak için kullandığınız önceden var olan AI olmayan süreçler üzerinde baskı oluşturması muhtemeldir. Gerekirse, girdi veya eğitim/test verisi olarak kullanılıp kullanılmadığına bakılmaksızın yanlış kişisel verileri düzeltin/silin.

Örnek kontroller

Risk Bildirimi

Yapay zeka sistemleri tarafından alınan hatalı çıktılar veya kararlar; kişiler için haksız/olumsuz sonuçlara ve adalet ilkesinin karşılanamamasına yol açabilir.

Önleyici

- Bir AI sistemi veya hizmetinin devam eden eğitimi, testi veya gelişimi için kullanılan tüm kişisel verilerin mümkün olduğunca doğru, kesin, amaca uygun, temsili, eksiksiz ve güncel olduğunu açıklayan bir veri yönetişim çerçevesini uygulamaya koyun.
- Kilit paydaşlar için eğitim sağlayın ve ilgili personelin nasıl belirlendiğini belgeleyin (örn. üst yönetim, risk yöneticileri, denetim).
- İstatistiksel doğruluğu etkileyen değişikliklerin yetkili kişiler tarafından yapıldığından ve imzalandığından emin olmak için, AI sistemlerinin geliştirilmesi ve devreye alınması için erişim yönetimi kontrollerinizi ve görevler ayrımını belgeleyin. Bu kontrollerin nasıl izlendiğine dair kanıt bulundurun.
- AI sistemlerinin geliştirilmesi/kullanımı için onay makamının seviyelerini belgelendirin. Uygun onayın kanıtını koruyun.
- DPIA'nıza, farklı hataların etkisinin/öneminin kapsamlı bir değerlendirmesini dahil edin.
- Üçüncü taraflarla ilgilenmek için belgelenmiş politikaları/süreçleri ve tamamlanan durum tespitine dair kanıtları koruyun. Özellikle, AI sistemlerini/hizmetlerini satın alırken, istatistiksel doğruluk gereksinimlerini karşıladıklarından ve düzenli yeniden test yapılmasına izin verdiklerinden emin olun.
- Herhangi bir AI sisteminin veya değişikliklerin canlıya alınmadan önce uygulama öncesi testini gerçekleştirmek için bir politika/süreç bulundurun ve belgeleyin. AI sisteminin dağıtımından önce testlerin tamamlandığına dair kanıtları ve test(ler)in sonuçlarını koruyun.

Tespit edici

- Uygulama sonrası testler yapın, testlerin sonuçlarını ve sonuç olarak alınan eylem(ler)i belgeleyin.
- Beklentilere göre raporların/performansın çıktısını izleyin.
- Örneğin nasıl seçildiği / ölçütlerin nasıl kullanıldığı da dahil olmak üzere istatistiksel doğruluk için AI kararlarının bir örneğini insan incelemesi olarak yapın.
- Özellikle 22. Madde ile ilgili olanlar da dahil olmak üzere, kişilerden gelen AI sistemlerinden elde edilen istatistiksel olarak yanlış çıktılara ilişkin kişisel hak taleplerini ve şikayetlerini belgeleyin. (hem bireysel düzeyde hem de daha geniş analizde).

- Beklentilere karşı performansın düzenli olarak gözden geçirilmesi ve sözleşme gerekliliklerine uyulması dahil olmak üzere herhangi bir üçüncü taraf tedarikçinin/işleyicinin sürekli gözetiminin yapıldığından emin olun.
- Aynı sonuca ulaşıldığını doğrulamak için AI sistemini yeni veri setlerinde test edin.

Düzeltici

- Yapay zeka sistemini yeniden eğitin (örneğin giriş verilerini iyileştirerek, yanlış pozitif ve negatiflerin farklı dengelerini veya farklı öğrenme algoritmalarını kullanarak).
- Ayrımcı model performansıyla ilgili olarak AI sistem geliştiricilerini yeniden eğitin.
- Yapay zeka tarafından verilen kararı değiştirin ve yanlışlıktan diğer kişilerin etkilenip etkilenmediğini değerlendirin.

Yanlılık (önyargı) ve ayrımcılık risklerini nasıl ele almalıyız?

Yapay zeka sistemleri, dengesiz olabilecek ve/veya ayrımcılığı yansıtabilecek verilerden öğrendiği için, cinsiyet, ırk, yaş, sağlık, din, engellilik, cinsel yönelim veya diğer özelliklerine göre insanlar üzerinde ayrımcı etkileri olan çıktılar üretebilir.

Yapay zeka sistemlerinin verilerden öğrenmesi, çıktılarının ayrımcı etkilere yol açmayacağını garanti etmez. Yapay zeka sistemlerini eğitmek ve test etmek için kullanılan veriler ile bunların tasarlanma ve kullanılma biçimleri, belirli gruplara daha az olumlu davranan yapay zeka sistemlerine yol açabilir.

Veri koruma kanunu, kişisel verilerin korunması hakkını toplumdaki işleviyle dengelemeyi amaçlamaktadır. Ayrımcılığa ve önyargıya yol açan kişisel verilerin işlenmesi, bu işlemenin adilliğini etkileyecektir. Bu, ayrımcılık yapmama hakkı da dahil olmak üzere, kişilerin hak ve özgürlüklerine yönelik risklerin yanı sıra adalet ilkesine uyum sorunlarına yol açar. Ayrıca GDPR, organizasyonların 'gerçek kişiler üzerinde ayrımcı etkileri' önlemek için önlemler alınması gerektiğini özellikle belirtmektedir.

Ek olarak, Birleşik Krallık'ın ayrımcılık karşıtı yasal çerçevesi, özellikle Birleşik Krallık Eşitlik Yasası 2010, veri koruma kanununun yanında yer alır ve bir dizi organizasyon için geçerlidir. Buna devlet daireleri, hizmet sağlayıcılar, işverenler, eğitim sağlayıcıları, ulaşım sağlayıcıları, dernekler ve üyelik organları ile kamu işlevlerinin sağlayıcıları dahildir. İster bir insan isterse otomatikleştirilmiş bir karar verme sistemi (veya ikisinin bir kombinasyonu) tarafından oluşturulmuş olsun, kişilere ayrımcılığa karşı koruma sağlar.

Bu bölümde, yapay zeka bağlamında, kişileri sınıflandırmak veya kişiler hakkında bir tahminde bulunmak için kullanılan makine öğrenimi (ML) sistemlerinin nasıl ayrımcılığa yol açabileceğine odaklanarak, bunun pratikte ne anlama geldiğini

araştırıyoruz. Ayrıca, bu riski yönetmek için benimseyebileceğiniz bazı teknik ve organizasyonel önlemleri araştırıyoruz.

Bir AI sistemi neden ayrımcılığa yol açabilir?

Varsayımsal bir senaryo alalım:

Örnek

Bir banka, potansiyel müşterilerin kredi riskini hesaplamak için bir yapay zeka sistemi geliştirir. Banka, kredi başvurularını onaylamak veya reddetmek için AI sistemini kullanacaktır.

Sistem, daha önce borç alanlar hakkında meslekleri, gelirleri, yaşları ve kredilerini geri ödeyip ödemedikleri gibi bir dizi bilgiyi içeren geniş bir veri seti üzerinde eğitilir.

Test sırasında banka, olası herhangi bir cinsiyet önyargısını kontrol etmek istiyor ve AI sisteminin kadınlara daha düşük kredi puanı verme eğiliminde olduğunu tespit ediyor.

Bunun olmasının iki ana nedeni vardır.

Biri dengesiz eğitim verileridir. Eğitim verilerindeki farklı cinsiyetlerin oranı dengeli olmayabilir. Örneğin, eğitim verileri, geçmişte daha az kadın kredi başvurusunda bulunduğundan ve bu nedenle bankanın kadınlar hakkında yeterli veriye sahip olmadığı için, erkek borçluların daha büyük bir oranını içerebilir.

AI algoritması, eğitildiği ve üzerinde test edildiği verilere en uygun olacak şekilde tasarlanmış bir istatistiksel model oluşturacaktır. Erkekler eğitim verilerinde fazla temsil ediliyorsa, model, erkekler için geri ödeme oranlarını öngören istatistiksel ilişkilere daha fazla ve kadınlar için farklı olabilecek geri ödeme oranlarını öngören istatistiksel kalıplara daha az dikkat edecektir.

Başka bir deyişle, **istatistiksel** olarak "daha az önemli" oldukları için, eğitim veri setindeki kadınların kredilerini erkeklerden daha fazla ödeme olasılığı ortalama olarak daha yüksek olsa bile, model sistematik olarak kadınlar için daha düşük kredi geri ödeme oranlarını öngörebilir.

Bu sorunlar, eğitim verilerinde yeterince temsil edilmeyen herhangi bir popülasyon için geçerli olacaktır. Örneğin, bir yüz tanıma modeli, belirli bir etnik kökene ve cinsiyete (örneğin beyaz erkeklere) ait orantısız sayıda yüz üzerinde eğitilirse, o gruptaki kişileri tanırken daha iyi ve diğerlerinde daha kötü performans gösterecektir.

Diğer bir sebep ise eğitim verilerinin geçmiş ayrımcılığı yansıtabilmesidir. Örneğin, geçmişte kadınlardan gelen kredi başvuruları

önyargı nedeniyle erkeklerden daha sık reddedildiyse, bu tür eğitim verilerine dayalı herhangi bir modelin aynı ayrımcılık modelini yeniden üretmesi muhtemeldir.

Ayrımcılığın tarihsel olarak önemli bir sorun olduğu belirli alanlarda, polisin genç siyah erkekleri durdurup araması veya geleneksel olarak erkek rolleri için işe alma gibi bu sorunu daha şiddetli yaşama olasılığı daha yüksektir.

Bu sorunlar, eğitim verileri cinsiyet veya ırk gibi korunan herhangi bir özellik içermese bile ortaya çıkabilir. Eğitim verilerindeki çeşitli özellikler genellikle korunan özelliklerle, örneğin meslekle yakından ilişkilidir. Bu "vekil değişkenler – proxy variables", tasarımcılar bunu istemese bile, modelin bu özelliklerle ilişkili ayrım kalıplarını yeniden üretmesini sağlar.

Bu problemler herhangi bir istatistiksel modelde ortaya çıkabilir. Bununla birlikte, daha fazla sayıda özellik içerebildikleri ve korunan özelliklerin proxyleri olan karmaşık özellik kombinasyonlarını tanımlayabildikleri için AI sistemlerinde ortaya çıkma olasılıkları daha yüksektir. Birçok modern ML yöntemi, yüksek boyutlu verilerde doğrusal olmayan kalıpları ortaya çıkarmada daha iyi oldukları için geleneksel istatistiksel yaklaşımlardan daha güçlüdür. Ancak bunlar aynı zamanda ayrımcılığı yansıtan kalıpları da içerir.

ML modellerinde ayrımcılık riskini azaltmak için teknik yaklaşımlar nelerdir?

Ayrımcılık, gerçekçi olarak teknoloji aracılığıyla genel bir "düzeltilemez" sorun olsa da, yapay zekaya dayalı ayrımcılığı azaltmak için çeşitli yaklaşımlar vardır. Bilgisayar bilimcileri ve diğerleri, ML modellerinin farklı gruplardan kişilere potansiyel olarak ayrımcı yollarla nasıl davrandığını ölçmek için farklı matematiksel teknikler geliştiriyorlar. Bu alan genellikle algoritmik 'adalet' olarak adlandırılır. Bu tekniklerin birçoğu geliştirmenin ilk aşamalarında olmasına ve piyasaya hazır olmamalarına rağmen, AI geliştiricilerinin sistemlerinden kaynaklanan potansiyel ayrımcılığı ölçmek ve azaltmak için alabilecekleri ve almaları gereken bazı temel yaklaşımlar vardır.

Dengesiz eğitim verileri söz konusu olduğunda, popülasyonun yetersiz/fazla temsil edilen alt setleri hakkında veri ekleyerek veya çıkararak (örneğin, kadınlardan alınan kredi başvurularına ilişkin daha fazla veri noktası ekleyerek) dengelemek mümkün olabilir.

Alternatif olarak, örneğin biri erkekler, diğeri kadınlar için ayrı modeller eğitebilir ve bunları her alt grupta mümkün olduğunca iyi performans gösterecek şekilde tasarlayabilirsiniz. Bununla birlikte, bazı durumlarda, farklı korunan sınıflar için farklı modeller oluşturmanın kendisi, ayrımcılık yapmama yasasının ihlali olabilir (örneğin, erkekler ve kadınlar için farklı araba sigortası primleri).

Eğitim **verilerinin geçmiş ayrımcılığı yansıttığ**ı durumlarda; verileri değiştirebilir, öğrenme sürecini değiştirebilir veya eğitimden sonra modeli değiştirebilirsiniz.

Bu tekniklerin etkili olması için, sonuçları ölçebileceğiniz bir veya daha fazla matematiksel "adalet" ölçüsü seçmeniz gerekir.

Bu önlemler üç geniş kategoride gruplandırılabilir:

- Anti-sınıflandırma
- Sonuç / hata paritesi
- Eşit kalibrasyon

Anti-sınıflandırma, bir modelin bir sınıflandırma veya tahmin yaparken korunan özellikleri hariç tutması durumunda adil olduğu durumdur. Bazı sınıflandırma karşıtı yaklaşımlar, korunan özellikler için proxy'leri belirlemeye ve hariç tutmaya çalışır (örneğin, tek cinsiyetli bir okula katılım). Tüm olası proxy'leri kaldırmak, çok az tahmin edilebilir kullanışlı özellik bırakabileceği için bu pratik olmayabilir. Ayrıca, daha fazla veri toplama ve analiz olmaksızın, belirli bir değişkenin (veya değişkenlerin kombinasyonunun) korunan bir özellik için bir proxy olup olmadığını bilmek genellikle zordur.

Sonuç / hata paritesi, farklı korunan grupların üyelerinin model tarafından nasıl ele alındığını karşılaştırır. **Sonuç paritesi** ile bir model, farklı gruplara eşit sayıda olumlu veya olumsuz sonuç veriyorsa adildir. **Hata paritesi** ile bir model, farklı gruplara eşit sayıda hata veriyorsa adildir. Hata paritesi, yanlış pozitif veya yanlış negatif paritesine bölünebilir (daha fazla ayrıntı için istatistiksel doğrulukla ilgili bölüm 2.3'e bakın).

Eşit kalibrasyon, kalibrasyon, modelin bir şeyin olma olasılığına ilişkin tahmininin, olayın gerçek sıklığıyla ne kadar yakından eşleştiğini ölçer. 'Eşit kalibrasyon'a göre bir model, farklı korunan grupların üyeleri arasında eşit olarak kalibre edilmişse adildir. Örneğin, bir sınıflandırma modeli kredi başvuru sahiplerini düşük, orta veya yüksek geri ödeme şansı olanlara ayırıyorsa, her bir risk kategorisinde **geri ödeme yapan** erkek ve kadın başvuru sahiplerinin eşit oranları olmalıdır. Bu, farklı risk kategorilerinde kadın ve erkek oranlarının eşit olması gerektiği anlamına gelmez. Örneğin, kadınlar gerçekten erkeklerden daha yüksek geri ödeme oranlarına sahipse, düşük risk kategorisinde erkeklerden daha fazla kadın olabilir.

Ne yazık ki, bu farklı önlemler çoğu zaman birbiriyle uyumsuz olabilir ve bu nedenle herhangi bir belirli yaklaşımı/yaklaşımları seçmeden önce herhangi bir çelişkiyi dikkatlice düşünmeniz gerekir. Örneğin:

- Sonuçların gerçek dağılımının farklı korunan gruplar arasında eşit olduğu nadir durumlar dışında, eşit kalibrasyon sonuç veya hata paritesi ile uyumsuzdur.

- Anti-sınıflandırma önlemlerin gerektirdiği gibi, korunan özellikleri kaldırırken sonuç denkliğine ulaşmaya çalışmak, öğrenme algoritmasının sonuçları eşitleyen bir model oluşturmak için alakasız proxy'leri bulmasına ve kullanmasına neden olabilir ve bu adil olmayabilir.

AI sistemlerinde ayrımcılığı değerlendirmek ve ele almak için özel kategori verilerini işleyebilir miyiz?

Yukarıda açıklanan tekniklerin çoğu, popülasyonun temsili bir örneğinin kişisel verilerini içeren bir veri setine erişiminizin olmasını gerektirir. Verilerde temsil edilen her kişi için, 2010 Eşitlik Yasası'nda ana hatlarıyla belirtilenler gibi korunan ilgili özelliklere ilişkin etiketlere ihtiyacınız vardır. Daha sonra, korunan özellikleri içeren bu veri setini, sistemin korunan her grupla nasıl performans gösterdiğini test etmek ve ayrıca potansiyel olarak modeli daha adil bir şekilde çalışması için yeniden eğitmek için kullanabilirsiniz.

Bu tür bir analiz yapmadan önce, verileri bu amaçlarla işlemek için uygun bir yasal dayanağa sahip olduğunuzdan emin olmanız gerekir. Test ettiğiniz ayrımcılık türlerine bağlı olarak farklı veri koruma hususları vardır. Yaşa veya cins / cinsiyete göre ayrımcı etki için bir sistemi test ediyorsanız, veri koruma kanununda 'özel kategori verileri' olarak sınıflandırılmadıkları için bu korunan özelliklerin işlenmesi için özel bir veri koruma koşulu yoktur. Yine de aşağıdakileri göz önünde bulundurmanız gerekir:

- İşlemlerin bir bütün olarak teşkil ettiği yasallık, adalet ve risklerle ilgili daha geniş sorular.
- Verilerin her durumda özel kategori verileri olma veya işleme sırasında böyle olma olasılığı (yani, işleme sağlık veya genetik durumla ilgili herhangi bir verinin analiz edilmesini veya çıkarılmasını içeriyorsa).

Ayrıca, bir kişiin fiziksel, fizyolojik veya davranışsal özelliklerine ilişkin belirli teknik işlemler sonnucunda ortaya çıkan kişisel verilerle uğraşırken ve kişinin benzersiz (unique) kimliğine izin verdiğiniz veya onayladığınız zaman, verilerin biyometrik veri olduğunu da unutmayın.

Bir kişiyi benzersiz şekilde tanımlamak **amacıyla** biyometrik verileri kullandığınızda, bu aynı zamanda özel kategori verileridir.

Bu nedenle, AI sisteminiz ayrımcılığı test etmek ve azaltmak için biyometrik veriler kullanıyorsa ancak veri setindeki kişilerin kimliğini doğrulamak veya bunlarla ilgili herhangi bir karar vermek amacıyla kullanmıyorsa, biyometrik veriler 9. Madde kapsamında yer almaz. Veriler hala GDPR kapsamında biyometrik veri olarak kabul edilir ancak özel kategori verileri değildir.

Benzer şekilde, kişisel veriler bir kişinin benzersiz kimliğine izin vermiyor veya onaylamıyorsa, biyometrik veri (veya özel kategori verisi) değildir.

Ancak, Eşitlik Yasası'nda özetlenen bazı korunan özellikler özel kategori verileri olarak sınıflandırılır. Bunlar arasında ırk, din veya inanç ve cinsel yönelim yer

almaktadır. Ayrıca, kişinin sağlığı hakkında bilgi verebileceği için şu ana kadar engellilik, hamilelik ve cinsiyet değiştirmeyi de içerebilirler. Benzer şekilde, hemcins birlikteliği yakın zamana kadar yalnızca aynı cinsiyetten çiftler için mevcut olduğundan, birinin hemcins birlikteliği içinde olduğunu gösteren veriler, dolaylı olarak cinsel yönelimlerini ortaya çıkarabilir.

Bu özelliklere dayalı olarak ayrımcı etki için bir AI sistemini test ediyorsanız, muhtemelen özel kategori verilerini işlemeniz gerekecektir. Bunu yasal olarak yapabilmek için, 6. Madde kapsamında yasal bir dayanağa sahip olmanın yanı sıra, GDPR'nin 9. Maddesindeki koşullardan birini karşılamanız gerekir. Bunlardan bazıları ayrıca, DPA 2018'in 1. Çizelgesi'nde bulunabilecek olan Birleşik Krallık Hukukunda ek dayanak veya yetkilendirme gerektirir.

