Yapay zekâ yazılı bir açıklamadan nasıl görsel üretir?

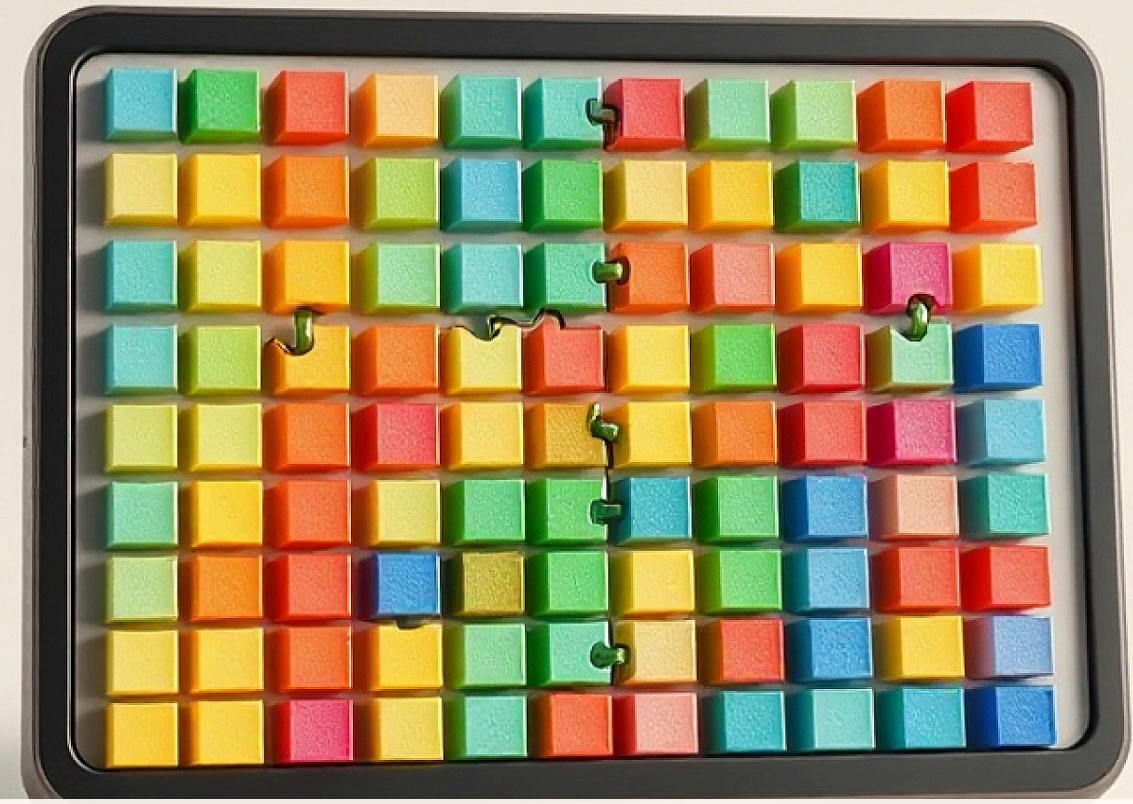


www.yzokul.com.tr

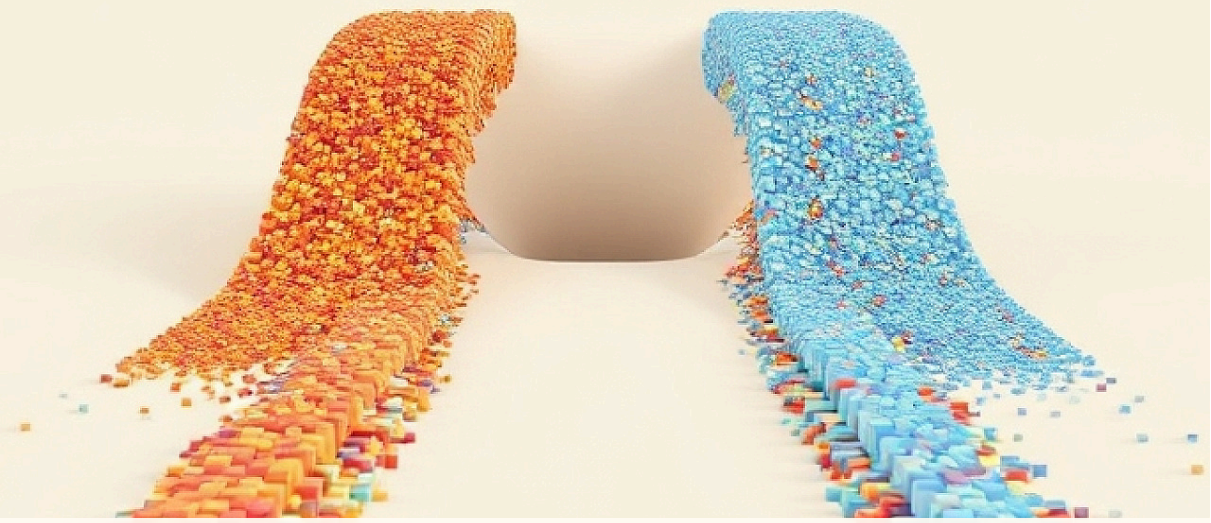
Yapay zekâ yazılı bir açıklamadan nasıl görsel üretir?



Yazılı bir açıklama yazarsın ve yapay zekâ buna uygun bir görsel üretir. Kısa yanıt: model eğitim sırasında metinler ile görseller arasındaki bağlantıları öğrenmiştir ve senin sözcüklerin bu bağlantılar aracılığıyla modeli yönlendirir. Bu kitap her parçayı sıfırdan adım adım açıklar. Sözcüklerin nasıl sayılara dönüştüğünü, eğitimin nasıl bağlantılar kurduğunu ve bir resmin nasıl oluşturulduğunu göreceksin.



