Yapay zekâ konuşmayı nasıl yazıya dönüştürür?



www.yzokul.com.tr

Yapay zekâ konuşmayı nasıl yazıya dönüştürür?



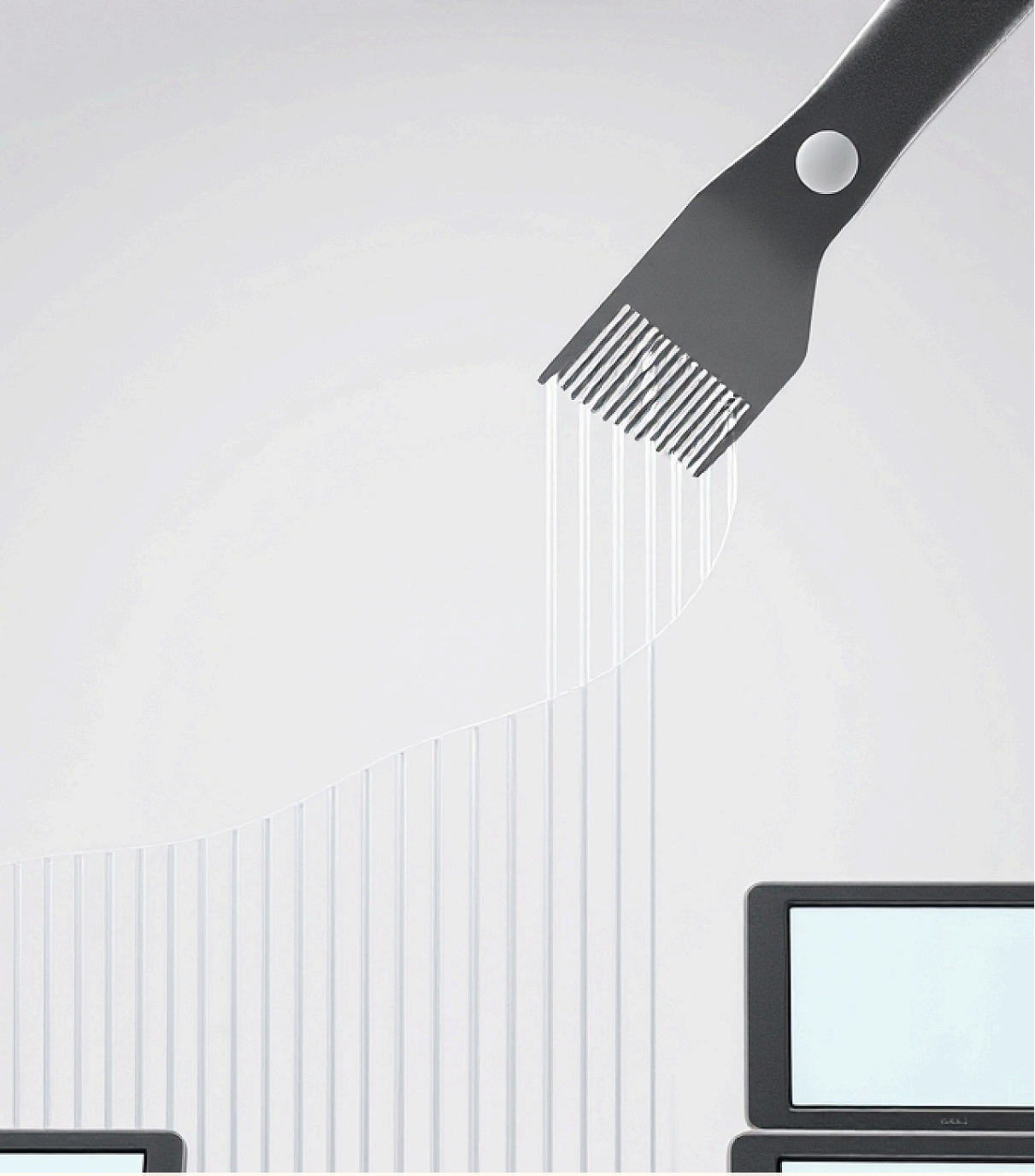
Yapay zekâ, sesi bir mikrofonla yakalayarak, bu sesi sayılara dönüştürerek ve ardından bu sayılardan sesleri ve sözcükleri tahmin ederek konuşmayı yazıya dönüştürür. Sistem bir insan gibi dinlemez; sayılardaki örüntüler üzerinde çalışır. Bu kitap, ilk ses dalgasından son yazılı metne kadar bu yolun her parçasını adım adım açıklar. Özniteliklerin, sözcük tahmininin ve cümle bağlamının nasıl birlikte çalıştığını görürsün.