Özel kategori verilerinin işlenmesi için bu koşullardan hangisinin (varsa) uygun olduğu, kişisel durumunuza bağlıdır.

Örnek: ayrımcılığı değerlendirmek, fırsat eşitliğini belirlemek ve teşvik etmek veya sürdürmek için AI'da özel kategori verilerinin kullanılması

İşe alım kararlarına yardımcı olmak için CV puanlı AI sistemi kullanan bir organizasyonun, sistemini dini veya felsefi inançlar tarafından ayrımcılık yapıp yapmadığını test etmesi gerekiyor.

Sistem gerçekten orantısız negatif sonuçlar mı yoksa hatalı tahminler mi ürettiğini değerlendirmek için iş başvurusunda bulunanların dini inançlarını toplar.

Organizasyon; Madde 9(2)(g)'deki önemli kamu yararı koşuluna ve DPA 2018 Ek 1 (8)'deki fırsat eşitliği veya muamele koşuluna güvenmektedir. Bu hüküm, belirli korunan gruplar arasında fırsat veya muamele eşitliğinin varlığını veya yokluğunu tanımlamak veya gözden geçirmek için kullanılabilir ve bu eşitliğin düzenlenmesi veya sürdürülmesini sağlamak amacıyla kullanılabilir.

Örnek: Araştırma amacıyla AI'da ayrımcılığı değerlendirmek için özel kategori verilerinin kullanılması

Bir üniversite araştırmacısı, bir araştırma projesinin parçası olarak piyasada bulunan yüz tanıma sistemlerinin farklı ırk veya etnik kökene sahip insanların yüzlerinde farklı performans gösterip göstermediğini araştırıyor.

Bunu yapmak için araştırmacı, sistemin test edileceği mevcut bir yüz veri setine ırk etiketleri atar ve böylece özel kategori verilerini işler. DPA 2018'in Çizelge 1 paragraf 4 ile okunan Madde 9(2)(j)'deki arşivleme, araştırma ve istatistik koşuluna güvenirler.

Son olarak, potansiyel olarak ayrımcı yapay zekayı değerlendirmek ve geliştirmek için kullandığınız korunan özellikler, orijinal olarak farklı bir amaç için işlenmişse, aşağıdakileri göz önünde bulundurmalısınız:

- Yeni amacınızın asıl amaçla uyumlu olup olmadığı.
- Gerekirse nasıl yeni onay alacağınız. Örneğin, veriler başlangıçta rızaya dayalı olarak toplandıysa, yeni amaç uyumlu olsa bile, yeni amaç için yine de yeni bir rıza almanız gerekir.
- Yeni amaç uyuşmuyorsa, nasıl rıza isteyeceğiniz.

Mevzuattaki ilgili hükümler

Bknz. GDPR'ın 9 madde ve 51 ve 56 gerekçesi (dış bağlantı)

Bknz. DPA 2018'in 1. Program (dış bağlantı)

İlave okuma – ICO kılavuzu

GDPR Kılavuzu'ndaki amaç sınırlaması ve özel kategori verileri hakkındaki kılavuzumuzu okuyun.

Peki ya özel kategori verileri, ayrımcılık ve otomatik karar alma süreci?

Yapay zeka sistemlerinin potansiyel ayrımcı etkilerini değerlendirmek için özel kategori verilerinin kullanılması, genellikle veri koruma kanunu kapsamında otomatik karar vermeyi teşkil etmez. Bunun nedeni, doğrudan kişiler hakkında herhangi bir karar vermeyi içermemesidir.

Benzer şekilde, ayrımcı etkilerini azaltmak için daha çeşitli bir popülasyondan alınan verilerle ayrımcı bir modeli yeniden eğitmek, kişilerle ilgili doğrudan kararlar vermeyi kapsamaz ve bu nedenle yasal veya benzer derecede önemli bir etkiye sahip bir karar olarak sınıflandırılmaz.

Ancak bazı durumlarda, yapay zeka modelini daha çeşitli bir eğitim setiyle yeniden eğitmek, bu modelin ayrımcı etkisini yeterince hafifletmek için yeterli olmayabilir. Bazı yaklaşımlar; bir tahmin yaparken korunan özellikleri **göz ardı** ederek bir modeli adil hale getirmeye çalışmak yerine, potansiyel olarak dezavantajlı grupların üyelerinin korunmasını sağlamak için sınıflandırma yaparken bu tür özellikleri doğrudan **dahil** eder.

Örneğin, iş başvurularını sıralamak için yapay zeka sistemi kullanıyorsanız, bir kişinin maluliyetini göz ardı eden bir model oluşturmaya çalışmak yerine, sistemin bunlara karşı ayrımcılık yapmaması için maluliyet durumlarını dahil etmek daha etkili olabilir. Engellilik durumunun otomatik karara bir girdi olarak dahil olmaması, sistemin engellilere karşı daha fazla ayrımcılık yapabileceği anlamına gelebilir, çünkü bu durum, durumlarının tahmin yapmak için kullanılan diğer özellikler üzerindeki etkisini etkilemez.

Bu yaklaşım, özel kategori verilerini kullanarak, önemli etkileri olan, tamamen otomatik bir şekilde kişiler hakkında kararlar almak anlamına gelir. Kişiden açık onayınız olmadıkça veya DPA'nın 1. Çizelgesinde belirtilen önemli kamu yararı koşullarından birini yerine getiremediğiniz sürece, bu GDPR kapsamında yasaktır.

Çizelge 1'deki hangi koşulların geçerli olabileceğini dikkatlice değerlendirmeniz gerekir. Örneğin, yukarıda bahsedilen fırsat eşitliği izleme koşuluna bu tür durumlarda güvenilemez, çünkü işleme, belirli bir kişi hakkında karar verme amacıyla gerçekleştirilmektedir. Bu nedenle, bu tür yaklaşımlar yalnızca, Çizelge 1'deki farklı bir, önemli kamu yararı koşuluna dayandığı takdirde yasal olacaktır.

Yapay zeka kullanımımız üzerinden kazara özel kategori verileri çıkarsa ne olur?

İçinde yaşadığınız posta kodu gibi korunmayan özelliklerin, ırk gibi korunan bir özelliğin proxy'leri olduğu birçok durum vardır. Makine öğrenimindeki "derin" öğrenme gibi son gelişmeler, yapay zeka sistemlerinin; dünyadaki görünüşte alakasız verilere yansıyan kalıpları, algılamasını daha da kolaylaştırdı. Ne yazık ki, bu aynı zamanda, açık olmayan yollarla korunan niteliklerle ilişkilendirilebilecek karmaşık özellik kombinasyonları kullanılarak ayrımcılık kalıplarının tespit edilmesini de içerir.

Örneğin, işe alım kararlarına yardımcı olmak amacıyla iş başvurularını puanlamak için kullanılan bir AI sistemi, daha önce başarılı olan adayların örnekleri üzerinden eğitilebilir. Uygulamanın kendisinde bulunan bilgiler ırk, engellilik veya akıll sağlığı gibi korunan özellikleri içermeyebilir.

Ancak, modeli eğitmek için kullanılan, çalışan örnekler bu gerekçelerle ayrımcılığa uğradıysa (örneğin performans incelemelerinde sistematik olarak yetersiz değerlendirilerek), algoritma, tasarımcının asla istememesine rağmen iş başvurusunda bulunan proxy verilerden bu özellikleri çıkararak bu ayrımcılığı yeniden üretmeyi öğrenebilir.

Bu nedenle, modelinizde korumalı özellikler kullanmasanız bile, bu korunan özelliklere dayalı olarak ayrımcılık kalıpları tespit eden ve çıktılarında bunları yeniden üreten bir modeli farkına varmadan kullanmanız çok olasıdır. Yukarıda açıklandığı gibi, bu korunan özelliklerden bazıları aynı zamanda özel kategori verileridir.

Özel kategori verileri, özel kategorileri 'açığa çıkaran veya ilgilendiren' kişisel veriler olarak tanımlanmaktadır. Model, özel bir kategoriyi yeterince açığa çıkaran belirli özellik kombinasyonlarını kullanmayı öğrenirse, model özel kategori verilerini işleme koyabilir.

Özel kategori verileriyle ilgili kılavuzumuzda belirtildiği gibi; özel kategori verileri çıkarmak **amacıyla** profilleme kullanıyorsanız, bu, çıkarımların yanlış olup olmadığına bakılmaksızın özel kategori verileridir.

Ayrıca, yukarıda belirtilen nedenlerden dolayı, modelinizin başka bir (özel kategori olmayan veri) çıkarıma ara adım olarak özel bir kategori oluşturması da söz konusu olabilir. Modelinizin bunu yapıp yapmadığını sadece modele ve ürettiği çıktılara bakarak anlayamayabilirsiniz. Bunu yapmak istememiş olsanız bile yüksek doğrulukla yapabilir.

Kişisel verilerle makine öğrenimi kullanıyorsanız, tahminlerde bulunmak için modelinizin korumalı özellikler ve/veya özel kategori verileri çıkarma olasılığını proaktif olarak değerlendirmeli ve sistemin yaşam döngüsü boyunca bu olasılığı aktif olarak izlemelisiniz. Potansiyel olarak çıkarılan özellikler özel kategori verileriyse, işleme için uygun bir Madde 9 koşuluna sahip olduğunuzdan emin olmalısınız.

Yukarıda belirtildiği gibi, böyle bir model yalnızca otomatik bir şekilde yasal veya benzer şekilde önemli kararlar almak için kullanılıyorsa, bu yalnızca kişinin rızasına sahipseniz veya önemli kamu yararı koşulunu (ve Çizelge-1'deki uygun bir hükmü) karşılıyorsanız yasaldır.

İlave okuma – ICO kılavuzu

Daha fazla bilgi için özel kategori verileriyle ilgili kılavuzumuzu okuyun

Bu riskleri azaltmak için ne yapabiliriz?

ML sistemlerinde ayrımcı sonuç riskini yönetmeye yönelik en uygun yaklaşım, faaliyet gösterdiğiniz belirli alana ve duruma bağlı olacaktır.

Herhangi bir AI uygulama yaşam döngüsünün en başından itibaren önyargı ve ayrımcılığın azaltılmasına yönelik yaklaşımınızı belirlemeli ve belgelemelisiniz, böylece tasarım ve inşa aşamasında uygun önlemleri ve teknik önlemleri hesaba katabilir ve uygulayabilirsiniz.

Yüksek kaliteli eğitim ve test verilerinin satın alınması ve yasal olarak işlenmesi için net politikalar ve iyi uygulamalar oluşturmak, özellikle dahili olarak yeterli veriye sahip değilseniz önemli olacaktır. İster dahili ister harici olarak temin edilmiş olsun, verilerin ML sistemine uyguladığınız popülasyonu temsil ettiğinden emin olmalısınız (yukarıda belirtilen nedenlerden dolayı bu, adaleti sağlamak için yeterli olmayacaktır). Örneğin, Birleşik Krallık'ta faaliyet gösteren bir ana cadde bankası için, eğitim verileri en son Nüfus Sayımı ile karşılaştırılabilir.

Üst yönetiminiz, ayrımcılık riskini yönetmek için seçilen yaklaşımın imzalanmasından ve veri koruma yasasına uygunluğundan sorumlu olmalıdır. Teknoloji liderlerinden ve diğer dahili veya harici konu uzmanlarından gelen uzmanlıktan yararlanabilseler de, sorumlu olmak için üst düzey liderlerinizin farklı yaklaşımların sınırlamaları ve avantajları hakkında yeterli bir anlayışa sahip olmaları gerekir. Bu aynı zamanda DPO'lar ve gözetim işlevlerindeki üst düzey personel için de geçerlidir, çünkü ayrımcılık riskini azaltmak için uygulamaya

konulan her türlü önlem ve önlemin uygunluğu konusunda sürekli tavsiye ve rehberlik sağlamaları beklenecektir.

Çoğu durumda, farklı risk yönetimi yaklaşımları arasında seçim yapmak, ödünleşmeler gerektirecektir ('AI ile ilgili ödünleşmeler nelerdir ve bunları nasıl yönetmeliyiz?' bölümüne bakın). Bu, farklı korunan özellikler ve gruplar için korumalar arasında seçim yapmayı içerir. Seçtiğiniz yaklaşımı belgelemeniz ve gerekçelendirmeniz gerekir.

Teknik yaklaşımlar tarafından yönlendirilen ödünleşmeler teknik olmayan personel için her zaman açık değildir, bu nedenle veri bilimcileri bunları işletme sahiplerine ve risk yönetimi ve veri koruma uyumundan sorumlu personele proaktif bir şekilde vurgulamalı ve açıklamalıdır. Teknik liderleriniz, algoritmik "adalet" yaklaşımlarını bilgilendirmek üzere, korunan özellikler için bilinen proxyler de dahil olmak üzere alana özgü bilgi arayışında proaktif olmalıdır.

Ayrımcılıkla mücadele önlemlerinin sağlam testlerini üstlenmeli ve makine öğrenimi sisteminizin performansını sürekli olarak izlemelisiniz. Risk yönetimi politikalarınız, hem devreye alımdan önce hem de uygun olduğunda bir güncellemeden sonra bir ML sisteminin nihai doğrulaması için hem süreci hem de sorumlu kişiyi açıkça belirtmelidir.

Ayrımcılığın izlenmesi amacıyla, organizasyon politikalarınız, seçilen Anahtar Performans Metriklerine karşı tüm sapma toleranslarını ve ayrıca yükseltme ve sapma araştırma prosedürlerini belirlemelidir. Ayrıca, üzerinde ML sisteminin kullanılmasını durdurması gereken varyans sınırlarını da net bir şekilde belirlemelisiniz.

Geleneksel karar verme sistemlerini yapay zeka ile değiştiriyorsanız, belirli bir süre için her ikisini de aynı anda çalıştırmayı düşünmelisiniz. İki sistem arasındaki farklı korunan gruplar için karar türlerindeki (örneğin, kredi kabul veya reddetme) önemli farklılıkları ve yapay zeka sisteminin nasıl performans göstereceği ve pratikte nasıl çalıştığına ilişkin farklılıkları araştırmalısınız.

Veri koruma yasasının gerekliliklerinin ötesinde, çeşitlilik içeren bir iş gücü; yapay zeka sistemlerinde ve daha genel olarak organizasyonda önyargı ve ayrımcılığın belirlenmesi ve yönetilmesinde güçlü bir araçtır.

Son olarak, bu, en iyi uygulama ve teknik yaklaşımların gelişmeye devam ettiği bir alandır. En iyi uygulamaları izlemeye devam ettiğinizden ve personelinizin sürekli olarak uygun şekilde eğitildiğinden emin olmak için zaman ve kaynaklara yatırım yapmalısınız. Bazı durumlarda yapay zeka, geleneksel karar verme süreçlerinde mevcut ayrımcılığı ortaya çıkarmak ve ele almak için bir fırsat sağlayabilir ve altta yatan ayrımcı uygulamaları ele almanıza izin verebilir.

Diğer kaynaklar

Equality Act 2010 (dış bağlantı)

European Charter of Fundamental Rights (dış bağlantı)

Kontrol örnekleri

Risk Bildirimi

AI sistemleri tarafından alınan ayrımcı çıktılar veya kararlar, belirli gruplardaki kişiler için istatistiksel olarak yanlış/haksız kararlara yol açabilir.

Önleyici

- Bir AI sisteminin sürekli eğitimi, test edilmesi veya değerlendirilmesi için kullanılan tüm kişisel verilerin mümkün olduğunca nasıl doğru, net, ilgili, temsili, eksiksiz ve güncel olduğunu açıklayan bir veri yönetişim çerçevesini uygulamaya koyun.
- Yapay zeka geliştiricilerinin, yapay zeka sistemlerindeki önyargı ve ayrımcılığı tanımlayıp ele alabilmeleri için eğitimi ve ilgili yetkinlik değerlendirmelerini tamamladıklarından emin olun.
- Kilit paydaşlar için eğitim sağlayın ve ilgili personelin nasıl belirlendiğini belgeleyin (örn. üst yönetim, risk yöneticileri, denetim).
- İstatistiksel doğruluğu etkileyen değişikliklerin yetkili kişiler tarafından yapıldığından ve imzalandığından emin olmak için AI sistemlerinin geliştirilmesi ve devreye alımı için erişim yönetimi kontrollerinizi ve görev ayrımını belgeleyin. Bu kontrollerin nasıl izlendiğine dair kanıtları koruyun.
- AI sistemlerinin geliştirilmesi/kullanımı için onay yetkisi seviyelerini belgeleyin. Uygun onayın kanıtını koruyun.
- DPIA'nıza, ayrımcılık riskinin kapsamlı bir değerlendirmesini ve bunu önlemek için yürürlükte olan hafifletici etkenleri/kontrolleri dahil edin.
- Üçüncü taraflarla ilişkiler için belgelenmiş politikaları/süreçleri ve tamamlanan durum tespitinin kanıtını sağlayın.
- Yapay zeka sistem tasarımının tipik örneği / meslektaş incelemesi için belgelenmiş bir süreç sürdürün. İncelemenin tamamlandığına dair kanıtları sağlayın.
- Herhangi bir yapay zeka sistemi veya değişikliğin canlıya alınmadan önce uygulama öncesi testini gerçekleştirmek için belgelenmiş bir politika/süreç uygulayın. Testin tamamlandığına dair kanıtları ve test(ler)in sonuçlarını sağlayın.
- AI sistemi içinde kullanımdan önce eğitim / test verilerinin çeşitliliğine / temsiline ilişkin onay ve tasdik düzeylerini belgeleyin. Uygun onayın kanıtını sağlayın.

Tespit Edici

- Uygun önlemleri kullanarak algoritmik adaleti düzenli olarak izleyin.
- AI sistemi içinde kullanımdan önce eğitim / test verilerinin çeşitliliğine / temsiline ilişkin onay ve tasdik düzeylerini belgeleyin. Uygun onayın kanıtını sağlayın.
- Model performansını en son verilerle düzenli olarak gözden geçirin.

Düzeltici

- Kapsamlı analiz / gerekçelendirme dahil olmak üzere, yetersiz / fazla temsil edilen gruplar hakkında veri ekleyin veya kaldırın.
- Modeli adalet kısıtlamalarıyla yeniden eğitin.
- Model tasarımcılarını ayrımcı model performansıyla ilgili olarak yeniden eğitin.

AI'da güvenliği ve veri minimizasyonunu nasıl değerlendirmeliyiz?

Genel bir bakış

Yapay zeka sistemleri, bilinen güvenlik risklerini artırabilir ve bunların yönetilmesini zorlaştırabilir. Ayrıca, veri minimizasyonu ilkesine uyum için zorluklar da ortaya koyarlar.

Yapay zekanın artırabileceği iki güvenlik riski şunlardır:

- AI sistemlerini eğitmek için genellikle gerekli olan büyük miktarda kişisel verinin kaybolması veya kötüye kullanılması potansiyeli.
- AI ile ilgili yeni kod ve altyapının tanıtılmasının bir sonucu olarak ortaya çıkacak yazılım güvenlik açıkları potansiyeli.

Default olarak, yapay zekayı geliştirmeye ve dağıtmaya yönelik standart uygulamalar, büyük miktarda verinin işlenmesini içerir. Bunun veri minimizasyon ilkesine uymama riski vardır. Hem veri minimizasyonunu hem de etkili yapay zeka geliştirme ve dağıtımını mümkün kılan bir dizi teknik mevcuttur.

Detaylı olarak

- Yapay zeka ne tür güvenlik riskleri getiriyor?
- Yapay zeka modelleri için ne tür gizlilik saldırıları uygulanır?
- AI modellerinde gizlilik saldırılarının risklerini yönetmek için hangi adımları atmalıyız?
- AI sistemleri için hangi veri minimizasyonu ve gizlilik koruma teknikleri uygundur?

Yapay zeka ne tür güvenlik riskleri getiriyor?

Kişisel verileri; yetkisiz veya hukuka aykırı olarak işlenmesine, yanlışlıkla kaybolmasına, yok olmasına veya zarar görmesine karşı uygun güvenlik seviyeleri sağlayacak şekilde işlemelisiniz. Bu bölümde, bilinen riskleri, daha kötü ve kontrol edilmesi daha zor hale getirerek yapay zekanın güvenliği nasıl olumsuz etkileyebileceğine odaklanıyoruz.