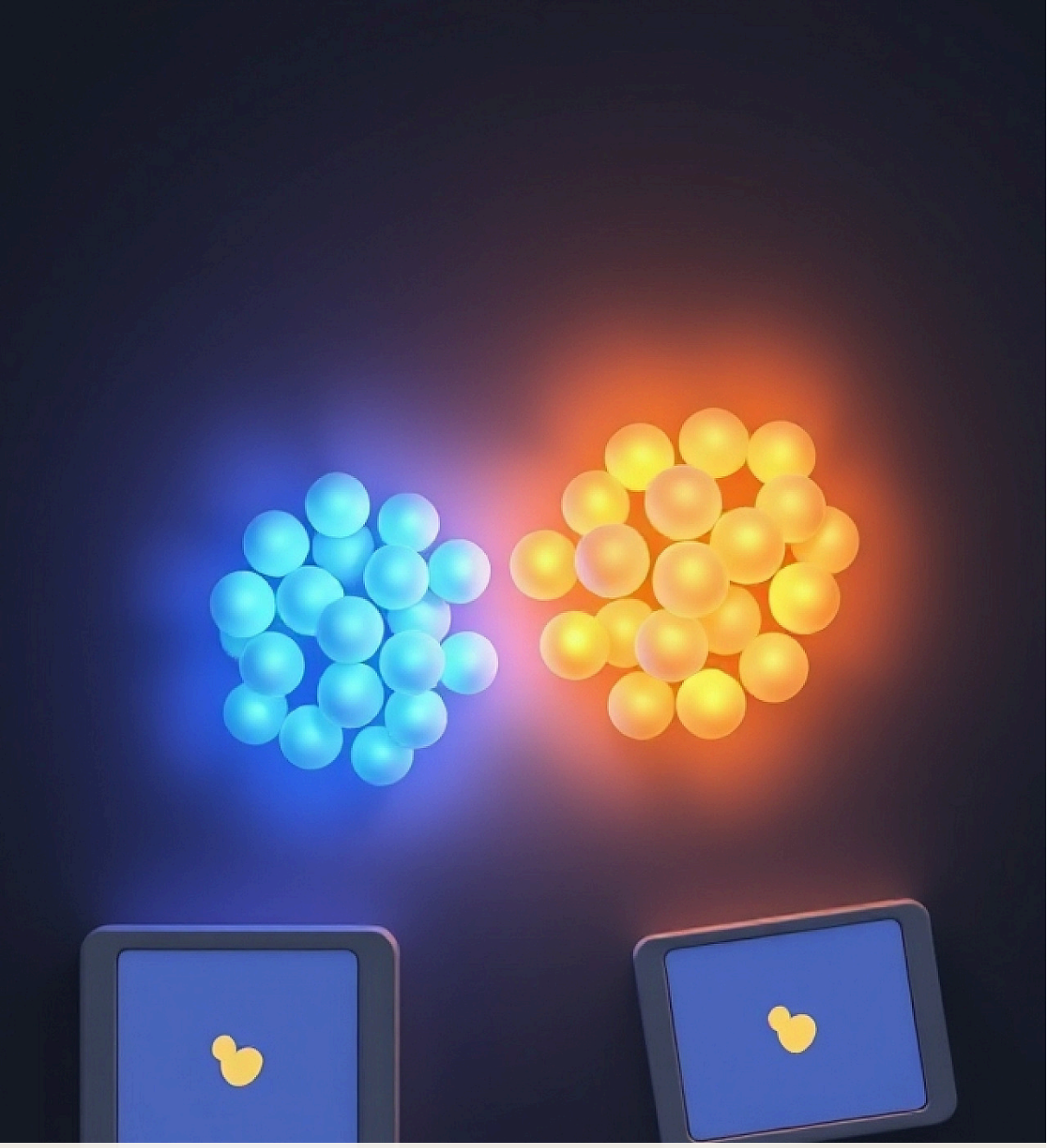
Bir bilgisayar görseli piksel adı verilen küçük renkli noktalardan oluşan bir ızgara olarak görür. Her piksel, rengini ve parlaklığını tanımlayan sayılar olarak saklanır. Yani bütün bir görsel, satırlar ve sütunlar halinde düzenlenmiş büyük bir sayı kümesinden ibarettir. Bir fotoğrafa baktığında bilgisayar sahneyi değil, yalnızca bu sayıları görür.