Konuşma, ses dalgası olarak yayılan bir hava titreşimidir. Konuştuğunda sesin ağzının çevresindeki havayı titreştirir ve bu titreşimler dışa doğru yayılır. Bir mikrofon bu dalgayı yakalar ve elektrik sinyaline dönüştürür. Bu sinyal saklanabilir ve daha sonra bir bilgisayar tarafından kullanılabilir. Sesin bir dalga olduğunu anlamak, mikrofonun aslında neyi kaydettiğini görmene yardımcı olur.



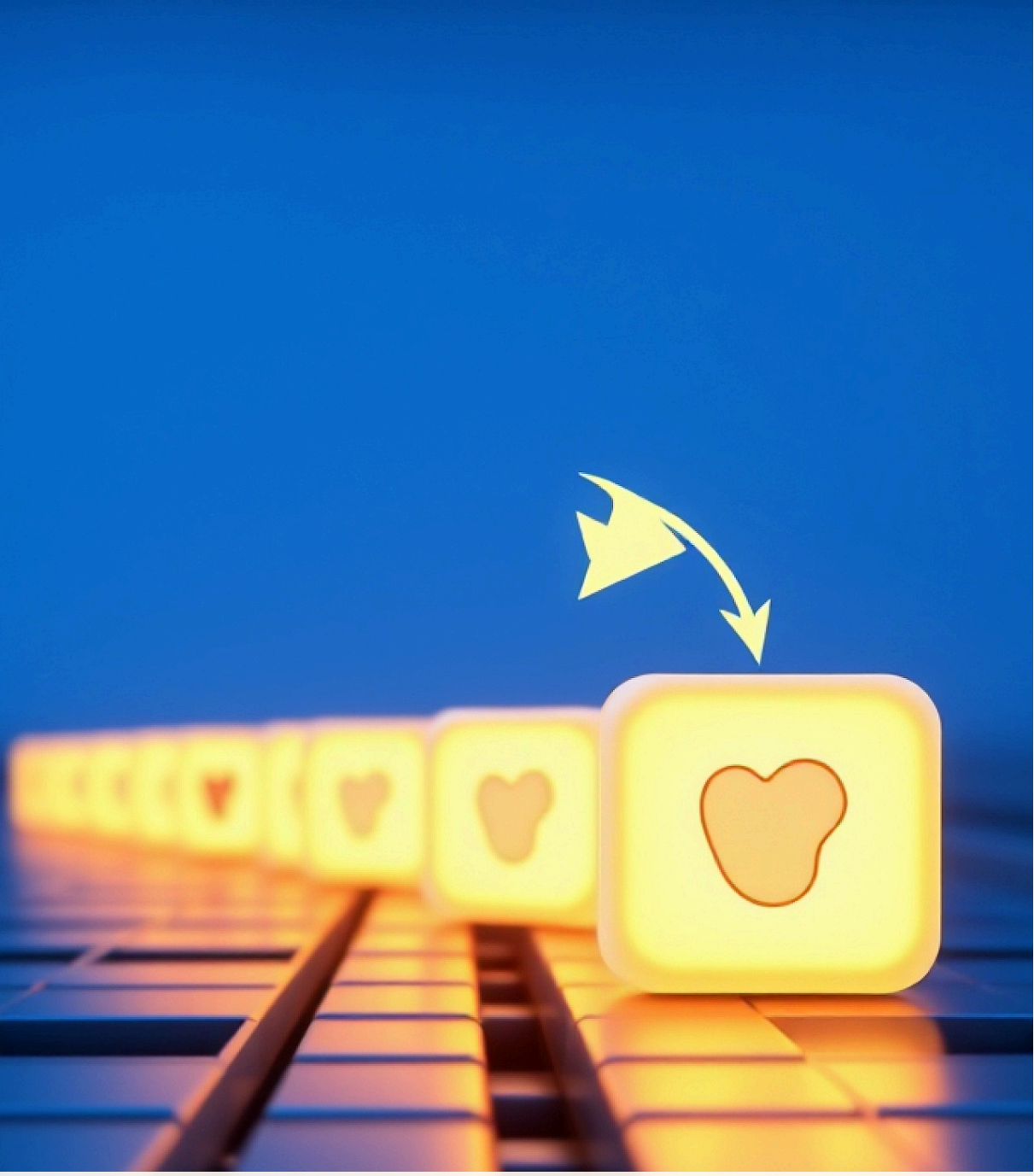
Bir mikrofon konuşan bir kişiyi kaydeder ve sesi dijital bir dosya olarak saklar. Mikrofon hava titreşimlerini elektrik sinyaline dönüştürür ve bilgisayar bu sinyali bir sayı dizisi olarak saklar. Bu kayıt, sistemin başladığı ham maddedir. Net bir kayıt olmadan sonraki adımların üzerinde çalışacağı malzeme azalır. Dijital dosya, analiz edilmeye hazır biçimde sesin tamamını tutar.