Güvenlik gereksinimlerimiz nelerdir?

Güvenlik için "herkese uygun" bir yaklaşım yoktur. Almanız gereken uygun güvenlik önlemleri, belirli işleme faaliyetlerinden kaynaklanan risklerin düzeyine ve türüne bağlıdır.

Herhangi bir kişisel veriyi işlemek için yapay zekayı kullanmanın güvenlik riski profiliniz üzerinde önemli etkileri vardır ve bunları dikkatli bir şekilde değerlendirmeniz ve yönetmeniz gerekir.

Bazı sonuçlar, örneğin makine öğrenimi modellerine yönelik düşmanca saldırılar gibi yeni risk türlerinin ortaya çıkmasıyla tetiklenebilir (aşağıdaki X bölümüne bakın).

İlave okuma – ICO kılavuzu

GDPR Kılavuzundaki güvenlik kılavuzumuzu ve veri koruma yasası kapsamında güvenlik hakkında genel bilgi için ICO/NCSC Güvenlik Sonuçlarını okuyun.

Bilgi güvenliği, AI denetim çerçevemizin önemli bir bileşenidir, ancak aynı zamanda bilgi hakları düzenleyicisi olarak işimizin merkezinde yer alır. ICO, yeni GDPR'de belirtilen ek gereksinimleri dikkate almak için genel güvenlik kılavuzunu genişletmeyi planlıyor.

Bu kılavuz, yapay zekaya özgü olmayacak olsa da, yazılım tedarik zinciri güvenliği ve açık kaynaklı yazılım kullanımının artırılması dahil olmak üzere, yapay zeka kullanan organizasyonlarla ilgili bir dizi konuyu kapsayacaktır.

'Geleneksel' teknolojilere kıyasla yapay zeka güvenliğinin farkı nedir?

Yapay zekanın benzersiz özelliklerinden bazıları, veri koruma yasasının güvenlik gereksinimlerine uyumun, hem teknolojik hem de insan açısından diğer daha yerleşik teknolojilere göre daha zor olabileceği anlamına gelir.

Teknolojik bir perspektiften bakıldığında, AI sistemleri, kullanmaya alışık olabileceğiniz daha geleneksel BT sistemlerinde bulunmayan yeni tür karmaşıklıklar sunar. Koşullara bağlı olarak, yapay zeka sistemlerini kullanmanız da büyük olasılıkla üçüncü taraf koduna ve/veya tedarikçilerle olan ilişkilere dayanmaktadır. Ayrıca, mevcut sistemlerinizin, aynı zamanda karmaşık bir şekilde birbirine bağlı olan birkaç yeni ve mevcut BT bileşeniyle entegre edilmesi gerekir.

Bu karmaşıklık, bazı güvenlik risklerini tanımlamayı ve yönetmeyi zorlaştırabilir ve kesinti riski gibi diğerlerini artırabilir.

İnsan perspektifinden bakıldığında, yapay zeka sistemlerinin oluşturulması ve uygulanmasına dahil olan kişilerin, geleneksel yazılım mühendisliği, sistem yönetimi, veri bilimciler, istatistikçiler ve alan uzmanları dahil olmak üzere normalden daha geniş bir geçmişe sahip olması muhtemeldir.

Güvenlik uygulamaları ve beklentileri önemli ölçüde farklılık gösterebilir ve bazıları için daha geniş güvenlik uyumluluğu gereksinimlerinin yanı sıra daha spesifik olarak veri koruma yasasının gereksinimleri konusunda daha az anlayış

olabilir. Kişisel verilerin güvenliği, özellikle daha önce kişisel olmayan verilerle veya araştırma kapasitesiyle yapay zeka uygulamaları geliştiriyorsa, her zaman kilit bir öncelik olmayabilir.

Veri bilimi ve yapay zeka mühendisliğinde kişisel verilerin nasıl güvenli bir şekilde işleneceğine dair yaygın uygulamalar hala geliştirilme aşamasında olduğu için daha fazla komplikasyon ortaya çıkıyor. Güvenlik ilkesine uygunluğunuzun bir parçası olarak, kişisel verileri bir yapay zeka bağlamında kullanırken en son güvenlik uygulamalarını aktif olarak izlediğinizden ve dikkate aldığınızdan emin olmalısınız.

Kişisel verileri işlemek için AI kullandığınızda daha da kötüleşebilecek bilinen tüm güvenlik risklerini listelemek mümkün değildir. Yapay zekanın güvenlik üzerindeki etkisi şunlara bağlıdır:

- Teknolojinin oluşturulma ve uygulanma şekli.
- Onu dağıtan organizasyonun karmaşıklığı.
- Mevcut risk yönetimi yeteneklerinin gücü ve olgunluğu.
- AI sistemi tarafından kişisel verilerin işlenmesinin niteliği, kapsamı, bağlamı, amaçları ve işlemenin sonucunda kişilere yönelik riskler.

Aşağıdaki varsayımsal senaryolar, yapay zekanın daha da kötüleştirebileceği bilinen bazı güvenlik riskleri ve zorlukları hakkında farkındalığı artırmayı amaçlamaktadır.

Anahtar mesajımız, bir AI bağlamında kişisel verilerin güvende olmasını sağlamak için risk yönetimi uygulamalarınızı gözden geçirmeniz gerektiğidir.

Vaka çalışması: eğitim verilerinin izini kaybetmek

ML sistemleri, orijinal işleme kaynaklarından kopyalanıp içe aktarılacak, üçüncü taraflar da dahil olmak üzere çeşitli formatlarda ve yerlerde saklanan, paylaşılan, büyük eğitim ve test verisi setlerine ihtiyaç duyar. Bu onları takip etmeyi ve yönetmeyi zorlaştırabilir.

Örnek

Bir organizasyon, işe alma sürecinin bir parçası olarak üçüncü taraf bir işe alım sorumlusu tarafından sunulan bir yapay zeka sistemini kullanmaya karar verir. Etkili olmak için, organizasyonun benzer önceki işe alım kararları hakkındaki verileri (örneğin satış müdürü) işe alan kişiyle paylaşması gerekir.

Önceden, organizasyon tamamen manuel bir CV tarama süreci kullanıyordu. Bu, bazı kişisel verilerin (örn. adayların CV'leri) paylaşılmasına yol açtı, ancak organizasyon ile işe alım görevlisi arasında büyük miktarda kişisel veri aktarımını içermiyordu.

Organizasyon, bu işlem için uygun bir yasal temele sahip olduğundan emin olmalıdır.

Bunun ötesinde, ek verilerin paylaşımı, aşağıdakiler gibi önemli güvenlik ve bilgi yönetişimi hususlarını gerektiren farklı konumlarda (aşağıya bakınız) depolanan farklı formatlarda birden çok kopya oluşturmayı gerektirebilir:

- Organizasyonun, işe alım firmasının üzerinde çalıştığı açık pozisyonlarla ilgili verileri incelemek ve seçmek için İK ve işe alım verilerini ayrı bir veritabanı sistemine kopyalaması gerekebilir.
- Seçilen veri alt setlerinin kaydedilip dosyalara aktarılması ve ardından sıkıştırılmış biçimde işe alım görevlisine aktarılması gerekir.
- Alındıktan sonra işe alım sorumlusu, dosyaları uzak bir konuma, örneğin buluta yükleyebilir.
- Bulutta, dosyalar temizlenerek bir programlama ortamına yüklenebilir ve AI sisteminin oluşturulmasında kullanılır.
- Hazır olduğunda, veriler daha sonra kullanılmak üzere yeni bir dosyaya kaydedilebilir.
- Hem organizasyon hem de işe alan sorumlu kişiler için; veriler her kopyalandığında ve farklı yerlerde depolandığında; yetkisiz işleme, kayıp, imha ve hasar dahil olmak üzere kişisel veri ihlali riski artar.

Bu örnekte, eğitim verilerinin tüm kopyalarının güvenlik politikaları doğrultusunda paylaşılması, yönetilmesi ve gerektiğinde silinmesi gerekecektir. Birçok işe alım firması hâlihazırda bilgi yönetişimi ve güvenlik politikalarına sahip olsa da, bunlar AI benimsendikten sonra artık amaca uygun olmayabilir ve dolayısı ile gözden geçirilmeli ve gerekirse güncellenmelidir.

Bu durumda ne yapmalıyız?

Teknik ekipleriniz, bir konumdan diğerine tüm hareketleri ve kişisel verilerin depolanmasını kaydetmeli ve belgelemelidir. Bu, uygun güvenlik riski kontrollerini uygulamanıza ve bunların etkinliğini izlemenize yardımcı olacaktır. Hesap verebilirlik ve dokümantasyon gereksinimlerini karşılamak için net denetim yolları da gereklidir.

Ayrıca, kişisel verileri içeren ara dosyaları, artık gerekmedikleri anda silmelisiniz, örneğin sistemler arasında veri aktarmak için oluşturulan dosyaların sıkıştırılmış sürümleri.

Kişilere yönelik riskin olasılığına ve ciddiyetine bağlı olarak, eğitim verileri için kaynağından çıkarılmadan ve dahili veya harici olarak paylaşılmadan önce kimlik gizleme tekniklerini uygulamanız gerekebilir.

Örneğin, başka bir organizasyonla paylaşmadan önce verilerden belirli özellikleri kaldırmanız veya gizlilik artırıcı teknolojiler (privacy enhancing technologies-PET'ler) uygulamanız gerekebilir.

Vaka çalışması: AI sistemleri oluşturmak için kullanılan harici olarak sağlanan yazılımların getirdiği güvenlik riskleri

Çok az sayıda organizasyon tamamen kendi bünyesinde yapay zeka sistemleri kurar. Çoğu durumda, AI sistemlerinin tasarımı, inşası ve çalıştırılması, en azından kısmen, organizasyonun her zaman bir sözleşme ilişkisine sahip olmayabileceği üçüncü taraflarca sağlanabilir.

Kendi makine öğrenimi mühendislerinizi işe alsanız bile, üçüncü taraf çerçevelere ve kod kitaplıklarına önemli ölçüde güvenebilirsiniz. En popüler ML geliştirme çerçevelerinin çoğu açık kaynaklıdır.

Üçüncü taraf ve açık kaynak kodu kullanmak geçerli bir seçenektir. Yapay zeka sisteminin tüm yazılım bileşenlerini sıfırdan geliştirmek, birçok organizasyonun karşılayamayacağı büyük bir zaman ve kaynak yatırımı gerektirir ve özellikle açık kaynak araçlarıyla karşılaştırıldığında, mevcut çerçevelerin etrafında kurulan zengin katkı ve hizmetler ekosisteminden yararlanmaz.

Bununla birlikte, önemli bir dezavantaj, bu standart ML çerçevelerinin genellikle bir BT sistemine halihazırda kurulu olan diğer yazılım parçalarına bağlı olmasıdır. İlgili riskler hakkında bir fikir vermek için, yakın tarihli bir araştırma, en popüler ML geliştirme çerçevelerinin 887.000'e kadar kod satırı içerdiğini ve 137 dış bağımlılığa dayandığını buldu. Bu nedenle, AI uygulamak, bir organizasyonun yazılım yığınında (ve muhtemelen donanımında) ek güvenlik riskleri getirebilecek değişiklikler gerektirecektir.

Örnek

İşveren, Python tabanlı bir ML çerçevesi kullanarak otomatik CV filtreleme sistemini oluşturmak için bir ML mühendisini işe alır. ML çerçevesi, işe alım görevlisinin BT sistemine indirilmesi gereken bir dizi uzman açık kaynaklı programlama kütüphanesine bağlıdır.

Bu kütüphanelerden biri, ham eğitim verilerini, ML modelini eğitmek için gereken formata dönüştürmek için bir yazılım işlevi içerir. Daha sonra işlevin bir güvenlik açığı olduğu keşfedilir. Güvenli olmayan default bir yapılandırma nedeniyle, bir saldırgan, kötü amaçlı kodu eğitim verileri olarak gizleyerek sisteme uzaktan girer ve çalıştırır.

Bu çok uzak bir örnek değil, Ocak 2019'da birçok makine öğrenimi geliştiricisi tarafından kullanılan Python programlama dili için popüler bir kütüphane olan 'NumPy'de böyle bir güvenlik açığı keşfedildi.

Bu durumda ne yapmalıyız?

Yapay zeka sistemleri ister kurum içinde, ister harici olarak veya her ikisinin bir kombinasyonu olsun, bunları güvenlik riskleri açısından değerlendirmeniz

gerekecektir. Şirket içinde geliştirilen herhangi bir kodun güvenliğini sağlamanın yanı sıra, harici olarak tutulan herhangi bir kodun ve çerçevenin güvenliğini de değerlendirmeniz gerekir.

Pek çok açıdan, kodun korunmasına ve güvenlik risklerinin yönetilmesine ilişkin standart gereksinimler, yapay zeka uygulamaları için geçerli olacaktır. Örneğin:

- Harici kod güvenlik önlemleriniz; güvenlik açıklarından haberdar olmak için güvenlik tavsiyelerine abone olmayı içermelidir.
- Dahili kod güvenlik önlemleriniz, kodlama standartlarına bağlı kalmayı ve kaynak kodu inceleme süreçleri oluşturmayı içermelidir.

Yaklaşımınız ne olursa olsun, personelinizin bu güvenlik risklerini ele almak için uygun beceri ve bilgiye sahip olduğundan emin olmanız gerekir.

Ek olarak, makine öğrenimi sistemleri geliştirirseniz, makine öğrenimi geliştirme ortamını mümkün olduğunda BT altyapınızın geri kalanından ayırarak üçüncü taraf koduyla ilişkili güvenlik risklerini daha da azaltabilirsiniz.

Bunu başarmanın iki yolu:

- 'sanal makineler' veya 'konteynerlar' kullanarak - içeride çalışan, ancak BT sisteminin geri kalanından izole edilmiş bir bilgisayar sisteminin emülasyonları. Bunlar, özellikle ML görevleri için önceden yapılandırılabilir. İşe alım örneğimizde, makine öğrenimi mühendisi bir sanal makine kullanmış olsaydı, güvenlik açığı kontrol altına alınabilirdi.
- Birçok ML sistemi, Python gibi bilimsel ve makine öğrenimi kullanımları için iyi geliştirilmiş, ancak en güvenli olmayan programlama dilleri kullanılarak geliştirilir. Ancak, bir ML modelini bir programlama dili (örn. Python) kullanarak eğitmek mümkündür, ancak daha sonra dağıtımdan önce modeli başka bir dile (örn. Java) dönüştürerek güvensiz kodlamayı daha az olası hale getirir. İşe alma örneğimize geri dönersek, makine öğrenimi mühendisinin CV filtreleme modeline yönelik kötü niyetli saldırı riskini azaltabileceği başka bir yol, dağıtımdan önce modeli farklı bir programlama diline dönüştürmek olurdu.

İlave okuma – ICO kılavuzu

Daha fazla bilgi için Çevrimiçi hizmetlerde kişisel verilerin korunması: başkalarının hatalarından öğrenme (PDF) hakkındaki raporumuzu okuyun. 2014 yılında yazılmış olmasına rağmen, raporun bu alandaki içeriği yine de size yardımcı olabilir.

ICO, veri koruma perspektifinden farklı olarak korunan kaynak kodun gözetimi ve gözden geçirilmesi için ek önerilerin yanı sıra tasarım gereği güvenlik ve veri korumasına yönelik etkilerini içerecek olan daha fazla güvenlik kılavuzu geliştiriyor.

Diğer kaynaklar

Ulusal Siber Güvenlik Merkezi'nden (NCSC- National Cyber Security Centre) kod depolarının bakımına ilişkin rehberlik de size yardımcı olabilir.

Yapay zeka modelleri için ne tür gizlilik saldırıları uygulanır?

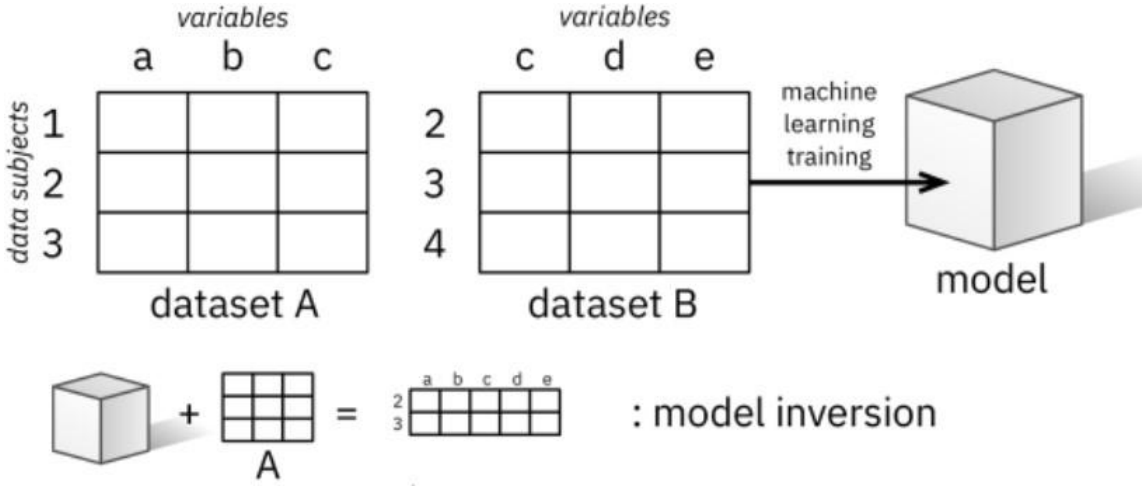
AI sisteminin eğitim aldığı kişilere ait kişisel veriler, sistemin kendisinin çıktıları tarafından istemeden ortaya çıkmış olabilir.

Normalde, verileri bir AI sistemini eğitmek için kullanılan kişilere ait kişisel verilerinin, sistemin, yeni girdilere yanıt olarak döndürdüğü tahminleri basitçe gözlemleyerek çıkarılamayacağı varsayılır. Ancak, makine öğrenimi modellerine yönelik yeni tür gizlilik saldırıları, bunun bazen mümkün olabileceğini gösteriyor.

Bu güncellemede, bu gizlilik saldırılarının iki türüne odaklanacağız – "model ters çevirme-model inversion" ve "üyelik çıkarımı-membership inference".

Model inversiyon-ters çevirme saldırıları (inversion attack) nedir?

Bir model ters çevirme saldırısında, saldırganlar eğitim verilerinde yer alan belirli kişilere ait bazı kişisel verilere zaten erişime sahiplerse, ML modelinin girdi ve çıktılarını gözlemleyerek aynı kişiler hakkında daha fazla kişisel bilgi elde edebilirler. Saldırganların öğrenebileceği bilgiler, benzer özelliklere sahip kişiler hakkında genel çıkarımların ötesine geçer.



Şekil 1. Model ters çevirme ve üyelik çıkarım saldırılarının gösterimi, Veale et al. 'Hatırlayan algoritmalar: model inversiyon saldırıları ve veri koruma kanunu

Birinci örnek – model ters çevirme saldırısı

Bu tür bir saldırının erken bir gösterimi, genetik biyobelirteçler dahil hasta verilerini kullanarak bir antikoagülanın doğru dozunu tahmin etmek için tasarlanmış bir tıbbi modelle ilgiliydi. Eğitim verilerine dahil edilen kişiler hakkında bazı demografik bilgilere erişimi olan bir saldırganın, temel eğitim verilerine erişimi olmamasına rağmen, modelden genetik biyobelirteçlerini çıkarabileceğini kanıtladı.