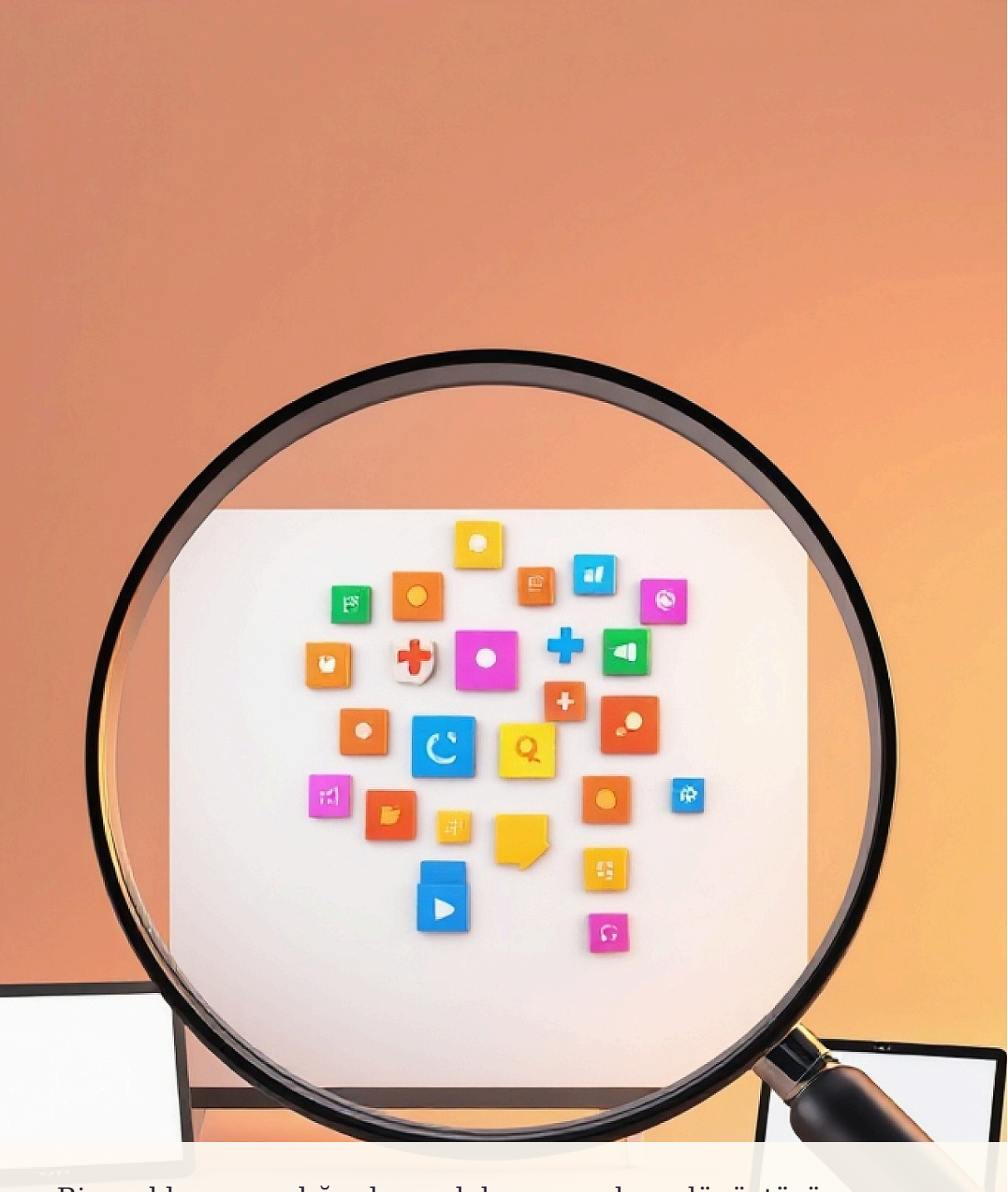
Bir bilgisayar yazılı açıklamanı bir sözcük listesi olarak görür. Her sözcük, bilgisayarın işleyebilmesi için genellikle vektör denen bir sayı listesine dönüştürülür. Aynısı görseller için de olur: onlar da sayılara dönüşür. Artık hem metin hem de görseller sayıdır ve model bunları karşılaştırıp örüntüler bulabilir.



