Introduction

Veri seçmeye ve modelleri eğitmeye başlamadan önce, bir yapay zekâ sisteminin hangi insan ihtiyaçlarına hizmet etmesi gerektiğini ve hatta inşa edilip edilmemesi gerektiğini dikkatlice değerlendirmek önemlidir.

**İnsan merkezli tasarım (İMT)**, insanların ihtiyaçlarına hizmet eden sistemler tasarlamaya yönelik bir yaklaşımdır.

Bu eğitimde, yapay zekâ sistemlerine İMT'yi nasıl uygulayacağınızı öğreneceksiniz. Ardından, bilginizi gerçek dünyadan ilginç senaryolardaki tasarım sorunlarına İMT'yi uygulayarak bir **egzersiz** ile test edeceksiniz.

# Approach[¶](https://www.kaggle.com/code/var0101/human-centered-design-for-ai" \l "Approach" \t "_self)

İnsan merkezli tasarım (İMT), insanları tasarım sürecinin her adımına dahil eder. Ekibiniz, yapay zekâya yönelik İMT yaklaşımını mümkün olduğunca erken benimsemelidir; ideal olarak, bir yapay zekâ sistemi inşa etme olasılığını düşünmeye başladığınız andan itibaren.

Aşağıdaki altı adım, yapay zekâ sistemlerinin tasarımına İMT'yi uygulamaya başlamanıza yardımcı olmayı amaçlamaktadır. Ancak, İMT'nin sizin için ne anlama geldiği, sektörünüze, kaynaklarınıza, organizasyonunuza ve hizmet etmeyi amaçladığınız insanlara bağlı olacaktır.

## **1. Understand people’s needs to define the problem**

İnsanlarla çalışarak, mevcut süreçlerindeki sıkıntı noktalarını anlamak, ele alınmamış ihtiyaçları bulmaya yardımcı olabilir. Bu, insanları mevcut araçlarda gezinirken gözlemleyerek, görüşmeler yaparak, odak grupları oluşturarak, kullanıcı geri bildirimlerini okuyarak ve diğer yöntemlerle gerçekleştirilebilir. Veri bilimcileri ve mühendisler de dahil olmak üzere tüm ekibiniz bu adıma dahil olmalıdır, böylece her ekip üyesi hizmet etmeyi umdukları insanları anlar. Ekibiniz, ırk, cinsiyet ve diğer özellikler açısından farklı bakış açılarına ve geçmişlere sahip insanları içermeli ve onları sürece dahil etmelidir. Problem tanımınızı keskinleştirin ve yaratıcı ve kapsayıcı çözümleri birlikte beyin fırtınası yaparak bulun.

Bir şirket, karaciğer nakli sonrası hastalara verilen immünosupresan ilaçlardaki dozaj hataları sorununu çözmek istiyor. Şirket, işe karaciğer nakli süreci boyunca doktorları, hemşireleri ve diğer hastane personelini gözlemleyerek başlar. Ayrıca, yayınlanmış yönergelere ve insan kararına dayanan mevcut dozaj belirleme süreci hakkında onlarla röportaj yapar ve röportajlardan alınan video klipleri tüm geliştirme ekibiyle paylaşır. Şirket ayrıca araştırma çalışmalarını gözden geçirir ve eski hastalar ile ailelerinden oluşan odak grupları oluşturur. Tüm ekip üyeleri, potansiyel çözümler için serbest bir beyin fırtınası oturumuna katılır.

## **2. Ask if AI adds value to any potential solution[¶](https://www.kaggle.com/code/var0101/human-centered-design-for-ai" \l "2.-Ask-if-AI-adds-value-to-any-potential-solution" \t "_self)**

Hangi ihtiyaca yönelik çalıştığınız ve bunu nasıl yapacağınız netleştikten sonra, yapay zekânın değer katıp katmadığını değerlendirin.