İkinci örnek – model ters çevirme saldırısı

Yakın tarihli bir başka örnek, saldırganların bir Yüz Tanıma Teknolojisi (FRT – Face Recognition Technology) sisteminin tanımak üzere eğitildiği yüzlerin görüntülerini yeniden oluşturabileceğini gösteriyor. FRT sistemleri genellikle üçüncü tarafların modeli sorgulamasına izin verecek şekilde tasarlanmıştır. Modele, yüzünü tanıdığı bir kişinin görüntüsü verildiğinde, model, kişinin adına ve ilişkili güven oranına ilişkin en iyi tahminini verir

Saldırganlar, rastgele oluşturulmuş birçok farklı yüz görüntüsü göndererek modeli araştırabilir. Model tarafından döndürülen adları ve güven puanlarını gözlemleyerek, eğitim verilerine dahil edilen kişilerle ilişkili yüz görüntülerini yeniden oluşturabildiler. Yeniden yapılandırılmış yüz görüntüleri kusurlu olsa da, araştırmacılar eğitim verilerindeki kişilerle (insan inceleyici tarafından) %95 doğrulukla eşleştirilebileceğini buldular (bkz. Şekil 2)

Şekil 2. Fredriksen ve diğerleri, 'Güven bilgisini istismar Eden Model Ters Çevirme Saldırılar' model ters çevirme saldırısı (solda) ve karşılık gelen eğitim seti görüntüsü (sağda) kullanılarak kurtarılan bir yüz görüntüsü.

Diğer Kaynaklar

'Hatırlayan algoritmalar: model inversiyon (ters çevirme) saldırıları ve veri koruma yasası'

Basit demografi genellikle insanları benzersiz bir şekilde tanımlar

'Güven bilgisinden ve temel karşı önlemlerden yararlanan model tersine çevirme saldırıları'

Üyelik çıkarım saldırıları (membership inference attacks) nelerdir?

Üyelik çıkarım saldırıları, kötü niyetli saldırganların, bir ML modelinin eğitim verilerinde belirli bir kişinin bulunup bulunmadığını belirlemesine olanak tanır. Ancak, model inversiyonundan farklı olarak, kişi hakkında herhangi bir ek kişisel veri öğrenmeleri gerekmez.

Örneğin, bir hastanın ne zaman taburcu olacağını tahmin eden bir modeli eğitmek için hastane kayıtları kullanılıyorsa, saldırganlar, eğitim verilerinin bir parçası olup olmadıklarını anlamak için bu modeli belirli bir kişiyle ilgili (zaten sahip oldukları) diğer verilerle birlikte kullanabilir. Bu, eğitim veri setinin kendisinden herhangi bir kişinin verilerini ortaya çıkarmaz, ancak pratikte, verilerin toplandığı dönemde eğitim verilerini oluşturan hastanelerden birini ziyaret ettiklerini ortaya çıkarır.

Daha önceki FRT örneğine benzer şekilde, üyelik çıkarım saldırıları, bir modelin tahminiyle birlikte sağlanan güven puanlarından yararlanabilir. Eğitim verisinde bir kişi varsa, model o kişiyle ilgili bir tahminde orantısız bir şekilde kendinden emin olacaktır çünkü onları daha önce görmüştür. Bu, saldırganın kişinin eğitim verilerinde olduğunu çıkarmasına olanak tanır.

Modellerin üyelik çıkarımına karşı savunmasızlığının sonuçlarının ciddiyeti, üyeliğin ne kadar hassas veya açıklayıcı olabileceğine bağlı olacaktır. Bir model, genel popülasyondan alınan çok sayıda insan üzerinde eğitilirse, üyelik çıkarım saldırıları daha az risk oluşturur. Ancak model, savunmasız veya hassas bir popülasyon (örneğin akıl hastalığı veya HIV hastaları) üzerinde eğitilmişse, o zaman yalnızca birinin bu nüfusun bir parçası olduğunu ifşa etmek ciddi bir mahremiyet riski olabilir.

Kara kutu ve beyaz kutu saldırıları nedir?

Modellere yönelik 'kara kutu' ve 'beyaz kutu' saldırıları arasında önemli bir ayrım vardır. Bu iki yaklaşım, farklı operasyonel modellere karşılık gelir.

Beyaz kutu saldırılarında, saldırgan modelin kendisine tam erişime sahiptir ve temel alınan kod ve özellikleri inceleyebilir (eğitim verileri olmasa da). Örneğin, bazı AI sağlayıcıları, üçüncü taraflara önceden eğitilmiş bir modelin tamamını verir ve lokal olarak çalıştırmalarına izin verir. Beyaz kutu saldırıları, bir

saldırganın modelden kişisel veriler çıkarmasına yardımcı olabilecek model türü ve kullanılan parametreler gibi ek bilgilerin toplanmasını sağlar.

Kara kutu saldırılarında saldırgan sadece modeli sorgulama ve girdiler ile çıktılar arasındaki ilişkileri gözlemleme yeteneğine sahiptir. Örneğin, birçok AI sağlayıcısı, girdi verilerini içeren sorgular göndermek ve modelin yanıtını almak için üçüncü tarafların çevrimiçi olarak bir ML modelinin işlevselliğine erişmesine olanak tanır. Bu vurguladığımız örneklerin ikisi de kara kutu saldırılarıdır.

Beyaz ve kara kutu saldırıları, sağlayıcıların müşterileri veya modelin kendisine ya da sorgu veya yanıt işlevselliğine yetki verilmiş ya da yetkisiz erişimi olan herkes tarafından gerçekleştirilebilir.

Tasarım gereği eğitim verilerini içeren modeller ne olacak?

Model inversiyonu ve üyelik çıkarımları, AI modellerinin istemeden kişisel veriler içerebileceğini göstermektedir. Ayrıca, tasarım gereği eğitim verilerinin parçalarını ham biçiminde içeren belirli türde ML modelleri olduğunu da unutmamalısınız. Örneğin, "destek vektör makineleri - support vector machines" (SVM'ler) ve "k-en yakın komşular- k-nearest neighbours" (KNN) modelleri, modelin kendisinde eğitim verilerinin bir kısmını içerir.

Bu gibi durumlarda, eğitim verilerinin kişisel veri olması durumunda, modele erişimin kendisi, modeli satın alan organizasyonun, daha fazla çaba sarf etmesine gerek kalmadan eğitim verilerinde yer alan kişisel verilerin bir alt setine zaten erişebileceği anlamına gelir. Bu tür ML modellerinin sağlayıcıları ve bunları tedarik eden üçüncü taraflar, bu şekilde kişisel veriler içerebileceklerinin farkında olmalıdır.

Model inversiyonu ve üyelik çıkarımının aksine, bunun gibi modellerde bulunan kişisel veriler bir saldırı vektörü değildir; bu tür modellerde yer alan herhangi bir kişisel veri, tasarım gereği orada olacaktır ve üçüncü şahıs tarafından kolayca alınabilir. Bu tür modellerin saklanması ve kullanılması bu nedenle kişisel verilerin işlenmesini oluşturur ve bu nedenle standart veri koruma hükümleri uygulanır.

AI modellerinde gizlilik saldırılarının risklerini yönetmek için hangi adımları atmalıyız?

Modelleri eğitir ve başkalarına sağlarsanız, bu modellerin kişisel veri içerip içermediğini veya saldırıya uğrarsa bunları ifşa etme riski altında olup olmadığını değerlendirmeli ve bu riskleri azaltmak için uygun adımları atmalısınız.

Eğitim verilerinin, doğrudan veya modele erişimi olan kişiler tarafından, kişilere ait tanımlanmış veya tanımlanabilir kişisel verileri içerip içermediğini değerlendirmelisiniz. Yukarıda açıklanan güvenlik açıkları ışığında makul olarak kullanılması muhtemel olan araçları değerlendirmelisiniz. Bu hızla gelişen bir alan olduğundan, hem saldırı hem de azaltma yöntemlerinde en son gelişmelerden haberdar olmalısınız.

Güvenlik ve makine öğrenimi araştırmacıları, hangi faktörlerin makine öğrenimi modellerini bu tür saldırılara karşı daha fazla veya daha az savunmasız hale getirdiğini ve etkili koruma ve azaltma stratejilerinin nasıl tasarlanacağını anlamaya çalışmaktadır.

ML modellerinin gizlilik saldırılarına karşı savunmasız olmasının olası bir nedeni "aşırı uyum- overfitting" olarak bilinir. Bu, modelin eğitim verilerinin ayrıntılarına çok fazla dikkat ettiği ve yalnızca genel kalıplardan ziyade eğitim verilerinden belirli örnekleri neredeyse etkili bir şekilde hatırladığı yerdir. Model inversiyonu ve üyelik çıkarımı saldırıları bundan yararlanabilir.

Aşırı uyumdan kaçınmak, hem mahremiyet saldırıları riskini azaltmada hem de modelin daha önce görmediği yeni örnekler üzerinde iyi çıkarımlar yapabilmesini sağlamada yardımcı olacaktır. Ancak fazla uyumdan kaçınmak riskleri tamamen ortadan kaldırmaz. Eğitim verilerine gereğinden fazla uymayan modeller bile gizlilik saldırılarına karşı savunmasız olabilir.

Yukarıdaki FRT örneğinde olduğu gibi bir ML sistemi tarafından sağlanan güven bilgilerin kullanılabileceği durumlarda, son kullanıcıya verilmeyerek risk azaltılabilir. Bunun, gerçek son kullanıcıların çıktısına güvenip güvenmeyeceğini bilme ihtiyacına karşı dengelenmesi gerekecek ve belirli kullanım durumu ve şartlarına bağlı olacaktır.

Bir Uygulama Programlama Arayüzü (API) aracılığıyla başkalarına bütün bir model sağlayacaksanız, bu şekilde beyaz kutu saldırılarına maruz kalmazsınız, çünkü API kullanıcıları modelin kendisine doğrudan erişemez. Ancak yine de kara kutu saldırılarına maruz kalabilirsiniz.

Bu riski azaltmak; API'nin şüpheli bir şekilde kullanılıp kullanılmadığını tespit etmek için API kullanıcılarından gelen sorguları izleyebilirsiniz. Bu, bir gizlilik saldırısını gösterebilir ve hızlı bir araştırma yapılmasını ve belirli bir kullanıcı hesabının potansiyel olarak askıya alınmasını veya engellenmesini gerektirebilir. Bu tür önlemler, "hız sınırlama" (belirli bir zaman sınırı içinde belirli bir kullanıcı tarafından gerçekleştirilebilecek sorgu sayısını azaltma) gibi diğer güvenlik tehditlerine karşı koruma sağlamak için kullanılan yaygın gerçek zamanlı izleme tekniklerinin bir parçası olabilir.

Modeliniz bir API aracılığıyla yalnızca erişilebilir olmak yerine üçüncü bir tarafa tamamen sağlanacaksa, "beyaz kutu" saldırıları riskini göz önünde bulundurmanız gerekir. Model sağlayıcı olarak, dağıtım sırasında modeli daha kolay izleyemeyecek ve böylece modele yönelik gizlilik saldırıları riskini değerlendirip azaltabileceksiniz.

Ancak, müşterilerinizin modeli uygulamaya koyma şeklinden dolayı modellerinizi eğitmek için kullanılan kişisel verilerin açığa çıkmamasını sağlamaktan siz sorumlu olursunuz. Belirli uygulamaya alma durumlarını ve ilişkili tehdit modellerini anlamak için müşterilerinizle işbirliği yapmadan bu riski tam olarak değerlendiremeyebilirsiniz.

Satın alma politikanızın bir parçası olarak, gerektiğinde ilgili değerlendirmelerinizi gerçekleştirmek için her taraf arasında yeterli bilgi paylaşımı olmalıdır. Bazı durumlarda, ML model sağlayıcıları ve müşterilerin ortak kontrol birimleri olur ve bu nedenle ortak bir risk değerlendirmesi yapması gerekir.

Modelin default olarak eğitim verilerinden örnekler içerdiği durumlarda (yukarıda bahsedilen SVM'lerde ve KNN'lerde olduğu gibi), bu bir kişisel veri aktarımıdır ve buna böyle muamele etmelisiniz.

Peki ya çelişkili örnekler?

AI ile ilgili temel veri koruma endişeleri, kişisel verilerin yanlışlıkla ifşa edilmesini içerirken, "çelişkili örnekler" gibi başka potansiyel yeni AI güvenlik riskleri de vardır.

Bunlar, güvenilir bir şekilde yanlış sınıflandırılmaları için kasıtlı olarak değiştirilmiş bir ML modeline beslenen örneklerdir. Bunlar, manipüle edilmiş görüntüler veya hatta öğenin yüzeyine yerleştirilmiş çıkartmalar gibi gerçek dünyadaki değişiklikler olabilir. Örnekler arasında, silah olarak sınıflandırılan kaplumbağa resimleri veya bir insanın anında 'DUR' olarak tanıyacağı, ancak bir görüntü tanıma modelinin tanımadığı, üzerlerinde çıkartma bulunan yol işaretleri sayılabilir.

Bu tür muhalif örnekler güvenlik açısından ilgili olsa da, kişisel verileri içermiyorlarsa, kendi başlarına veri koruma endişelerini dile getirmeyebilirler. Güvenlik ilkesi, kişisel verilerin güvenliğini ifade eder - yetkisiz işlemlere karşı korur. Bununla birlikte, çelişkili saldırılar, kişisel verilerin yetkisiz olarak işlenmesini gerektirmez, yalnızca sisteme bir uzlaşma sağlar.

Ancak, karşıt örneklerin kişilerin hak ve özgürlüklerine yönelik risk oluşturabileceği durumlar olabilir. Örneğin, yüz tanıma sistemlerine bazı saldırılar yapıldı. Bir kişinin yüz görüntüsü hafifçe bozularak, yüz tanıma sistemi kandırılarak, onları başka biri olarak yanlış sınıflandırması için kandırabilir (bir insan yine de çarpık görüntüyü doğru kişi olarak tanıyacak olsa bile). Bu, özellikle sistem kişiler hakkında yasal veya benzer şekilde önemli kararlar veriyorsa, sistemin istatistiksel doğruluğu hakkında endişelere yol açacaktır.

Ayrıca, 2018 NIS yönergesi kapsamındaki yükümlülüklerinizin bir parçası olarak çekişmeli örnekler riskini de göz önünde bulundurmanız gerekebilir. ICO, NIS kapsamında 'ilgili dijital hizmet sağlayıcılar' için yetkili makamdır. Bunlara çevrimiçi arama motorları, çevrimiçi pazar yerleri ve bulut bilişim hizmetleri dahildir. Bir "NIS olayı", ağ ve bilgi sistemleri tarafından depolanan verileri ve sağladıkları ilgili hizmetleri tehlikeye atan olayları içerir. Bu muhtemelen AI bulut bilişim hizmetlerini içerecektir. Dolayısıyla, bir düşman saldırısı kişisel verileri içermese bile, yine de bir NIS olayı olabilir ve bu nedenle ICO'nun görev alanı içinde olabilir.

İlave okuma – ICO kılavuzu

İlgili bir dijital hizmet sağlayıcı olarak nitelikli olup olmadığınız da dahil olmak üzere NIS Düzenlemeleri hakkında daha fazla bilgi için NIS Kılavuzumuzu okuyun.

Kontrol örnekleri

Risk Bildirimi

AI sistemlerinin altyapısı ve mimarisi; kişisel verilere yetkisiz erişim, değişiklik veya imha olasılığını artırır.

Önleyici

- Güvenlik açıkları ile ilgili uyarılar almak için güvenlik önerilerine abone olun.
- Harici güvenlik sertifikalarına veya planlarına karşı bir AI sistemine uyum sağlayın veya bu sistemi değerlendirin.
- Yazılımı, bir veya daha fazla kişinin kaynak kodunun bölümlerini görüntüleyip okuduğu bir kalite incelemesine tabi tutun. İnceleyenlerden en az biri kodun yazarı olmamalıdır.
- AI geliştirme ortamının diğer BT ağı/altyapısından ayrılması için politika/süreç belgeleyin. Ayrılığa uyulduğuna / gerçekleştiğine dair kanıt bulundurun.
- Maliyetlerin, risklerin, hizmet/performans ve sürdürülebilirliğin optimizasyonuna yönelik koordineli bir yaklaşım sağlamak için varlık yönetimi yaklaşımına sahip olun.
- Üçüncü taraflarla yapılan sözleşmeleri, üçüncü tarafların rol ve sorumlulukları hakkında net olduğunu belgeleyin.
- Üçüncü taraflarla ilişkilere ilişkin politikaları/süreçleri ve bilgi güvenliğinin durum tespitinin tamamlandığına dair kanıtı belgeleyin.
- İhlal raporlaması ve eskalasyon için politika/süreçleri belgeleyin.
- Politikaya / sürece bağlı kalın.
- Model bir yönetişim politikasına sahip olun.
- Eğitilen modelin daha güvenli uygulamalarını değerlendirin ve bunları geliştirme sonrasında, ancak dağıtım öncesinde uygun şekilde uygulayın.
- En son gizlilik artırıcı teknikleri gözden geçirmek, tekniğin kendi durumlarına uygulanabilirliğini değerlendirmek ve uygun olduğunda uygulamak için süreçlere sahip olun.
- Bir saldırı olasılığını ve etkisini azaltmak için güvenlik risklerinin kapsamlı bir değerlendirmesini ve azaltılması/kontrollerini içeren bir DPIA'yı belgeleyin.
- Şüpheli etkinliği belirlemek ve bildirmek için isteklerin hacmini ve modellerini izleyen bir API erişim politikasına sahip olun.
- Personelin, ihlal raporlama politikasını ve hangi prosedürlerin izlenmesi gerektiğini anlama konusunda eğitilmesini sağlayın.

Tespit edici

- Şüpheli istekleri tespit etmek ve sonuç olarak harekete geçmek için API isteklerini izleyin.
- Yerleştirilen güvenlik önlemlerinin etkinliğini düzenli olarak test edin, değerlendirin ve ölçün (örn. sızma testi gibi tekniklerle).
- Şikayetleri izleyin ve etkilenebilecek diğer kişileri belirlemek için daha geniş analizler de dahil olmak üzere sonuç olarak harekete geçin.

Düzeltici

- Gelecekteki saldırı riskini azaltmak için analiz / gerekçelendirme dahil olmak üzere AI sistem tasarımında yapılan değişiklikleri katıtlayın.

AI sistemleri için hangi veri minimizasyonu ve gizlilik koruma teknikleri uygundur?

Veri minimizasyon ilkesine ilişkin hangi hususları dikkate almamız gerekiyor?

Veri minimizasyonu ilkesi, amacınızı gerçekleştirmek için ihtiyaç duyduğunuz minimum kişisel veri miktarını belirlemenizi ve yalnızca bu bilgileri işlemenizi gerektirir, daha fazlasını değil. Örneğin, GDPR Madde 5(1)(c) şöyle der:

Alıntı

'1. Kişisel veriler; işlendikleri amaçlarla ilgili olarak yeterli, alakalı ve gerekli olanlarla sınırlı olacaktır (veri minimizasyonu)'

Ancak, yapay zeka sistemleri genellikle büyük miktarda veri gerektirir. Bu nedenle, ilk bakışta, AI sistemlerinin veri minimizasyon ilkesine nasıl uyum sağlayabileceğini görmek zor olabilir - ancak işlemenizin bir parçası olarak AI kullanıyorsanız, yine de bunu yapmanız gerekir.

Zor gibi görünse de pratikte durum böyle olmayabilir. Veri minimizasyonu ilkesi, "hiçbir kişisel veri işlemeyiz" veya "daha fazla iş yaparsak kanunu çiğneriz" anlamına gelmez. Önemli olan, yalnızca ihtiyacınız olan kişisel verileri amacınız için işlemenizdir.

Neyin 'yeterli, alakalı ve sınırlı' olduğunu belirleme konusunda nasıl bir yol izleyeceğiniz, bu nedenle koşullarınıza özel olacaktır ve veri minimizasyonuna ilişkin mevcut kılavuzumuz, atmanız gereken adımları ayrıntılı olarak açıklamaktadır.

Yapay zeka sistemleri bağlamında, "yeterli, alakalı ve sınırlı" olan şey bu nedenle duruma özeldir. Ancak, işlevsel kalırken yalnızca ihtiyacınız olan verileri işleyen yapay zeka sistemleri geliştirmek için uygulayabileceğiniz birkaç teknik vardır.

Bu bölümde, şu anda kullanımda olan en yaygın yapay zeka türü olan denetimli Makine Öğrenimi (ML) sistemleri için en alakalı tekniklerden bazılarını inceliyoruz.