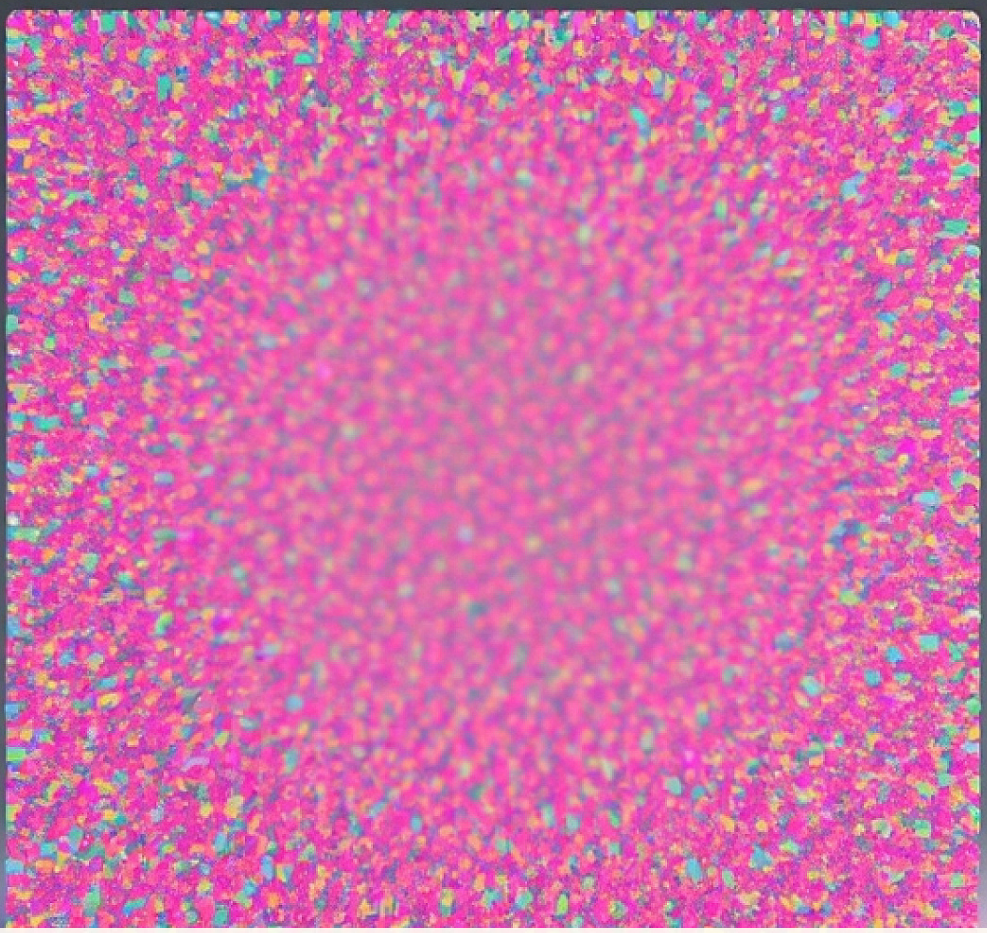
Eğitim sırasında modele milyonlarca çift gösterilir: bir görsel ve o görselin açıklaması. Bu tür pek çok çifti görerek hangi sözcüklerin hangi resimlerle birlikte gitme eğiliminde olduğunu öğrenir. Örneğin köpek veya gün batımı gibi sözcüklerle etiketlenmiş pek çok görsel görür. Zamanla metinler ile görseller arasındaki istatistiksel ilişkileri kavrar.



Model, eşleşen metin ve görsel çiftlerinin temsilinde birbirine yakın konumlanması için iç sayılarını kademeli olarak ayarlar. Sözcükler ile resimler arasındaki ilişkileri bu şekilde saklar. Bir sözcük ile bir görsel sık sık birlikte göründüğünde, sayı örüntüleri birbirine benzer hale gelir. Bu öğrenilmiş bağlantılar, modelin sözcüklerini görsel içerikle nasıl ilişkilendirdiğinin özüdür.



Bir açıklama yazdığında model onu sayılara dönüştürür ve görselin ne içermesi gerektiğine karar vermek için öğrenilmiş bağlantıları kullanır. Açıklaman bir dizi talimat ya da bir kılavuz gibi işlev görür. Örneğin kırmızı araba yazarsan, modelin kırmızı ve araba için iç sayıları görseli bu özelliklere doğru çeker. Sözcükler üretim sürecini yönlendirir.



Bazı modeller ekrandaki karıncalanma gibi rastgele gürültülü bir görselle başlar. Bu gürültü yalnızca rastgele sayılardır. Görev, bu gürültüyü açıklamana uyan bir resme dönüştürmektir. Gürültüden başlamak, modelin var olan bir görseli kopyalamak yerine yeni bir görsel oluşturmasını sağlar. Gürültü ham maddedir.