* Genellikle insanlar, ulaşmaya çalıştığınız sonucun iyi bir sonuç olduğu konusunda hemfikir midir?
* Oluşturulması, denetlenmesi ve sürdürülmesi daha kolay olan kural tabanlı çözümler gibi yapay zekâ dışı sistemler, bir yapay zekâ sisteminden **önemli ölçüde** daha mı az etkili olurdu?
* Yapay zekâyı kullandığınız görev, insanların sıkıcı, tekrarlayıcı veya odaklanması zor bulacağı bir görev mi?
* Yapay zekâ çözümleri, geçmişte benzer kullanım durumları için diğer çözümlerden daha iyi olduğunu kanıtladı mı?

Bu sorulardan herhangi birine hayır cevabı verdiyseniz, bir yapay zekâ çözümü gerekli veya uygun olmayabilir.

Bir afet yardım kuruluşu, seller gibi afetlerde insanları kurtarma süresini azaltmak için ilk müdahale ekipleriyle birlikte çalışıyor. Mahsur kalmış insanları bulmak için drone ve uydu fotoğraflarının zaman ve emek yoğun insan incelemesi, kurtarma süresini artırıyor. Daha hızlı kurtarmalar daha fazla hayat kurtarabileceği için, fotoğraf incelemesini hızlandırmanın iyi bir sonuç olacağı konusunda herkes hemfikir. Kuruluş, bu görev için bir yapay zekâ görüntü tanıma sisteminin, yapay zekâ dışı otomatik bir sistemden muhtemelen daha etkili olacağına karar veriyor. Ayrıca, yapay zekâ tabanlı görüntü tanıma araçlarının tarım gibi diğer sektörlerde de hava çekimlerini incelemek için başarıyla uygulandığının farkında. Bu nedenle, kuruluş yapay zekâ tabanlı bir çözüm olasılığını daha fazla araştırmaya karar veriyor.

## **3. Consider the potential harms that the AI system could cause[¶](https://www.kaggle.com/code/var0101/human-centered-design-for-ai" \l "3.-Consider-the-potential-harms-that-the-AI-system-could-cause" \t "_self)**

Bir yapay zekâ sisteminin tasarım aşamasından (veri toplama ve etiketlemeden, modelin eğitilmesine ve sistemin devreye alınmasına kadar) yapay zekâyı kullanmanın faydalarını potansiyel zararlarına karşı tartın. Kullanıcılar ve toplum üzerindeki etkisini göz önünde bulundurun. Gizlilik ekibiniz, gizli kalmış gizlilik sorunlarını ortaya çıkarmaya ve **ayrımlı gizlilik** (differential privacy) veya **birleşik öğrenme** (federated learning) gibi gizliliği koruyucu tekniklerin uygun olup olmadığını belirlemeye yardımcı olabilir. Veri seçiminde, model eğitiminde ve sistemin çalışmasında insanları - ve dolayısıyla insan muhakemesini - daha etkili bir şekilde dahil ederek zararları azaltmak için adımlar atın. Eğer zararların faydalardan daha ağır basacağını düşünüyorsanız, sistemi inşa etmeyin.

Bir çevrimiçi eğitim şirketi, öğrenci kompozisyonlarını "okumak" ve otomatik olarak puanlamak için bir yapay zekâ sistemi kullanmak istiyor. Aynı zamanda şirket personelini rastgele kompozisyonları tekrar kontrol etmeye ve yapay zekâ sisteminin zorlandığı kompozisyonları incelemeye yönlendirmeyi planlıyor. Sistem, şirketin öğrencilere puanları hızlı bir şekilde geri göndermesini sağlayacak. Şirket, bir zarar inceleme komitesi oluşturur ve bu komite, sistemin inşa edilmemesini tavsiye eder. Komite tarafından belirlenen başlıca zararlardan bazıları şunlardır: yapay zekâ sisteminin eğitim verilerinden belirli dil kalıplarına karşı yanlılık (önyargı) edinme ve bunu güçlendirme potansiyeli (bu dil kalıplarını kullanan gruplardaki insanlara zarar vererek), öğrencileri kompozisyonlarını geliştirmek yerine algoritmayı "hack'lemeye" (ona göre davranmaya) teşvik etme ve eğitim uzmanlarının sınıftaki rolünü azaltıp teknoloji uzmanlarının rolünü artırma potansiyeli.