Organizasyonunuz içinde, risk yönetiminden ve AI sistemlerinin uyumluluğundan sorumlu kişilerin, bu tür tekniklerin var olduğunun farkında olması ve teknik personeli ile farklı yaklaşımları tartışabilmesi ve değerlendirebilmesi gerekir. Örneğin, veri bilimcilerinin yapay zeka sistemleri tasarlama ve inşa etme konusundaki default yaklaşımı, aynı amaçlara daha az veriyle nasıl ulaşabileceklerini düşünmeden mümkün olduğunca çok veri toplamayı ve kullanmayı içerebilir.

Bu nedenle, veri minimizasyonunun ve ilgili tüm minimizasyon tekniklerinin tasarım aşamasından itibaren tamamen dikkate alınmasını sağlamak için tasarlanmış risk yönetimi uygulamalarını uygulamalısınız. Benzer şekilde, yapay zeka sistemleri satın alırsanız ve/veya üçüncü taraflarca işletilen sistemler uygularsanız, bu hususlar tedarik süreci durum değerlendirmesinin bir parçasını oluşturmalıdır.

Ayrıca, veri minimizasyonu ilkesine uymanıza yardımcı olsalar da burada açıklanan tekniklerin diğer risk türlerini ortadan kaldırmadığını da bilmelisiniz.

Ayrıca, bazı teknikler veri minimizasyonu gereksinimlerine uymak için herhangi bir uzlaşma gerektirmezken, diğerleri, örneğin istatistiksel olarak daha doğru ve ayrımcı olmayan ML modelleri yapmak gibi, veri minimizasyonunu diğer uyumluluk veya fayda hedefleriyle dengelemenize ihtiyaç duyabilir. İlave ayrıntı için ödünleşmeler bölümüne bakın.

Veri minimizasyonuna uyum için atmanız gereken ilk adım, kişisel verilerin kullanılabileceği tüm ML süreçlerini anlamak ve haritasını çıkarmaktır.

Mevzuattaki ilgili hükümler

Bknz. GDPR 5(1)(c) madde ve 39 gerekçe, ile 16 madde (düzeltme hakkı) ve 17 madde (silme hakkı) (dış bağlantı)

İlave okuma – ICO kılavuzu

GDPR Kılavuzu'ndaki veri minimizasyonu ilkesine ilişkin kılavuzumuzu okuyun.

Denetimli ML modellerinde kişisel verileri nasıl işlemeliyiz?

Denetimli ML algoritmaları, modelleri tanımlamak ve modelden sınıflandırması veya tahmin etmesi istenecek örnek türlerinin geçmiş örneklerini içeren veri setlerinden ("eğitim verileri") modeller oluşturmak için eğitilebilir. Spesifik

olarak, eğitim verileri hem "hedef" değişkeni, yani modelin tahmin etmeyi veya sınıflandırmayı amaçladığı şeyi hem de birkaç "tahmin edici" değişkeni, yani tahmini yapmak için kullanılan girdiyi içerir.

Örneğin, bir bankanın kredi riski ML modeli için eğitim verilerinde, tahmin değişkenleri yaş, gelir, meslek ve önceki müşterilerin bulunduğu yeri içerebilirken, hedef değişken müşterilerin kredilerini geri ödeyip ödemediği olacaktır.

ML sistemleri bir kez eğitildikten sonra, sistemin daha önce hiç görmediği örnekleri içeren yeni verilere dayalı olarak sınıflandırabilir ve tahminlerde bulunabilir. ML modeline, yeni bir örnek için tahmin değişkenlerini içeren bir sorgu gönderilir (örneğin, yeni bir müşterinin yaşı, geliri, mesleği, vb.). Model, bu yeni durum için hedef değişkene ilişkin en iyi tahminiyle yanıt verir (örneğin, yeni müşterinin bir krediyi temerrüde düşüp düşmeyeceği).

Denetimli ML yaklaşımları bu nedenle verileri iki ana aşamada kullanır:

1. Eğitim verilerinin geçmiş örneklere dayalı modeller geliştirmek için kullanıldığı **eğitim aşaması.**
2. Modelin yeni örnekler hakkında bir tahmin veya sınıflandırma yapmak için kullanıldığı **çıkarım aşaması**.

Model, bireysel kişiler hakkında tahminler veya sınıflandırmalar yapmak için kullanılıyorsa, kişisel verilerin hem eğitim hem de çıkarım aşamalarında kullanılması çok muhtemeldir.

ML uygulamaları tasarlarken kişisel verileri en aza indirmek için hangi teknikleri kullanmalıyız?

ML uygulamaları tasarlarken ve oluştururken, veri bilimcileri genellikle sistemin eğitiminde, test edilmesinde ve çalıştırılmasında kullanılan tüm verilerin merkezi bir şekilde toplanacağını ve AI sistem yaşam döngüsü boyunca birden çok yerde tek bir varlık tarafından tam ve orijinal biçiminde tutulacağını varsayar.

Ancak bunun kişisel veri olduğu durumlarda, amacınız/amaçlarınız için işlemenin gerekli olup olmadığını değerlendirmeniz gerekir. Aynı sonucu daha az kişisel veri işleyerek elde edebiliyorsanız, tanım gereği veri minimizasyonu ilkesi bunu yapmanızı gerektirir.

İşlemeniz gereken kişisel veri miktarını en aza indirmenize yardımcı olabilecek bir dizi teknik mevcuttur.

Eğitim aşamasında kişisel verileri nasıl minimize etmeliyiz?

Açıkladığımız gibi, eğitim aşaması, tahmin veya sınıflandırma oluşturmak için kullanılan her bir kişi için bir dizi özellik içeren bir veri setine bir öğrenme algoritması uygulamayı içerir.

Ancak, bir veri setinde yer alan tüm özellikler mutlaka amacınızla ilgili olmayacaktır. Örneğin, tüm finansal ve demografik özellikler kredi riskini tahmin etmek için faydalı olmayacaktır. Bu nedenle, hangi özelliklerin - ve dolayısıyla hangi verilerin - amacınıza uygun olduğunu değerlendirmeniz ve yalnızca bu verileri işlemeniz gerekir.

Bir modele dahil edilmek için faydalı olacak özellikleri seçmek için veri bilimcileri tarafından kullanılan çeşitli standart özellik seçim yöntemleri vardır. Bu yöntemler veri biliminde iyi uygulamalardır, ayrıca veri minimizasyon ilkesini karşılama yolunda da bir yol kat ederler.

Ayrıca, ICO'nun AI ve Büyük Veri ile ilgili önceki raporunda tartışıldığı gibi, süreçte bazı verilerin daha sonra tahmin yapmak için yararlı olabileceği gerçeği, bu amaçla neden saklamanız gerektiğini belirlemek için yeterli değildir, verilerin toplanmasını, kullanılmasını veya saklanmasını geriye dönük olarak haklı çıkarmaz. Gerçekleşmeyecek öngörülebilir bir olay için bilgi tutabilmeniz mümkün olsa da kişisel verileri gelecekte yararlı olma ihtimaline karşı toplamamalısınız, sadece bunu gerekçelendirebiliyorsanız tutabilirsiniz.

İlave okuma – ICO kılavuzu

Büyük Veri, yapay zeka, makine öğrenimi ve veri koruması hakkındaki raporumuz okuyun.

Hangi gizliliği artıran yöntemleri dikkate almalıyız?

Ayrıca, eğitim aşamasında işlenmekte olan kişisel verileri en aza indirmek için kullanabileceğiniz, gizliliği artırmaya yönelik çeşitli teknikler de vardır:

- Bozuculuk (perturbation) veya 'gürültü' eklenmesi.
- Birleşik öğrenme (federated learning).

Bu tekniklerin bazıları, iyi performans gösteren modelleri eğitmek amacıyla kullanımını korurken, belirli kişilere kadar izlenebilme derecesini azaltmak için eğitim verilerinin değiştirilmesini içerir.

Bu tür gizliliği artıran teknikleri, zaten topladıktan sonra eğitim verilerine uygulayabilirsiniz. Ancak, mümkün olduğunda, büyük veri setlerinin kişilere yönelik risklerini azaltmanın bir parçası olarak, herhangi bir kişisel veri toplamadan önce bunları uygulamalısınız.

Bu mahremiyeti artıran tekniklerin, kişilerin mahremiyeti ile bir ML sisteminin faydasını dengelemedeki etkinliğini, differansiyel gizlilik gibi yöntemleri kullanarak matematiksel olarak ölçebilirsiniz.

Diferansiyel gizlilik, bir ML algoritması tarafından oluşturulan bir modelin, onu eğitmek için kullanılan herhangi bir kişinin verilerine önemli ölçüde bağlı olup

olmadığını ölçmenin bir yoludur. Teoride matematiksel olarak titiz olsa da, pratikte anlamlı bir şekilde diferansiyel gizliliği uygulamak hala zordur.

Bu yöntemlerdeki gelişmeleri izlemeli ve uygulamaya başlamadan önce, kendi durumunuzda anlamlı veri minimizasyonu sağlayıp sağlayamayacaklarını değerlendirmelisiniz.

- **Bozuculuk (Perturbation)**

Değişiklik, kişilere ait veri noktalarının değerlerinin, bu özelliklerin bazı istatistiksel özelliklerini koruyacak şekilde rastgele değiştirilmesini içerebilir - "bozucu" veya verilere "gürültü" eklenmesi olarak bilinir.

Genel olarak konuşursak, "gürültülü verilerden" ne kadar öğrenebileceğinizin açık sonuçlarıyla ne kadar gürültü enjekte edeceğinizi seçebilirsiniz.

Örneğin, akıllı telefon tahmini metin sistemleri, kullanıcıların daha önce yazdıkları sözcükleri temel alır. Her zaman bir kullanıcının gerçek tuş vuruşlarını toplamak yerine, sistem rastgele "gürültülü" (yani yanlış) kelimeler oluşturacak şekilde tasarlanabilir. Bu, hangi kelimelerin "gürültü" olduğunu ve hangi kelimelerin belirli bir kullanıcı tarafından gerçekten yazıldığını büyük ölçüde daha az kesinleştirdiği anlamına gelir.

Sistemin yeterli kullanıcısı olması koşuluyla veriler kişisel düzeyde daha az doğru olsa da yine de kalıpları gözlemleyebilir ve bunları ML modelinizi toplu düzeyde eğitmek için kullanabilirsiniz. Ne kadar fazla gürültü enjekte ederseniz, verilerden o kadar az şey öğrenebilirsiniz, ancak bazı durumlarda, verileri anlamlı bir koruma düzeyi sağlayacak şekilde takma ad (pseudonymous) haline getirmek için yeterli gürültü enjekte edebilirsiniz.

- **Birleşik öğrenme**

İlgili bir gizliliği koruma tekniği, birleşik öğrenmedir. Bu, birden fazla farklı tarafın modelleri kendi verileri ('lokal' modeller) üzerinde eğitmesine olanak tanır. Daha sonra, bu modellerin tanımladığı ("gradyanlar" olarak bilinir) bazı kalıpları, birbirleriyle herhangi bir eğitim verisi paylaşmak zorunda kalmadan tek, daha doğru bir "küresel" modelde birleştirirler.

Birleşik öğrenme nispeten yenidir ancak birkaç büyük ölçekli uygulaması vardır. Bunlar, akıllı telefonlarda otomatik düzeltme ve tahmine dayalı metin modellerinin yanı sıra birden fazla hasta veri tabanında analiz içeren tıbbi araştırmaları içerir.

Lokal olarak eğitilmiş bir modelden elde edilen gradyanı paylaşmak, eğitim verilerinin kendisini paylaşmaktan daha düşük bir gizlilik riski teşkil ederken, bir gradyan, özellikle modelin çok sayıda ince taneli (fine-grained) değişkenle karmaşık olması durumunda, türetildiği kişilerle ilgili bazı kişisel bilgileri açıklayabilir. Bu nedenle yeniden tanımlama riskini yine de değerlendirmeniz gerekir. Birleştirilmiş öğrenme durumunda, katılımcı organizasyonlar,

birbirlerinin verilerine erişememelerine rağmen ortak denetleyiciler olarak kabul edilebilirler.

İlave okuma

Yapay zekada denetleyicilik hakkında daha fazla bilgi için denetleyici/işlemci ilişkileri bölümünü okuyun.

Bir bozuculuk (perturbation) örneği için 'Rappor (randomised aggregatable privacy preserving ordinal responses)' konusuna bakın

Çıkarım (inference) aşamasında kişisel verileri nasıl en aza indirmeliyiz?

Bir kişi hakkında bir tahmin veya sınıflandırma yapmak için, ML modelleri genellikle o kişinin sorguya dahil edilmesi için tam bir tahmin değişkenleri setini gerektirir. Eğitim aşamasında olduğu gibi, çıkarım aşamasında da kişisel verileri en aza indirmek ve/veya bu verilere yönelik riskleri azaltmak için kullanabileceğiniz çeşitli teknikler bulunmaktadır:

- Kişisel verileri daha az "insan tarafından okunabilir" biçimlere dönüştürmek.
- Lokal olarak çıkarımlar yapmak.
- Gizliliği koruyan sorgu yaklaşımları.

Bu yaklaşımları aşağıda ele alıyoruz.

- **Kişisel verileri daha az "insan tarafından okunabilir" biçimlere dönüştürmek**

Çoğu durumda, verileri bir model tarafından sınıflandırılmasına izin veren bir formata dönüştürme süreci, onu minimize etmeye yönelik bir yolda ilerler. Ham kişisel verilerin genellikle tahmin amacıyla daha soyut bir formata dönüştürülmesi gerekecektir. Örneğin, insan tarafından okunabilen kelimeler normalde bir dizi sayıya çevrilir ("özellik vektörü" olarak adlandırılır).

Bu, eğer bir AI modeli devreye alırsanız; sorguda yer alan kişisel verilerin, insan tarafından yorumlanabilir sürümünü işlemeniz gerekmeyebileceği anlamına gelir; örneğin, dönüşüm kullanıcının cihazında gerçekleşirse.

Ancak, kolayca insan tarafından yorumlanamaz olması, dönüştürülen verilerin artık kişisel olmadığı anlamına gelmez. Örneğin, Yüz Tanıma Teknolojisini (FRT) düşünün. Bir yüz tanıma modelinin çalışması için, sınıflandırılan yüzlerin dijital görüntülerinin 'yüz izlerine' dönüştürülmesi gerekir. Bunlar, alttaki yüzlerin geometrik özelliklerinin matematiksel temsilleridir – örneğin bir kişinin burnu ile üst dudağı arasındaki mesafe.

Yüz görüntülerini sunucularınıza göndermek yerine, fotoğraflar doğrudan kişilerin cihazında yüz izlerine dönüştürülebilir ve bu da onları sorgulama için

modele göndermeden önce yakalar. Bu yüz izleri, herhangi bir insan tarafından yüz fotoğraflarından daha kolay teşhis edilebilir olacaktır.

Bununla birlikte, yüz izleri hala kişisel (aslında biyometrik) verilerdir ve bu nedenle onları kullanan yüz tanıma modellerinin kullandığı durumda çok fazla tanımlanabilir - ve bir kişiyi benzersiz bir şekilde tanımlama amacıyla kullanıldığında, veri koruma yasası kapsamında özel kategori verileri olacaktır.

- **Lokal olarak çıkarımlar yapmak**

Tahmin değişkenlerini paylaşmanın içerdiği riskleri azaltmanın bir başka yolu; ML modelini, sorgunun oluşturulduğu ve bireyin kişisel verilerini halihazırda toplayan ve depolayan cihazda barındırmaktır. Örneğin, bir ML modeli kullanıcının kendi cihazına kurulabilir ve bir bulut sunucusunda barındırılmak yerine 'lokal olarak' çıkarımlar yapabilir.

Örneğin, bir kullanıcının hangi haber içeriğiyle ilgilenebileceğini tahmin etmeye yönelik modeller, akıllı telefonlarında lokal olarak çalıştırılabilir. Kullanıcı haber uygulamasını açtığında, telefona günün haberleri gönderilir ve lokal model, kullanıcının kişisel alışkanlıklarına veya cihazın kendisinde izlenen ve depolanan profil bilgilerine göre kullanıcıya göstermek için en uygun hikayeleri seçer ve içerik sağlayıcı veya uygulama mağazası ile paylaşılmaz.

Kısıtlama, makine öğrenimi modellerinin kullanıcının kendi donanımında çalışması için yeterince küçük ve hesaplama açısından verimli olması gerektiğidir. Ancak, akıllı telefonlar ve gömülü cihazlar için amaca yönelik donanım alanındaki son gelişmeler, bunun giderek daha uygulanabilir bir seçenek olduğu anlamına geliyor.

Lokal işlemenin mutlaka veri koruma yasasının kapsamı dışında olmadığına dikkat etmek önemlidir. Eğitime dahil olan kişisel veriler, kullanıcının cihazında işleniyor olsa bile, modeli oluşturan ve devreye alan organizasyon, işleme araçlarını ve amaçlarını belirlediği ölçüde yine de bir denetleyicidir.

Benzer şekilde, kullanıcının cihazındaki kişisel verilere daha sonra üçüncü bir tarafça erişilirse, bu faaliyet söz konusu verilerin 'işlenmesini' teşkil edecektir.

- **Gizliliği koruyan sorgu yaklaşımları**

Modeli lokal olarak devreye almak mümkün değilse, ML modeline gönderilen bir sorguda ortaya çıkan verileri en aza indirmek için diğer gizliliği artıran teknikler mevcuttur. Bunlar, bir tarafın tüm bu bilgileri modeli çalıştıran tarafa ifşa etmeden bir tahmin veya sınıflandırma almasına izin verir; basit bir ifadeyle, soruyu tam olarak açıklamak zorunda kalmadan bir yanıt almanızı sağlar.

İlave okuma

Bkz. "Privad: çevrimiçi reklamcılıkta pratik gizlilik (Privad: practical privacy in online advertising)" [dış bağlantı] ve Yerel olarak çıkarımlar yapmak için konsept

kanıtı örnekleri için 'El cihazında hedeflenen reklamcılık: gizlilik ve güvenlik zorlukları (Targeted advertising on the handset: privacy and security challenges)' [dış bağlantı].

Gizliliği koruyan sorgu yaklaşımlarına örnek için "TAPAS: güvenilir, gizliliğe duyarlı katılımcı algılama - TAPAS: trustworthy privacy-aware participatory sensing" konusuna bakın.

Anonimleştirmenin bir rolü var mı?

Veri minimizasyonu ve anonimleştirme arasında kavramsal ve teknik benzerlikler vardır. Bazı durumlarda, gizliliği koruma tekniklerinin uygulanması, makine öğrenimi sistemlerinde kullanılan belirli verilerin takma ad veya anonim olarak işlendiğini gösterir.

Bununla birlikte, takma ad kullanmanın esasen bir güvenlik ve risk azaltma tekniği olduğunu ve takma adlı kişisel veriler için veri koruma yasasının hâlâ geçerli olduğunu unutmamalısınız. Buna karşılık, 'anonim bilgi', söz konusu bilgilerin artık kişisel veri olmadığı ve veri koruma kanununun buna uygulanmadığı anlamına gelir.

İlave okuma

ICO şu anda bu alandaki yeni gelişmeleri ve teknikleri dikkate almak için anonimleştirme konusunda yeni bir kılavuz geliştirmektedir.

Eğitim verilerini depolamak ve sınırlamak için ne yapmalıyız?

Bazen, örneğin yeni modelleme yaklaşımları kullanıma sunulduğunda ve hata ayıklama gibi durumlarda, modeli yeniden eğitmek için eğitim verilerinin tutulması gerekebilir. Bununla birlikte, bir modelin oluşturulduğu ve yeniden eğitilmesi veya değiştirilmesinin muhtemel olmadığı durumlarda, eğitim verilerine artık ihtiyaç duyulmayabilir. Model yalnızca son 12 aylık verileri kullanacak şekilde tasarlandıysa, bir veri saklama ilkesi 12 aydan eski verilerin silineceğini belirtmelidir.

İlave okuma

Avrupa Birliği Ağ ve Bilgi Güvenliği Ajansı (ENISA - The European Union Agency for Network and Information Security), araştırma raporları da dahil olmak üzere PET'ler hakkında bir dizi yayına sahiptir. (dış bağlantı)

Kontrol örnekleri

Risk Bildirimi

AI geliştiricileri, AI sistemleri için kullanılan kişisel verilerin; yeterliliğini, gerekliliğini ve uygunluğunu düzgün bir şekilde değerlendirmez ve bu da veri minimizasyon ilkesine uyumsuzlukla sonuçlanır.