Model gürültülü görseli pek çok küçük adımda temizler. Her adımda, açıklamanın yönlendirmesiyle biraz gürültü kaldırır. Böylece resim yavaş yavaş ortaya çıkar, giderek daha net ve ayrıntılı hale gelir. Her adım, modelin görselin sözcüklerine uyması için nasıl görünmesi gerektiğini düşünmesine dayanan küçük bir ayarlamadır.



Açıklamayı ay ışığında seyreden bir gemi olarak düşün. Gemi sözcüğü modeli gemi benzeri bir biçim oluşturmaya yönlendirir. Ay ışığı koyu mavi ve gümüş renkleri ve gece ortamını çağrıştırır. Model eğitiminden gelen bu ipuçlarını tutarlı bir sahne oluşturmak için birleştirir. Sözcükler nesnelerin, renklerin ve atmosferin seçimini yönlendirir.



Gemi ve ay ışığı gibi sözcükler, görseli modelin eğitim sırasında öğrendiği belirli biçimlere ve renklere doğru çeker. Model, bu öğrenilmiş çağrışımlara dayanarak en iyi uyanı seçer. Örneğin gemi, tekne gövdeleri ve yelkenlerle ilişkilendirilirken ay ışığı soluk ışık ve karanlık gökyüzüyle ilişkilendirilir. Model tek bir görsel üretmek için tüm sözcük etkilerini dengeler.



Pek çok iyileştirme adımından sonra gürültü net bir resme dönüşür. Sonuç, modelin açıklamana uyduğuna karar verdiği tek bir görseldir. Model gürültünün çoğunu kaldırdığında ve görsel makul bir eşleşme gibi göründüğünde durur. Sözcüklerini izlemede modelin en iyi çabası olan son görseli görürsün.



Üretilen görselin fiziksel veya tarihsel olarak doğru olacağı garanti edilmez. Gerçek dünya gerçekleriyle değil, eğitimden gelen örüntülerle eşleşir. Örneğin imkânsız arma donanımına sahip bir gemi ya da yanlış yönden gelen ay ışığı gösterebilir. Model gerçekliği değil, eğitim verilerine dayalı makullüğü hedefler. Önemli ayrıntıları her zaman kontrol et.



Yaygın bir hata, modelin var olan bir fotoğrafı aradığını düşünmektir. Hiçbir şeyi araştırmaz; öğrenilmiş örüntülerden yeni bir görsel kurar. İçinden görsel çektiği bir veritabanı yoktur. Bunun yerine, açıklaman ve eğitimi tarafından yönlendirilerek pikselleri adım adım üretir. Sonuç kopya değil, yeni bir şeydir.



Bir başka hata, modelin sözcükleri insanlar gibi gerçekten anladığına inanmaktır. Gerçek bir kavrayışla değil, pek çok örnekten öğrenilmiş sayı örüntüleriyle çalışır. Geminin insani anlamda ne olduğunu bilmez; yalnızca eğitim verilerinde gemi sözcüğüne hangi görsel özelliklerin eşlik etme eğiliminde olduğunu bilir. Anlayışı kavramsal değil, istatistikseldir.



Aynı açıklama her zaman aynı görseli vermez. Bazı modeller rastgele gürültüden başladığı için sonuçlar her seferinde farklı olabilir. Başlangıç gürültüsündeki küçük değişiklikler bile farklı son görsellere yol açar. Yani ay ışığında bir gemiyi iki kez üretirsen iki farklı resim elde edebilirsin. Bu değişkenlik normaldir ve beklenir.



Eğitim metin-görsel bağlantılarını öğretir, açıklaman süreci yönlendirir ve bazı modeller gürültüyü adım adım iyileştirir. Sonucun öğrenilmiş örüntüleri izlediğini, doğruluğun garanti edilmediğini unutma. Model araştırmaz ya da gerçekten anlamaz; istatistiksel çağrışımlara dayanarak üretir. Bu bilgiyle yapay zekâ tarafından üretilen görselleri daha iyi yorumlayabilir ve kullanabilirsin.