## **4. Prototype, starting with non-AI solutions[¶](https://www.kaggle.com/code/var0101/human-centered-design-for-ai" \l "4.-Prototype,-starting-with-non-AI-solutions" \t "_self)**

Yapay zekâ sisteminizin yapay zekâ içermeyen bir prototipini hızlıca geliştirin ve insanların onunla nasıl etkileşimde bulunduğunu görün. Bu, prototiplemeyi daha kolay, daha hızlı ve daha az maliyetli hale getirir. Ayrıca, kullanıcıların sisteminizden ne beklediği ve etkileşimlerini nasıl daha ödüllendirici ve anlamlı hale getirebileceğiniz hakkında erken bilgi edinmenizi sağlar.

Prototipinizin kullanıcı arayüzünü, insanların sisteminizin nasıl çalıştığını kolayca öğrenmelerini, ayarları değiştirmelerini ve geri bildirimde bulunmalarını sağlayacak şekilde tasarlayın.

Geri bildirimde bulunan kişilerin, ırk, cinsiyet, uzmanlık ve diğer özellikler dahil olmak üzere çeşitli geçmişlere sahip olması önemlidir. Ayrıca, neye yardımcı olduklarını ve bunun nasıl gerçekleştiğini anlamalı ve buna rıza göstermelidirler.

Bir film yayın hizmeti başlangıç şirketi, kullanıcıların belirttikleri tercihleri ve izleme geçmişlerine dayanarak onlara film önermek için yapay zekâ kullanmak istiyor. Ekip ilk olarak, çeşitli kullanıcı grubunu, belirttikleri tercihleri ve izleme geçmişlerini bir film meraklısıyla paylaşmaya davet eder. Film meraklısı daha sonra kullanıcıların beğenebileceği filmleri önerir. Bu konuşmalara ve kullanıcıların önerilen hangi filmlerden keyif aldığına dair geri bildirimlere dayanarak, ekip filmlerin nasıl kategorize edildiğine dair yaklaşımını değiştirir. Çeşitli bir kullanıcı grubundan erken aşamada geri bildirim almak ve sık sık yineleme yapmak, ekibin ürününü daha sonra pahalı düzeltmeler yapmak yerine erken aşamada iyileştirmesini sağlar.

## **5. Provide ways for people to challenge the system[¶](https://www.kaggle.com/code/var0101/human-centered-design-for-ai" \l "5.-Provide-ways-for-people-to-challenge-the-system" \t "_self)**

Yapay zekâ sisteminiz kullanıma sunulduğunda, kullanıcılar onun önerilerine itiraz edebilmeli veya kullanımı kolayca devre dışı bırakabilmelidir. İtirazları kabul etmek, izlemek ve ele almak için sistemler ve araçlar oluşturun.

Kullanıcılarla konuşun ve bir kullanıcının bakış açısından düşünün: Sistem önerilerinden merak duyar veya memnuniyetsiz kalırsanız, şunları yaparak itiraz etmek ister miydiniz:

* Öneriye nasıl ulaştığına dair bir açıklama talep ederek?
* Girdiğiniz bilgide değişiklik talep ederek?
* Belirli özellikleri kapatarak?
* Sosyal medya üzerinden ürün ekibine ulaşarak?
* Başka bir eylemde bulunarak?