Önleyici

- Modele dahil edilen kişisel veri setleri de dahil olmak üzere AI sistemlerinin geliştirilmesi/kullanımı için onay yetkilisinin seviyelerini belgeleyin. Uygun onay kanıtını belgeleyin.
- Model geliştirmenin her aşamasında, verilerin tutulması için ayrıntılı gerekçe ve alakasız verilerin kaldırıldığının/silindiğinin teyidi de dahil olmak üzere, kişisel verilerin uygunluğunu gözden geçirin.
- Veri minimizasyon ilkesinin koşullarına bağlı olarak, AI yaşam döngüsünün farklı aşamalarında verileri ayırın.
- Bir saklama politikası/programı ve programın bağlı olduğuna dair kanıtları belgeleyin (kişisel veriler programa uygun olarak silinir veya programın dışında tutulması gerekçelendirilir ve onaylanır).
- Model girdisi/çıktısının özellikle kişisel veri girdilerinin uygunluğuyla ilgili bağımsız bir inceleme yapın.
- Değerlendirilen PET'lerin kapsamlı bir değerlendirmesini ve neden uygun bulunup bulunmadığını da içeren bir DPIA'yı belgeleyin.

Tespit edici

- Bir AI modelindeki özelliklerin hala alakalı olup olmadıklarını kontrol etmek için periyodik inceleme(ler) yapın, örneğin işlenen kişisel veri miktarını azaltmak amacıyla aynı sonuçların elde edilip edilemeyeceğini görmek için daha az özelliğe sahip diğer sistemlere karşı test edin.
- Sonuç olarak alınan önlemler de dahil olmak üzere kişilerden gelen kişisel hak taleplerini ve şikayetlerini izleyin (hem kişisel düzeyde hem de sınır analizinde).
- Modelin üçüncü taraflarca kullanıldığında veri minimizasyon süreçleriyle uyumlu olup olmadığını periyodik olarak değerlendirin.

Düzeltici

- Gerekli olmayan özellikleri kaldırın / silin.
- Değişiklik için kapsamlı gerekçeler de dahil olmak üzere daha az istilacı bir model seçin.
- Artık gerekli olmayan (örneğin, veri olmadığı için artık tahmine dayalı olarak kullanışlı olmayan) eğitim verilerini kaldırın / silin.
- Uygun PET'leri uygulayın.

Yapay zeka sistemlerimizde kişisel hakları nasıl etkinleştiririz?

Genel bir bakış

AI sistemlerinin geliştirilme ve devreye alınma şekli, kişisel verilerin genellikle olağandışı şekillerde yönetildiği ve işlendiği anlamına gelir. Bu, kişisel hakların bu tür verilere ne zaman ve nasıl uygulanacağını anlamayı zorlaştırabilir ve kişilerin bu hakları kullanmaları için etkili mekanizmalar uygulamak daha zordur.

Detaylı olarak

- Kişisel haklar, AI yaşam döngüsünün farklı aşamalarına nasıl uygulanır?
- Kişisel haklar, bir AI modelinin kendisinde bulunan kişisel verilerle nasıl ilişkilidir?
- Yalnızca yasal veya benzer etkiye sahip otomatikleştirilmiş kararlarla ilgili kişisel hakları nasıl etkinleştiririz?
- İnsan gözetiminin rolü nedir?

Kişisel haklar, AI yaşam döngüsünün farklı aşamalarına nasıl uygulanır?

Veri koruma kanunu kapsamında bireyler, kişisel verileriyle ilgili bir takım haklara sahiptir. Yapay zekada bu haklar; kişisel verilerin, yapay zeka sisteminin geliştirme ve devreye alma yaşam döngüsündeki çeşitli noktalarda kullanıldığı her yerde geçerlidir. Bu nedenle bu, aşağıdaki kişisel verileri kapsar:

- Eğitim verilerinde yer alan.
- Devreye alma esnasında bir tahmin yapmak için kullanılan ve tahminin kendisinin sonucu olan.
- Modelin kendisinde bulunabilen.

Bu bölüm, AI'yı geliştirirken ve devreye alırken; kişisel bilgi, erişim, düzeltme, silme, işlem kısıtlaması, veri taşınabilirliği ve itiraz haklarına uymaya çalışırken karşılaşabileceğiniz hususları açıklamaktadır (GDPR'ın 13-21. maddelerinde bahsedilen haklar). Her hakkı ayrıntılı olarak kapsamaz, ancak bu hakları bir AI bağlamında kolaylaştırmaya yönelik genel zorlukları tartışır ve uygun olduğunda belirli haklara yönelik zorluklardan bahseder.

Kişilerin kendilerini yasal veya benzer şekilde önemli şekillerde etkileyen yalnızca otomatik kararlarla ilgili sahip oldukları haklar, "İnsan gözetiminin rolü nedir?"

bölümünde daha ayrıntılı olarak tartışılmaktadır, çünkü bu haklar AI kullanırken belirli zorluklar doğurmaktadır.

Eğitim verileri için kişisel hak taleplerini nasıl etkinleştirmeliyiz?

ML modelleri oluştururken veya kullanırken, bu modelleri eğitmek için her zaman veri elde etmeniz gerekir.

Örneğin, geçmiş işlemlere dayalı olarak tüketici alımlarını tahmin etmek için bir model oluşturan bir perakendeci, modeli eğitmek için büyük bir müşteri işlemleri veri setine ihtiyaç duyar.

Eğitim verilerinin ilgili olduğu kişileri belirlemek, haklarını sağlamak için potansiyel bir zorluktur. Tipik olarak, eğitim verileri yalnızca geçmiş işlemler, demografik bilgiler veya konum gibi tahminlerle ilgili bilgileri içerir, ancak iletişim bilgilerini veya benzersiz müşteri tanımlayıcılarını içermez. Eğitim verileri aynı zamanda, ML algoritmalarına daha uygun hale getirmek için tipik olarak çeşitli önlemlere tabi tutulur.

Bununla birlikte, bir müşterinin satın alımlarının ayrıntılı bir zaman çizelgesi, işlem geçmişlerindeki iniş ve çıkışların bir özetine dönüştürülebilir.

İstatistiksel bir modeli eğitmek için kullanmadan önce verileri dönüştürme işlemine (örneğin, sayıları 0 ile 1 arasındaki değerlere dönüştürme) genellikle "ön işleme" denir. Bu, "işleme"nin kişisel veriler üzerinde gerçekleştirilen herhangi bir işlem veya işlem dizisi anlamına geldiği, veri korumada terminoloji konusunda kafa karışıklığı yaratabilir. Dolayısıyla "ön işleme" (makine öğrenimi terminolojisinde) hala "işleme"dir (veri koruma terminolojisinde) ve bu nedenle veri koruması hala geçerlidir.

Bu süreçler, kişisel verilerin bir formdan diğerine, potansiyel olarak daha az ayrıntılı bir forma dönüştürülmesini içerdiğinden, eğitim verilerinin belirli bir adlandırılmış kişiye bağlanmasını potansiyel olarak çok daha zor hale getirebilirler. Ancak, veri koruma kanununda bu, verilerin kapsam dışına alınması için mutlaka yeterli görülmemektedir. Bu nedenle, kişilerin haklarını kullanma taleplerine yanıt verirken yine de bu verileri göz önünde bulundurmanız gerekir.

Verilerde ilişkili tanımlayıcılar veya iletişim bilgileri içermese de ve ön işleme yoluyla dönüştürülmüş olsa bile, eğitim verileri yine de kişisel veri olarak kabul edilebilir. Bunun nedeni, ilgili olduğu kişiyi tek başına veya işleyebileceğiniz diğer verilerle birlikte (bir müşterinin adıyla ilişkilendirilemese bile) 'seçmek' için kullanılabilmesidir.

Örneğin, bir satın alma tahmin modelindeki eğitim verileri, bir müşteriye özgü satın alma modelini içerebilir.

Bu örnekte, bir müşteri kendi talebinin bir parçası olarak son satın alımlarının bir listesini sağlayacaksa, organizasyon eğitim verilerinin o kişile ilgili kısmını tanımlayabilir.

Bu tür durumlarda, kimliğini doğrulamak için makul önlemleri aldığınızı ve başka hiçbir istisnanın geçerli olmadığını varsayarak, bir kişinin talebine yanıt vermekle yükümlüsünüz.

Tanımlanabilirlik hakkında daha fazla bilgi için kişisel verinin ne olduğunu belirleme konusunda rehberimize başvurmalısınız.

- **Erişim hakkı**

Eğitim verilerine erişim, düzeltme veya silme taleplerini; bunların gerçekleştirilmesi daha zor olabileceğinden veya talep etme isteği tipik olarak alabileceğiniz diğer erişim taleplerine kıyasla belirsiz olabileceğinden, açıkça asılsız veya aşırı olarak kabul etmemelisiniz.

Yalnızca GDPR'a (Madde 11 uyarınca) uymak amacıyla eğitim verileri içindeki kişileri tanımlamanıza olanak sağlamak için ek kişisel veriler toplamanız veya saklamanız gerekmez. Bu nedenle, eğitim verilerinde kişiyi tanımlayamayacağınız (kişi, kimliğinin belirlenmesini sağlayacak ek bilgiler sağlayamadığı) ve bu nedenle bir talebi yerine getiremeyeceğiniz zamanlar olabilir.

- **Düzeltme hakkı**

Düzeltme hakkı, bir AI sistemini eğitmek için kişisel verilerin kullanımı için de geçerli olabilir. Düzeltme ile ilgili olarak atmanız gereken adımlar, işlediğiniz verilere ve ayrıca bu işlemenin niteliğine, kapsamına, bağlamına ve amacına bağlıdır.

Bir yapay zeka sistemi için eğitim verileri söz konusu olduğunda, işlemenin bir amacı, büyük veri setlerindeki genel kalıpları bulmak olabilir. Bu bağlamda, bir kişi hakkında işlem yapmak için kullanabileceğiniz kişisel verilerle karşılaştırıldığında, birçok veri noktasından yalnızca biri olduğundan, eğitim verilerindeki kişisel yanlışlıkların modelin performansını etkilemesi olası değildir.

Örneğin, yanlış kaydedilmiş bir müşteri teslimat adresini düzeltmenin, eğitim verilerinde aynı yanlış adresi düzeltmekten daha önemli olduğunu düşünebilirsiniz. Gerekçeniz, muhtemelen birincisinin başarısız bir teslimatla sonuçlanabileceği, ancak ikincisinin modelin genel doğruluğunu neredeyse hiç etkilemeyeceği yönündedir.

Ancak uygulamada, düzeltme hakkı, amaçlarınız için daha az önemli olduğunu düşündüğünüz herhangi bir talebi göz ardı etmenize izin vermez.

- **Silme hakkı**

Ayrıca, eğitim verilerinde yer alan kişisel verilerin silinmesine ilişkin talepler de alabilirsiniz. Silme hakkı mutlak olmamasına rağmen, verileri yasal bir zorunluluk veya kamu görevi (her ikisinin de yasal dayanak olması muhtemel değildir) temelinde işlemiyorsanız, aldığınız herhangi bir silme talebini dikkate almanız

gerektiğini unutmayın. AI sistemlerini eğitmek için - daha fazla bilgi için yasal dayanaklar bölümüne bakın).

Bir kişinin kişisel verilerinin eğitim verilerinden silinmesinin, bir AI sisteminin eğitim amaçlarını yerine getirme yeteneğinizi etkilemesi olası değildir. Bu nedenle, kişisel verileri, eğitim veri setinizden silme talebini yerine getirmemeniz için bir gerekçeniz olması olası değildir.

Eğitim verilerinin silinmesi talebine uymak; modellerin kendileri bu verileri içermediği veya çıkarım yapmak için kullanılamadığı sürece (aşağıdaki bölümde ele alacağımız durumlar) bu verilere dayalı tüm ML modellerinin silinmesini gerektirmez.

- **Veri taşınabilirliği hakkı**

Kişiler, işlemin yasal dayanağının rıza veya sözleşme olduğu durumlarda; bir denetleyiciye "sağladıkları" veriler için veri taşınabilirliği hakkına sahiptir. 'Sağlanan veriler', kişinin bir forma bilinçli olarak girdiği verileri değil, aynı zamanda bir hizmeti kullanma sürecinde toplanan davranışsal veya gözlemsel verileri de içerir.

Çoğu durumda, bir modeli eğitmek için kullanılan veriler (örneğin demografik bilgiler veya harcama alışkanlıkları), kişi tarafından "sağlanan" veriler olarak sayılır. Bu nedenle veri taşınabilirliği hakkı, bu işlemenin rıza veya sözleşmeye dayalı olduğu durumlarda geçerli olacaktır.

Bununla birlikte, yukarıda tartışıldığı gibi, genellikle verileri orijinal biçiminden makine öğrenimi algoritmaları tarafından daha etkili bir şekilde analiz edilebilecek bir şeye önemli ölçüde değiştiren ön işleme yöntemleri uygulanır. Bu dönüşümün önemli olduğu durumlarda, elde edilen veriler artık "sağlandı" olarak sayılmayabilir.

Bu durumda veriler, veri taşınabilirliğine tabi olmayacaktır, ancak yine de kişisel veri olarak teşkil eder ve bu nedenle erişim hakkı gibi diğer veri koruma hakları hala geçerlidir. Bununla birlikte, önceden işlenmiş verilerin türetildiği verilerin orijinal biçimi, veri taşınabilirliği hakkına tabidir (kişi tarafından rıza veya sözleşme kapsamında sağlanmışsa ve otomatik yollarla işlenmişse).

- **Bilgilendirilme hakkı**

Kişisel veriler bir AI sistemini eğitmek için kullanılacaksa, kişileri bilgilendirmelisiniz. Bazı durumlarda, eğitim verilerini kişiden almamış olabilirsiniz ve bu nedenle, bunu yaptığınız sırada onları bilgilendirme fırsatınız olmayabilir. Bu gibi durumlarda, 14. maddede belirtilen bilgileri kişiye en geç bir ay olmak üzere makul bir süre içinde vermelisiniz.

Bir yapay zeka sistemini eğitmek amacıyla bir kişinin verilerini kullanmak, normalde yasal veya benzer şekilde önemli etkileri olan yalnızca otomatik bir karar vermeyi teşkil etmediğinden, bu kararları alırken yalnızca bilgi vermeniz gerekir. Ancak yine de ana şeffaflık gereksinimlerine uymanız gerekecektir.

Yukarıda belirtilen nedenlerle, kişisel verileri eğitim verilerinde yer alan kişilerin, kimliğinin tespit edilmesi ve iletişim kurulması zor olabilir. Örneğin, eğitim verileri kimliğinden ve iletişim adreslerinden (hala kişisel veriler olarak kalırken) çıkarılmış olabilir. Bu gibi durumlarda, kişiye doğrudan bilgi sağlamak imkansız olabilir veya orantısız bir çaba gerektirebilir.

Bu nedenle, bunun yerine kişinin hak ve özgürlüklerini ve meşru menfaatlerini korumak için uygun önlemleri almalısınız. Örneğin, AI sisteminizi eğitmek için kullandığınız verileri nereden elde ettiğinizi açıklayan halka açık bilgiler sağlayabilirsiniz.

AI çıktıları için kişisel hak taleplerini nasıl etkinleştirmeliyiz?

Tipik olarak, bir kez uygulandıktan sonra, bir AI sisteminin çıktıları, bir kişinin profilinde depolanır ve onlar hakkında bazı eylemlerde bulunmak için kullanılır.

Örneğin, bir müşterinin bir web sitesinde gördüğü ürün teklifleri, profillerinde depolanan tahmine dayalı modelin çıktısı tarafından yönlendirilebilir. Bu tür veriler kişisel veri oluşturduğunda, erişim, düzeltme ve silme haklarına tabidir. Eğitim verilerindeki kişisel yanlışlıklar ihmal edilebilir bir etkiye sahip olabilirken, bir modelin hatalı çıktısı kişiyi doğrudan etkileyebilir.

Model çıktılarının (veya bunların dayandığı kişisel veri girdilerinin) düzeltilmesi taleplerinin bu nedenle eğitim verilerinin düzeltilmesi taleplerinden daha olasıdır. Bununla birlikte, yukarıda belirtildiği gibi, tahminler, gerçek ifadelerin aksine tahmin puanları olarak tasarlandıysa, yanlış değildir. Kişisel veriler yanlış değilse, düzeltme hakkı geçerli değildir.

Sağlanan verilerin daha fazla analizinden kaynaklanan kişisel veriler, taşınabilirlik hakkına tabi değildir. Bu, AI modellerinin kişilerle ilgili tahminler ve sınıflandırmalar gibi çıktılarının taşınabilirlik hakkı kapsamı dışında olduğu anlamına gelir.

Bazı durumlarda, modeli eğitmek için kullanılan özelliklerin bir kısmı veya tamamı, kişisel verilerin önceki bazı analizlerinin sonucu olabilir. Örneğin, kişinin finansal verilerine dayalı istatistiksel analizin sonucu olan bir kredi puanı daha sonra bir ML modelinde bir özellik olarak kullanılabilir. Bu durumlarda kredi notu, diğer özellikler olsa dahi veri taşınabilirliği hakkı kapsamında değildir.

İlave okuma – ICO kılavuzu

Aşağıdakiler dahil olmak üzere GDPR Kılavuzu'ndaki kişisel haklar konusundaki kılavuzumuzu okuyun:

- bilgilendirilme hakkı;
- erişim hakkı;
- silme hakkı;
- düzeltme hakkı; ve
- veri taşınabilirliği hakkı.

Kişisel haklar, bir AI modelinin kendisinde bulunan kişisel verilerle nasıl ilişkilidir?

Kişisel veriler bir modelin girdi ve çıktılarında kullanılmasının yanı sıra bazı durumlarda modelin kendisinde de yer alabilmektedir. Bölüm 3.2'de açıklandığı gibi, iki nedenden dolayı olabilir; tasarım veya tesadüfen.

Tasarım gereği veri içeren modellerle ilgili talepleri nasıl yerine getirmeliyiz?

Kişisel veriler tasarım gereği modellere dahil edildiğinde, bunun nedeni Destek Vektör Makineleri (SVM- Support Vector Machines) gibi belirli model türlerinin devreye alma esnasında yeni örnekleri ayırt etmeye yardımcı olmak için eğitim verilerinden bazı önemli örnekler içermesidir. Bu durumlarda, modelin iç mantığında bir yerde küçük bir dizi kişisel örnek bulunur.

Eğitim seti tipik olarak yüz binlerce örnek içerir ve bunların yalnızca çok küçük bir yüzdesi doğrudan modelde kullanılır. Bu nedenle ilgili kişilerden birinin bir talepte bulunma olasılığı çok düşüktür; ama mümkün olmaya devam eder.

ML modelinin uygulandığı belirli programlama kütüphanesine bağlı olarak, bu örnekleri kolayca almak için yerleşik bir işlev olabilir. Bu gibi durumlarda, bir kişinin talebine yanıt vermeniz pratik olarak mümkün olabilir. Bunu sağlamak için, tasarım gereği kişisel veri içeren modelleri kullandığınızda, bunları bu örneklerin kolayca bulunabilmesini sağlayacak şekilde uygulamanız gerekir.

İstek verilere erişim içinse, modeli değiştirmeden bunu gerçekleştirebilirsiniz. Talep verilerin düzeltilmesi veya silinmesi için ise, bu, model yeniden eğitilmeden (ya düzeltilen verilerle ya da silinen veriler olmadan) veya model tamamen silinmeden mümkün olmayabilir.

İyi organize edilmiş bir model yönetim sisteminiz ve devreye alma düzeniniz varsa, bu tür istekleri karşılamak ve AI modellerinizi buna göre yeniden eğitmek ve yeniden devreye almak aşırı maliyetli olmamalıdır.

Modellerde yer alan verilerle ilgili istekleri tesadüfen nasıl yerine getirmeliyiz?

Tasarım gereği eğitim verilerinden örnekler içeren SVM'ler ve diğer modellerin yanı sıra; bazı modeller kazara kişisel verileri 'sızdırabilir'. Bu durumlarda, yetkisiz taraflar, eğitim verilerinin öğelerini kurtarabilir veya modelin çalışma şeklini analiz ederek içinde kimin olduğunu çıkarabilir.

Bu senaryolarda erişim, düzeltme ve silme haklarının uygulanması ve yerine getirilmesi zor veya imkansız olabilir. Kişi, modelden kişisel verilerinin çıkarılabileceğine dair kanıt sunmadıkça, kişisel verilerin çıkarılıp

çıkarılamayacağını ve dolayısıyla talebin herhangi bir dayanağı olup olmadığını belirleyemeyebilirsiniz.

Kişisel verilerin en son teknolojinin ışığında modellerden çıkarılma olasılığını düzenli ve proaktif olarak değerlendirmelisiniz, böylece yanlışlıkla ifşa etme riskini en aza indirirsiniz.

Yalnızca yasal veya benzer etkiye sahip otomatikleştirilmiş kararlarla ilgili kişisel hakları nasıl etkinleştiririz?

Veri koruma; kişiler üzerinde yasal veya benzer şekilde önemli bir etkisi olan, yalnızca otomatikleştirilmiş kararlar almak için kişisel verileri işlediğinizde uygun önlemleri sağlamanızı gerektirir. Bu güvenceler, kişilerin aşağıdakileri yapma hakkını içerir:

- İnsan müdahalesi sağlamak.
- Bakış açılarını ifade etmek.
- Haklarında verilen karara itiraz etmek.
- Kararın mantığı hakkında bir açıklama sağlamak.

DPA 2018'in 2. Bölümü kapsamına giren yalnızca otomatik karar vermeyi içeren işlem için, bu tür bir işlemenin yasal dayanağı kanunen bir gereklilik veya yetki ise, bu önlemler GDPR'dekilerden farklıdır.

DPA 2018 Bölüm 3 kapsamına giren, yalnızca otomatikleştirilmiş karar vermeyi içeren işlemler için; bir kişi, kararı yeniden gözden geçirmenizi veya yalnızca otomatik işlemeye dayanmayan yeni bir karar almanızı talep etme hakkına sahip olsa da, uygulanabilir korumalar, otomatik karar vermeye yetki veren ilgili yasada sağlanan düzenlemelere bağlı olacaktır.

Bu korumalar simgesel jestler olamaz. Avrupa Veri Koruma Kurulu (EDPB) tarafından yayınlanan kılavuz, insan müdahalesinin kararın gözden geçirilmesini içermesi gerektiğini belirtir, ki o:

Alıntı

"kararı değiştirmek için uygun yetki ve yeteneğe sahip biri tarafından gerçekleştirilmelidir"

İnceleme ayrıca şunları içermelidir:

Alıntı

"Veri sahibi tarafından sağlanan herhangi bir ek bilgi de dahil olmak üzere ilgili tüm verilerin kapsamlı bir değerlendirmesi."

İnsan müdahalesinin anlamlı olarak nitelendirildiği koşullar, yalnızca otomatik olmayan bir kararı veren koşullara benzer (önceki bölüme bakın). Bununla birlikte, temel fark, tamamen otomatikleştirilmiş durumlarda; bir sistemin yalnızca otomatik olmayan olarak nitelendirilmesi için her kararda anlamlı insan müdahalesi gerekirken, insan müdahalesi yalnızca kişinin haklarını korumak için vaka bazında gereklidir.

Otomatik kararlarla ilgili haklar neden yapay zeka sistemleri için özel bir konu olabilir?

Yalnızca otomatikleştirilmiş kararlar almaya dahil olan sistemlerin türü ve karmaşıklığı, kişilerin veri koruma haklarına yönelik riskin niteliğini ve şiddetini etkiler ve farklı hususların yanı sıra uyumluluk ve risk yönetimi zorluklarını gündeme getirir.

Nispeten az sayıda açıkça yazılmış kuralı otomatikleştiren temel sistemlerin AI olarak değerlendirilmesi pek olası değildir (örneğin, bir müşterinin bir ürün için uygunluğunu belirlemek için açıkça ifade edilen bir dizi "eğer-o zaman" kuralı). Ancak, sonuçta ortaya çıkan kararlar yine de veri koruma kanunu anlamında otomatik karar almayı teşkil edebilir.

Ayrıca, sistemin yüksek yorumlanabilirliği nedeniyle bir karara bir kişi itiraz ederse, inceleyen bir insanın herhangi bir hatayı tespit etmesi ve düzeltmesi nispeten kolay olmalıdır.

Ancak, makine öğrenimine dayalı olanlar gibi diğer sistemler daha karmaşık olabilir ve anlamlı insan incelemesi için daha fazla itiraz sunabilir. ML sistemleri, veri kalıplarına dayalı olarak insanlar hakkında tahminler veya sınıflandırmalar yapar. Son derece istatistiksel olarak doğru olsalar bile, kişisel bir durumda zaman zaman yanlış karara varacaklar. Gözden geçiren bir insanın hataları tanımlaması, anlaması veya düzeltmesi kolay olmayabilir.

Bir kişiden gelen her itiraz, kararın bozulmasıyla sonuçlanmasa da, birçoğunun olabileceğini beklemelisiniz. ML sistemlerinde bunun böyle olmasının iki özel nedeni vardır:

- **Kişi bir "aykırı değer -outlier"dir,** yani koşulları, AI sistemini oluşturmak için kullanılan eğitim verilerinde dikkate alınanlardan önemli ölçüde farklıdır. ML modeli, benzer kişiler hakkında yeterli veri üzerinde eğitilmediği için yanlış tahminler veya sınıflandırmalar yapabilir.
- **AI tasarımındaki varsayımlara itiraz edilebilir,** örneğin yaş gibi sürekli bir değişken, modelleme sürecinin bir parçası olarak 20-39 gibi farklı yaş aralıklarına bölünmüş ('binned') olabilir. Daha ince taneli "binler", farklı yaşlardaki insanlar için önemli ölçüde farklı tahminlere sahip farklı bir modele neden olabilir. Bu veri ön işleme ve diğer tasarım seçeneklerinin geçerliliği, ancak bir kişiin itiraz etmesi sonucu olarak söz konusu olabilir.

Otomatik karar vermeyle ilgili hakları yerine getirmek için hangi adımları atmalıyız?

Aşağıdakileri yapmalısınız:

- Tasarım aşamasından itibaren anlamlı bir insan incelemesini desteklemek için gerekli sistem gereksinimlerini göz önünde bulundurun. Özellikle, insan incelemeleri ve müdahalelerini desteklemek için, yorumlanabilirlik gereksinimleri ve etkili kullanıcı arayüzü tasarımı.
- İnceleyen insanlar için uygun eğitim ve desteği tasarlayın ve sunun.
- Personele, kişilerin endişelerini ele almak veya iletmek için uygun yetkiyi, teşvikleri ve desteği verin ve gerekirse AI sisteminin kararını geçersiz kılın.

Ancak, bilmeniz gereken bazı ek gereksinimler ve hususlar vardır.

ICO'nun ExplAIn kılavuzu, karmaşık AI sistemlerinin kişilere anlamlı açıklamalar sağlama yeteneğinizi nasıl ve ne ölçüde etkileyebileceğini inceler. Bununla birlikte, karmaşık yapay zeka sistemleri, diğer zorunlu korumaların etkinliğini de etkileyebilir. Bir sistem açıklamak için çok karmaşıksa, anlamlı bir şekilde itiraz etmek, müdahale etmek, gözden geçirmek veya karşı alternatif bir bakış açısı koymak için de çok karmaşık olabilir.

Örneğin, bir AI sistemi bir tahmin yapmak için yüzlerce özellik ve karmaşık, doğrusal olmayan bir model kullanıyorsa, o zaman bir kişinin hangi değişkenlere veya korelasyonlara itiraz edeceğini belirlemesi zor olabilir. Bu nedenle, yalnızca otomatikleştirilmiş yapay zeka sistemleri etrafındaki güvenlik önlemleri karşılıklı olarak destekleyicidir ve bütünsel olarak ve kişi göz önünde bulundurularak tasarlanmalıdır.

Bir sistemin mantığı hakkındaki bilgiler ve kararların açıklamaları, kişilere insan müdahalesini talep edip etmeyeceklerine ve hangi gerekçelerle istediklerine karar vermeleri için gerekli şartları vermelidir. Bazı durumlarda yetersiz açıklamalar, kişileri gereksiz yere başka haklara başvurmaya sevk edebilir. Müdahale taleplerinin, görüşlerin ifade edilmesinin veya itirazların, kişilerin karara nasıl ulaşıldığına dair yeterli bir anlayışa sahip olmadıklarını hissetmeleri durumunda gerçekleşmesi daha olasıdır.

Kişilerin haklarını kullanma süreci basit ve kullanıcı dostu olmalıdır. Örneğin, yalnızca otomatik kararın sonucunu bir web sitesi aracılığıyla iletirseniz, sayfa, kişinin herhangi bir gereksiz gecikme veya komplikasyon olmadan, müdahale edebilecek bir personel ile iletişim kurmasına olanak tanıyan bir bağlantı veya net bilgi içermelidir.

Ayrıca, hesap verme sorumluluğunuzun ve belgeleme yükümlülüklerinizin bir parçası olarak bir yapay zeka sistemi tarafından verilen tüm kararların kaydını tutmanız gerekmektedir. Bu ayrıca, bir kişinin insan müdahalesi talep edip etmediğini, herhangi bir görüş ifade edip etmediğini, karara itiraz edip etmediğini ve sonuç olarak kararı değiştirip değiştirmediğinizi de içermelidir.

Bu verileri izlemeli ve analiz etmelisiniz. Kişilerin haklarını kullanmalarına karşılık kararlar düzenli olarak değiştirilirse, sistemlerinizi buna göre nasıl değiştireceğinizi düşünmelisiniz. Sisteminizin makine öğrenimine dayalı olduğu durumlarda, bu, gelecekte benzer hataların daha az olasılıkla olması için düzeltilmiş kararları yeni eğitim verilerine dahil etmeyi içerebilir.

Daha önemli olarak, hatalı karara yol açan boşlukları doldurmak için daha fazla veya daha iyi eğitim verisi toplama ihtiyacını belirleyebilir veya model oluşturma sürecini, yani özellik seçimini değiştirerek modifiye edebilirsiniz.

Bir uyumluluk gerekliliği olmasının yanı sıra, bu aynı zamanda yapay zeka sistemlerinizin performansını iyileştirmeniz ve karşılığında kişilerin onlara güvenini oluşturmanız için bir fırsattır. Ancak, ciddi veya sık yapılan hatalar tespit edilirse, altta yatan sorunları anlamak ve düzeltmek için acil adımlar atmanız ve gerekirse otomatik sistemin kullanımını askıya almanız gerekebilir.

Ayrıca, döngüde bir insana sahip olmanın aşağıdakilere yol açan ödünleşmeleri de vardır: (eğer gözden geçiren insanların yapay zeka tarafından üretilen bir çıktıyı doğrulamak veya reddetmek için ek kişisel verileri dikkate almaları gerekiyorsa), ya mahremiyetin daha fazla bozulmaya uğraması açısından, ya da otomatik bir sürecin sonunda olası insan önyargılarının yeniden ortaya çıkması açısından.

İlave okuma – ICO kılavuzu

GDPR Kılavuzu'ndaki belgeler hakkındaki kılavuzumuzu okuyun.

İlave okuma – Avrupa Veri Koruma Kurulu

29. Madde Çalışma Grubunun (WP29) yerini alan Avrupa Veri Koruma Kurulu (EDPB - The European Data Protection Board), her AB üye devletinin veri koruma yetkililerinden temsilciler içerir. GDPR gerekliliklerine uymak için yönergeleri benimser.

WP29, Şubat 2018'de otomatik karar verme ve profil oluşturma hakkında yönergeler yayınladı. EDPB, Mayıs 2018'de bu yönergeleri onayladı.

Yapay zeka odaklı otomatik bir kararda yer alan mantığı nasıl açıklamalıyız?

Kişiler ayrıca yapay zeka güdümlü otomatik bir kararda yer alan mantık hakkında anlamlı bilgilere sahip olma hakkına sahiptir.

Bu hakka nasıl uyulacağı konusunda daha fazla ayrıntı için, lütfen Alan Turing Enstitüsü ile işbirliği içinde hazırladığımız son "ExplAIn" kılavuzumuza bakın. (Turing).

Kontrol örnekleri

Risk bildirimi

Yapay zeka sistemlerinin altyapısı ve mimarisi; kişisel hak taleplerini tanıma, bunlara yanıt verme ve bunlara göre hareket etme yeteneğini engeller.

Önleyici

- Yapay zeka sistem geliştiricilerinizin, dalında kişisel hakları (IR – Individual Rights) dikkate alma gereksinimini içeren eğitim almalarını sağlayın.
- Modelde IR'nin dikkate alınması da dahil olmak üzere AI sistemlerinin geliştirilmesi/kullanımı için onay yetkisi düzeylerini belirleyin ve belgeleyin. Uygun onayların kanıtını koruyun.
- Veri işleme düzenindeki IR talepleriyle ilgilenmek için bir politika/süreç uygulayın ve belgeleyin, özellikle yanıt vereceğiniz ve vermeyeceğiniz durumları tanımlayın, örneğin sonuça göre eğitim için kullanılan veriler ve kişi üzerinde bir etkinin olduğu yerler.
- Daha karmaşık taleplerin nasıl iletileceği de dahil olmak üzere 'müşteriyle yüz yüze olan' bireyler için kişisel haklar eğitimi sağlayın.
- DPIA'nıza kapsamlı bir veri akışı eşlemesi ekleyin.
- IR talepleri bekleniyorsa, sistem tasarımının bir parçası olarak veri indekslemeyi ve ortak tanımlayıcıları kullanarak sistemlerinizi aranabilir hale getirmeyi düşünün.
- Tam otomatik karar verme için, bir kararı araştırması gerekebilecek kişilerin, kararı araştırmak ve geçersiz kılmak için becerilere, araçlara ve özerkliğe sahip olduğundan emin olun.
- Veri öznelerinin, bir yapay zeka sistemini eğitmek veya profillerini çıkarmak amacıyla verilerinin işlenmesi hakkında bilgilendirildiğinden emin olun. Eğitim için kullanılan verilerin başka bir organizasyona ait olduğu ve ilgili kişiyle doğrudan bir ilişkinin bulunmadığı ve doğrudan bilgilendirilmesinin orantısız bir çaba gerektireceği durumlarda; bu bilgileri kamuya açık hale getirdiğinizden ve verileri temin ettiğiniz organizasyonların veri konularını işleme hakkında bilgilendirmek için süreçlere sahip olduğundan emin olun.
- AI sistemlerinizde tanımlanmaları için, kişilere ek veri sağlama araçlarının verildiğinden emin olun.
- Üçüncü taraflarla, özellikle de roller ve sorumluluklar (denetçi / işleyici) ile iş yapmak için belgelenmiş politikaları / süreçleri koruyun.

Tespit edici

- Tüm eylemlerin gerektiği gibi tamamlandığından emin olmak için meslektaş incelemeleri yapın.
- Doğru ve eksiksiz olduğundan ve reddedildiği durumlarda gerekçenin (örneğin açık olarak asılsız) uygun olduğundan emin olmak için örnek IR taleplerini periyodik olarak gözden geçirin.

- Potansiyel olarak daha karmaşık olan sistemleri belirlemek için isteklere yanıt vermede geçen süreyi sistematik olarak izleyin.
- Süreci test etmek ve sonuçları ölçmek için "sahte" IR istekleri gönderin.

Düzeltici

- AI sistemini / veri depolamayı / eğitim verilerinin indekslenmesini yeniden tasarlayın.
- Bir talep durumunda ek verilerin veri konularını belirlemeye yardımcı olup olmayacağını değerlendirin.
- Herhangi bir yanlış kişisel veriyi düzeltin ve gerçekle ilgili olarak yanıltıcı olmaması için çıkarılan verileri ele alın.
- Gerekirse kişisel verileri silin.
- IR taleplerinin belirlenmesi ve yürütülmesinden sorumlu çalışanları yeniden eğitin.
- İnsanları yeniden eğitin / kaynak gereksinimlerinin yeniden değerlendirin (örneğin, insanlar y zamanında x karar vermeleri için baskı altındaysa).
- Yapay zekayı yeniden tasarlayın, örneğin basitleştirin/uyarıları/açılır pencereleri dahil edin.
- Değişiklik için kapsamlı gerekçeler de dahil olmak üzere daha uygun bir model seçin.
- Yeniden gözden geçirin; bozma kararları (örn. bir hileli incelemeci) ve kişiler üzerindeki etkilerin daha geniş bir değerlendirmesi de dahil olmak üzere sonuç olarak alınan herhangi bir eylem.

İnsan gözetiminin rolü nedir?

AI, kişiler hakkında yasal veya benzer şekilde önemli kararları bildirmek için kullanıldığında, bu kararların uygun insan gözetimi olmadan alınma riski vardır. Bu, GDPR'nin 22. Maddesini ihlal eder.

Bu riski azaltmak için, insan gözetimi sağlamakla görevlendirilen kişilerin ilgili ve kritik durumda kalmasını ve uygun olan her yerde sistemin çıktılarına mücadele edebilmesini sağlamalısınız.

Tamamen otomatik ve kısmen otomatik karar verme arasındaki fark nedir?

AI sistemlerini iki şekilde kullanabilirsiniz:

- Sistemin otomatik olarak karar verdiği **otomatik karar verme** (ADM) için.
- **Karar desteği** olarak, sistemin yalnızca kendi görüşlerinde bir insan karar vericiyi **desteklediği** durumlarda.

Örneğin; yapay zekayı, bir finansal krediyi otomatik olarak onaylayan ya da reddeden bir sistemde veya yalnızca bir kredi başvurusunu verip vermemeye

karar verirken bir kredi görevlisini desteklemeye yönelik ek bilgi sağlamak için kullanabilirsiniz.

Tamamen otomatik yapay zeka odaklı karar alma, yapay zeka destekli insan karar alma mekanizmasından belirli koşullara bağlı olarak daha fazla veya daha az riskli olabilir. Bu nedenle bunu kendi durumunuza göre değerlendirmeniz gerekir.

Göreceli değerlerine bakılmaksızın, otomatik kararlar, veri koruma yasasında insan kararlarından farklı şekilde ele alınır, özellikle, GDPR'nin 22. Maddesi, kişiler üzerinde yasal veya benzer şekilde önemli etkileri olan tam otomatik kararları daha sınırlı bir dizi yasal dayanakla sınırlandırır ve belirli önlemlerin alınmasını gerektirir.

Buna karşılık, yalnızca insan karar vermesini **destekleyen** karar destek araçlarının kullanımı bu koşullara tabi değildir. Bu kısıtlamaların ve güvencelerin bir sonucu olarak, otomatik karar vermenin insan karar vermesinden daha yüksek bir risk taşıdığı tartışılabilir (bazı durumlarda insan karar verme risklerinin bazılarını hafifletebilse de).

Yapay zekayı yalnızca insan karar verme sürecini **desteklemek** için kullanmaya karar verirseniz, bir insan karara "kauçuk damga- bakmadan imzalama- rubber-stamped" vurdu diye kararın 22. Madde kapsamı dışında kalmadığını bilmelisiniz. İnsan girdisinin **anlamlı** olması gerekir. Bir kişi hakkında nihai bir karar verilmeden önce insan incelemesinin ve müdahalesinin derecesi ve kalitesi; bir AI sisteminin otomatik karar verme için mi yoksa yalnızca karar desteği olarak mı kullanıldığını belirlemede kilit faktörlerdir.

Bu durumlarda insan girdisinin anlamlı olmasını sağlamak sadece sistemi kullanan kişinin sorumluluğunda değildir. Diğerlerinin yanı sıra kıdemli liderler, veri bilimcileri, işletme sahipleri ve gözetim işlevlerinin, AI uygulamalarının amaçlandığı şekilde tasarlanmasını, oluşturulmasını ve kullanılmasını sağlamada aktif bir rol oynamaları beklenir.