Bir çevrimiçi video konferans şirketi, video görüşmeleri sırasında arka planı otomatik olarak bulanıklaştırmak için yapay zekâ kullanıyor. Şirket, ürününü farklı etnik kökenlerden çeşitli insan gruplarıyla başarıyla test etti. Ancak, videonun bir kişinin yüzüne düzgün odaklanamadığı durumlar olabileceğinin farkında. Bu nedenle, arka planı bulanıklaştırma özelliğini isteğe bağlı hale getiriyor ve müşterilerin sorunları bildirebilmesi için bir düğme ekliyor. Şirket ayrıca, sosyal medyayı ve diğer çevrimiçi forumları kullanıcı şikayetleri için izlemek üzere bir müşteri hizmetleri ekibi kuruyor.

## **6. Build in safety measures[¶](https://www.kaggle.com/code/var0101/human-centered-design-for-ai" \l "6.-Build-in-safety-measures" \t "_self)**

Güvenlik önlemleri, kullanıcıları zararlara karşı korur. Bir sistemin güvenilir bir şekilde yüksek kaliteli sonuçlar vermesini sağlayarak, istenmeyen davranışları ve kazaları sınırlamayı amaçlar. Bu ancak kapsamlı ve sürekli değerlendirme ve test etme yoluyla elde edilebilir. Yapay zekâ sisteminizin performansını, amaçlanan faydaların sunulmasını, zararların azaltılmasını, adalet ölçümlerini ve insanların sistemi **gerçekte** nasıl kullandığındaki herhangi bir değişikliği sürekli olarak izlemek için süreçler tasarlayın.

Sisteminizin ihtiyaç duyduğu güvenlik önlemlerinin türü, amacına ve neden olabileceği zarar türlerine bağlıdır. İşe, benzer yapay zekâ dışı ürün veya hizmetlere yerleştirilmiş güvenlik önlemlerinin listesini gözden geçirerek başlayın. Ardından, sisteminizde yapay zekâ kullanımının potansiyel zararlarına ilişkin daha önceki analizinizi (3. Adıma bakın) inceleyin.

Yapay zekâ sisteminizin insan gözetimi çok önemlidir:

* Sisteminizi kasıtlı olarak istenmeyen davranışlara yönlendirmeye çalışan bir kişinin rolünü üstlenmesi için insanlardan oluşan bir **'kırmızı ekip'** oluşturun. Ardından, sisteminizi bu tür manipülasyonlara karşı güçlendirin.
* Sistem kullanıma sunulduğunda, kuruluşunuzdaki kişilerin sistemin güvenliğini en iyi nasıl izleyebileceğini belirleyin.
* Yapay zekâ sisteminizin zorlu bir durumla karşılaştığında bir insanı hızlıca nasıl uyarabileceğini araştırın.
* Kullanıcıların ve diğer kişilerin potansiyel güvenlik sorunlarını işaretleyebilmesi için yollar oluşturun.

Yaygın olarak kullanılan yapay zekâ destekli bir sesli asistan geliştiren bir şirket, ürününün güvenliğini artırmak için, sesli asistanı manipüle etmek isteyen kötü niyetli kişilerin rolünü üstlenecek kalıcı bir şirket içi **'kırmızı ekip'** kurar. Kırmızı ekip, sesli asistanı kandırmak için düşmanca girdiler geliştirir. Şirket daha sonra, ürünü benzer düşmanca girdilere karşı korumak için **'düşmanca eğitim'** (adversarial training) kullanır ve böylece güvenliğini artırır.

# Learn more[¶](https://www.kaggle.com/code/var0101/human-centered-design-for-ai" \l "Learn-more" \t "_self)

To dive deeper into the application of HCD to AI, check out these resources:

* Lex Fridman’s [introductory lecture](https://www.youtube.com/watch?v=bmjamLZ3v8A) on Human-Centered Artificial Intelligence
* Google’s People + AI Research (PAIR) [Guidebook](https://pair.withgoogle.com/guidebook/)
* Stanford Human-Centered Artificial Intelligence (HAI) [research](https://hai.stanford.edu/research)