Karar destek araçları olarak tasarlanmış ve bu nedenle Madde 22'nin kapsamı dışında olması amaçlanan AI sistemlerini devreye alıyorsanız, bu konularda hem ICO'dan hem de EDPB'den gelen mevcut rehberin farkında olmalısınız.

Önemli hususlar şunlardır:

- İnceleyen insanlar, sistemin tavsiyesini kontrol etmede yer almalı ve otomatik tavsiyeyi sadece rutin bir şekilde bir kişiye uygulamamalıdır.
- İnceleyenlerin katılımı, yalnızca simgesel bir jest değil, aktif olmalıdır. Tavsiyeye karşı çıkmak için "yetki ve yeterlilik" dahil olmak üzere karar üzerinde gerçek "anlamlı" bir etkiye sahip olmalıdırlar.
- İnceleyenler tavsiyeyi 'tartmalı' ve 'yorumlamalı', mevcut tüm girdi verilerini dikkate almalı ve ayrıca diğer ek faktörleri de hesaba katmalıdır.

Mevzuattaki ilgili hükümler

Bknz. GDPR'ın 22 madde ve 71 gerekçe (dış bağlantı)

Bknz. DPA 2018 14, 49 and 50 bölümleri (dış bağlantı)

İlave okuma – ICO kılavuzu

GDPR Kılavuzu'nda otomatik karar verme ve profil oluşturma konusundaki kılavuzumuzu okuyun.

İlave okuma – Avrupa Veri Koruma Kurulu

29. Madde Çalışma Grubunun (WP29) yerini alan Avrupa Veri Koruma Kurulu (EDPB - The European Data Protection Board), her AB üye devletinin veri koruma yetkililerinden temsilciler içerir. GDPR gerekliliklerine uymak için yönergeleri benimser.

WP29, Şubat 2018'de otomatik karar verme ve profil oluşturma hakkında yönergeler yayınladı. EDPB, Mayıs 2018'de bu yönergeleri onaylad

Yapay zeka sistemlerindeki ek risk faktörleri nelerdir?

Ne kadar basit olursa olsun, kullandığınız herhangi bir otomatik karar verme sisteminde insan girdisinin anlamlılığını göz önünde bulundurmalısınız.

Bununla birlikte, daha karmaşık yapay zeka sistemlerinde; karar desteği olarak tasarlanan bir sistemin, anlamlı insan girdisi sağlamada istemeden başarısız olmasına ve dolayısıyla 22. Maddenin kapsamına girmesine potansiyel olarak neden olabilecek iki ek faktör vardır. Bunlar:

- Otomasyon yanlılığı.
- Yorumlanabilirlik eksikliği.

'Otomasyon yanlılığı' ne anlama geliyor?

AI modelleri matematik ve verilere dayanmaktadır. Bu nedenle, insanlar bunları nesnel olarak düşünmeye ve ne kadar doğru olursa olsun çıktılarına güvenme eğilimindedir.

Otomasyon yanlılığı veya **otomasyon kaynaklı kayıtsızlık** terimleri, insan kullanıcılarının bir karar destek sistemi tarafından üretilen çıktıya rutin olarak nasıl güvendiğini ve kendi yargılarını kullanmayı veya çıktının yanlış olup olmadığını sorgulamayı nasıl bıraktığını tanımlar.

'Yorumlanabilirlik eksikliği' ne anlama geliyor?

Bazı yapay zeka sistem türleri, örneğin karmaşık, yüksek boyutlu 'derin öğrenme' modellerine dayananlar gibi, bir insan incelemesinin yorumlaması zor olan çıktılara sahip olabilir.

Yapay zeka sistemlerinin çıktıları kolayca yorumlanamıyorsa ve diğer açıklama araçları mevcut veya güvenilir değilse, bir insanın bir yapay zeka sisteminin çıktısını anlamlı bir şekilde değerlendirememe ve bunu kendi karar verme süreçlerine dahil edememe riski vardır.

Anlamlı incelemeler mümkün değilse, gözden geçiren kişi, yargılamadan veya sorgulamadan sistemin tavsiyelerine katılmaya başlayabilir. Bu, sonuçta ortaya çıkan kararların etkin bir şekilde "yalnızca otomatikleştirilmiş" olduğu anlamına gelir.

Yalnızca otomatikleştirilmiş AI sistemlerini, otomatikleştirilmiş olmayanlardan ayırmalı mıyız?

Evet. Herhangi bir AI sisteminin kullanım amacı hakkında en baştan net bir görüş almalısınız. Yapay zekayı insan karar verme sürecini desteklemek veya geliştirmek için mi yoksa yalnızca otomatik kararlar almak için mi kullandığınızı açıkça belirtmeli ve belgelemelisiniz.

Üst yönetiminiz, organizasyonunuzun risk iştahıyla uyumlu olduğundan emin olarak herhangi bir yapay zeka sisteminin kullanım amacını gözden geçirmeli ve imzalamalıdır. Bu, üst yönetimin her bir seçeneğe ilişkin temel risk etkileri konusunda sağlam bir anlayışa sahip olması ve zorluk derecesine uygun olarak hazır ve donanımlı olması gerektiği anlamına gelir.

Ayrıca, başlangıçtan itibaren net hesap verebilirlik çizgileri ve etkin risk yönetimi politikalarının yürürlükte olduğundan emin olmalısınız. AI sistemleri yalnızca insan kararlarını desteklemeyi amaçlıyorsa, politikalarınız özellikle otomasyon yanlılığı ve yorumlanabilirlik eksikliği gibi ek risk faktörlerini ele almalıdır.

Aşağıdakiler mümkündür:

- Kısmen veya tamamen otomatikleştirilmiş bir yapay zeka uygulamasının ihtiyaçlarınızı en iyi şekilde karşılayıp karşılamayacağını önceden bilemeyebilirsiniz.
- Tam otomatik bir yapay zeka sisteminin, işleminizin amaçlanan sonucunu daha tam olarak elde edeceğine, ancak kişiler için kısmen otomatik bir sisteme göre daha fazla risk taşıyabileceğinin farkında olun.

Bu durumlarda, risk yönetimi politikalarınız ve DPIA'larınız bunu açıkça yansıtmalı ve AI sisteminin yaşam döngüsü boyunca her seçenek için risk ve kontrolleri içermelidir.

Otomasyon yanlılığı risklerini nasıl ele alabiliriz?

İnsan inceleyicilerin eğitiminin ve izlenmesinin etkinliğini artırarak, otomasyon yanlılığını ele alabileceğinizi düşünebilirsiniz. Eğitim, etkili AI risk yönetiminin kilit bir bileşeni olsa da, kapsam belirleme ve tasarım aşamalarının yanı sıra geliştirme ve devreye alma dahil olmak üzere, projenin başlangıcından itibaren otomasyon yanlılığını azaltmak için kontrollere de sahip olmalısınız.

Tasarım ve inşa aşamasında, organizasyonunuzun tüm ilgili bölümleri (örneğin işletme sahipleri, veri bilimcileri ve varsa gözetim işlevleri), başlangıçtan itibaren anlamlı bir insan incelemesini destekleyen tasarım gereksinimleri geliştirmek için birlikte çalışmalıdır.

AI sisteminin hangi özellikleri dikkate almasını beklediğinizi ve kararlarını kesinleştirmeden önce insan incelemelerin hangi ek faktörleri dikkate alması gerektiğini düşünmelisiniz. Örneğin, yapay zeka sistemi, bir iş başvurusunda bulunan kişinin kaç yıllık deneyime sahip olduğu gibi niceliksel olarak ölçülebilir özellikleri dikkate alırken, bir insan inceleme uzmanı bir uygulamanın diğer yönlerini niteliksel olarak değerlendirir (örneğin, yazılı iletişim).

İnsan yorumcular, yalnızca AI sistemi tarafından kullanılan verilere erişebiliyor veya bunları kullanabiliyorsa, muhtemelen diğer ek faktörleri hesaba katmıyorlardır. Bu, incelemelerin yeterince anlamlı olmayabileceği ve kararın "tamamen otomatik" olarak değerlendirilebileceği anlamına gelir.

Gerektiğinde, insan inceleyiciler tarafından değerlendirilmek üzere ek faktörleri nasıl yakalayacağınızı düşünmelisiniz. Örneğin, bu tür bilgileri toplamak için kararın verildiği kişiyle doğrudan etkileşime girebilirler.

Bir AI sisteminin ön arayüzünü tasarlamaktan sorumlu olanlar, insan inceleyicilerin ihtiyaçlarını, düşünce sürecini ve davranışlarını anlamalı ve etkin bir şekilde müdahale etmelerini sağlamalıdır. Bu nedenle, erken aşamada insan inceleyicilere danışmak ve seçenekleri test etmek faydalı olabilir.

Ancak, kullandığınız yapay zeka sistemlerinin özellikleri, mevcut verilere, seçilen model(ler)in türüne ve diğer sistem oluşturma seçeneklerine de bağlıdır. AI sistemi eğitildikten ve oluşturulduktan sonra tasarım aşamasında yapılan varsayımları test etmeniz ve onaylamanız gerekir.

Yorumlanabilirlik risklerini nasıl ele alabiliriz?

Tasarım aşamasından itibaren yorumlanabilirliği de göz önünde bulundurmalısınız. Bununla birlikte, yorumlanabilirliği mutlak terimlerle tanımlamak zordur ve farklı şekillerde ölçülebilir. Örneğin, insan inceleyicisi şunları yapabilir:

- Farklı girdiler verildiğinde sistemin çıktılarının nasıl değişeceğini tahmin edin.
- Belirli bir çıktıya katkıda bulunan en önemli girdileri belirleyin.

- Çıktının ne zaman yanlış olabileceğini belirleyin.

Bu nedenle, yorumlanabilirliğin ne anlama geldiğini ve bunun nasıl ölçüleceğini, kullanmak istediğiniz her bir AI sistemin ve sistemin işleyeceği kişisel verilerin özel durumuna göre tanımlamanız ve belgelemeniz önemlidir.

Bazı AI sistemleri diğerlerinden daha yorumlanabilir. Örneğin, az sayıda insan tarafından yorumlanabilir özellik (örneğin yaş ve ağırlık) kullanan modellerin yorumlanması, çok sayıda özellik kullanan modellerden daha kolay olabilir.

Girdi özellikleri ile modelin çıktısı arasındaki ilişki de basit veya karmaşık olabilir. Karar ağaçlarında olduğu gibi, belirli çıkarımların yapılabileceği koşulları belirleyen basit kuralların yorumlanması daha kolaydır.

Benzer şekilde, (çıktı değerinin girdiyle orantılı olarak arttığı) doğrusal ilişkilerin yorumlanması, doğrusal olmayan (çıktı değerinin girdiyle orantılı olmadığı) veya monoton olmayan (çıkış değerinin arttığı durumlarda) ilişkilerden daha kolay olabilir. (girdi arttıkça, çıktı değeri artabilir veya azalabilir).

Düşük yorumlanabilirliği ele almak için bir yaklaşım; genel olarak modelden ziyade belirli bir çıktının açıklamasını sağlayan Lokal Yorumlanabilir Model-agnostik Açıklama (LIME- Local Interpretable Model-agnostic Explanation) gibi yöntemleri kullanarak 'lokal' açıklamaların kullanılmasıdır.

LIME'ler, yorumlamaya çalıştığınız sistemdekilere benzer girdi ve çıktı çiftleri arasındaki ilişkileri özetlemek için daha basit bir vekil model kullanır. Kişisel tahminlerin özetlerine ek olarak, LIME'ler bazen hataları tespit etmeye yardımcı olabilir (örneğin, bir görüntünün hangi belirli bölümünün, bir modelin onu yanlış sınıflandırmasına neden olduğunu görmek için).

Ancak, AI sisteminin ve çıktılarının altında yatan mantığı temsil etmezler ve yanlış kullanıldığında yanıltıcı olabilirler. Bu nedenle, kendi bağlamınızda, LIME ve benzer yaklaşımların, insan karar vericinin AI sistemini ve çıktısını anlamlı bir şekilde yorumlamasına yardımcı olup olmayacağını değerlendirmelisiniz.

Birçok istatistiksel model, her bir çıktının yanında, bir insan inceleyicinin kendi karar vermesinde yardımcı olabilecek bir güven puanı sağlamak üzere tasarlanabilir. Daha düşük bir güven puanı, insan inceleyicisinin nihai karar için daha fazla girdiye sahip olması gerektiğini gösterir. (Bkz. ‘İstatistiksel doğruluk için ne yapmamız gerekiyor?’)

Yorumlanabilirlik gereksinimlerinin değerlendirilmesi, tasarım aşamasının bir parçası olmalı ve gerektiğinde sistemin bir parçası olarak açıklama araçları geliştirmenize izin vermelidir.

Bu nedenle risk yönetimi politikalarınız, yapay zeka kullanan her işleme işlemi için sağlam, risk tabanlı ve bağımsız bir onay süreci oluşturmalıdır. Ayrıca, devreye alınmadan önce sistemin test edilmesinden ve nihai onaylanmasından kimin sorumlu olduğunu da açıkça belirtmelidirler. Bu kişiler, yorumlanabilirlik ve insan incelemelerinin etkinliği üzerindeki herhangi bir olumsuz etkiden

sorumlu olmalı ve yalnızca AI sistemleri benimsenen risk yönetimi politikasına uygunsa onay sağlamalıdır.

Bu riskleri ele almak için personelimizi nasıl eğitmeliyiz?

Personelinizi eğitmek, bir yapay zeka sisteminin yalnızca otomatik olmayan bir sistem olarak kabul edilmesini sağlamak için çok önemlidir. Başlangıç noktası olarak, insan inceleyicilerinizi aşağıdakiler için eğitmelisiniz (veya yeniden eğitmelisiniz):

- Bir AI sisteminin nasıl çalıştığını ve sınırlamalarını anlama.
- Sistemin ne zaman yanıltıcı veya yanlış olabileceğini ve nedenini tahmin edebilme.
- Onlara dikkate almaları gereken faktörlerin bir listesini sağlayın ve kendi uzmanlıklarının sistemi nasıl tamamlaması gerektiğini anlamalarını sağlayın.
- Kendi uzmanlıklarının sistemi nasıl tamamlaması gerektiğini anlamaları ve onlara dikkate almaları gereken faktörlerin bir listesini sağlayın.
- AI sisteminin çıktısını (sorumlu olmaları gereken bir karar) reddetmek veya kabul etmek için anlamlı açıklamalar sağlayın. Ayrıca, yerinde net bir eskalasyon politikanız olmalıdır.

Eğitimin etkili olabilmesi için şunlar önemlidir:

- İnsan inceleyiciler, AI sistemi tarafından üretilen çıktıyı geçersiz kılma yetkisine sahiptir ve bunu yaptıkları için cezalandırılmayacaklarından emindirler. Bu yetki ve güven, yalnızca politikalar ve eğitimle oluşturulamaz: destekleyici bir organizasyon kültürü de çok önemlidir.
- Herhangi bir eğitim programı, teknolojik gelişmeler ve süreçlerdeki değişiklikler doğrultusunda güncel tutulur ve uygun olduğunda, insan inceleyicilere belirli aralıklarla 'tazeleme' eğitimi verilir.

Burada insan inceleyicilerin eğitimine odaklandık; ancak, başka herhangi bir işlevin etkin gözetim sağlamak için ek eğitim gerektirip gerektirmediğini de (örneğin risk veya iç denetim) dikkate almanız gerektiğini belirtmekte fayda var.

Hangi izlemeyi yapmalıyız?

Bir insan inceleyicinin yapay zeka sisteminin çıktısını neden ve kaç kez kabul ettiğinin veya reddettiğinin analizi, etkili bir risk izleme sisteminin önemli bir parçasıdır.

Risk izleme raporları, insan inceleyicilerinizin rutin olarak AI sisteminin çıktılarını kabul ettiğini gösteriyorsa ve bunları gerçekten değerlendirdiklerini gösteremiyorsa, kararları GDPR kapsamında etkin bir şekilde yalnızca otomatikleştirilmiş olarak sınıflandırılabilir.

Riski hedef seviyelerde tutmak için kontrollere sahip olmanız gerekir. Sonuçlar hedef seviyelerin ötesine geçtiğinde, uyumluluğu hızla değerlendirmek ve gerekirse harekete geçmek için süreçleriniz olmalıdır. Bu, insan denetimini geçici olarak artırmayı veya karar vermenin etkin bir şekilde tam otomatik hale gelmesi durumunda uygun bir yasal temele ve güvencelere sahip olmanızı sağlamayı içerebilir.

İlave okuma – ICO kılavuzu

Yapay zeka sistemlerini açıklama ve yorumlama yöntemleri hakkında daha fazla bilgi için explAIn taslak kılavuzumuzu okuyun.

Kontrol örnekleri

Risk Bildirimi

Yanlış bir şekilde tam otomatik değil olarak sınıflandırılan yapay zeka sistemleri, anlamlı insan gözetimi eksikliğine ve DP mevzuatına uyumsuzluk potansiyeline neden olur.

Önleyici

- Onay makamı seviyesi de dahil olmak üzere, Madde 22 ile ilgili olarak AI sistemlerinin sınıflandırılması hakkında bir politika/süreç uygulayın ve belgeleyin. Karar verme süreci ve uygun imza/onay kanıtlarını muhafaza edin.
- AI sistem kararına itiraz etme ve bağımsız bir inceleme sağlama dahil olmak üzere anlamlı bir gözetim sağlamak için istihdam edilen insanlara eğitim sağlayın.
- Yapay zeka sistem geliştiricilerinin, yapay zeka sistemini tasarlarken insan inceleyicilerin becerilerini, deneyimini ve yeteneklerini anladığından emin olun.
- Anlamlı olduğundan emin olmak için insan incelemelerinin bir değerlendirmesini uygulama öncesi testlerine dahil edin.
- İnsan inceleyicilerin yorumlayabilmesini ve sorgulayabilmesini sağlamak için, özellikle model karmaşıklığıyla ilgili olarak, AI sistemlerinin geliştirilmesi/kullanımı için onay yetkisi düzeylerini belirleyin ve belgeleyin. Uygun onayın kanıtını koruyun.
- Bir insanın anlamlı bir şekilde gözden geçirmesi için beklenen sürenin analizini yapın ve belgeleyin.

Tespit edici

- Uygulama sonrası testleri gerçekleştirin ve testlerin sonuçlarını ve sonuç olarak alınan eylem(ler)i belgeleyin.

- İnsanın doğru kararı verdiğinden emin olmak için bir karar örneğini test edin. Numunenin nasıl seçildiği / kullanılan kriterler de dahil olmak üzere bu tür testleri belgeleyin.
- Yapay zeka tarafından verilen kararları izleyin ve bunları insan kararlarıyla karşılaştırın ve tanımlanan toleransların dışına çıkan performans sonucunda gerçekleştirilen her türlü eylemi belgeleyin.
- Girdilerin anlamlı olduğundan emin olmak için, insanların AI ile aynı fikirde olmaması gereken kasıtlı olarak yanıltıcı verileri periyodik olarak sağladığınız 'gizli alışveriş' alıştırmaları yapın ve belgeleyin.
- Sonuç olarak alınan önlemler de dahil olmak üzere (hem kişisel düzeyde hem de sınır analizinde) özellikle 22. Madde ile ilgili olarak kişilerden gelen kişisel hak talep ve şikayetleri izleyin.
- AI sonucunu tersine çevirmede insan güveninin periyodik bir değerlendirmesini yapın.
- Aykırı değerleri ve sonuç olarak alınan eylemi belirlemek için kişilerin performansını izleyin.

Düzeltici

- İnsan karar vericileri yeniden eğitin, kaynak gereksinimlerini yeniden değerlendirin (örneğin, insanlar kısa sürede çok fazla karar vermeleri için baskı altındaysa).
- Yapay zekayı yeniden tasarlayın, örneğin basitleştirme/uyarıların/açılır pencerelerin dahil edilmesi.
- Daha uygun bir model seçin ve değişiklik için kapsamlı bir gerekçe ekleyin.
- Kararları (örn. Sahte bir gözden geçirmeniz varsa) ve bunun sonucunda alınan her türlü eylemi, kişiler üzerindeki etkilerin daha kapsamlı bir şekilde değerlendirilmesi de dahil olmak üzere yeniden gözden geçirin.