Bilgisayar ses dalgasını doğrudan kullanamaz, bu yüzden dalgayı saniyede birçok kez ölçer. Her ölçüm bir sayıya dönüşür ve konuşma uzun bir sayı listesi hâline gelir. Bu sürece örnekleme denir. Sayı listesi, hava basıncının zaman içinde nasıl değiştiğini temsil eder. Bu noktadan itibaren sistem yalnızca sayılarla çalışır, özgün dalgayla değil.



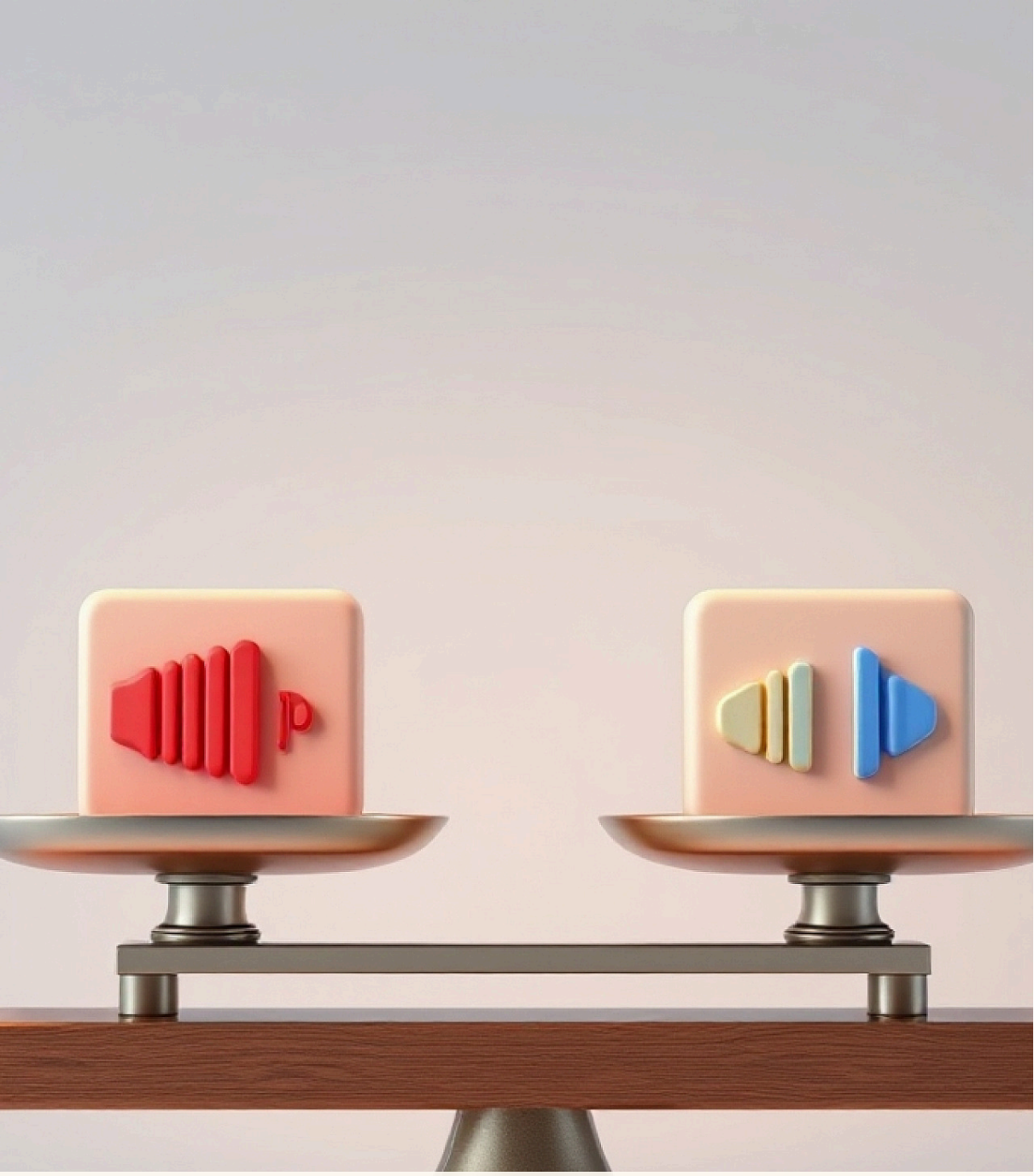
Sistem her ham sayıyı tutmaz, bunun yerine öznitelik adı verilen yararlı örüntüler çıkarır. Öznitelikler, sesin önemli niteliklerini, örneğin enerjinin farklı frekanslara nasıl dağıldığını, sıkışık biçimde tanımlar. Frekans, havanın ne kadar hızlı titreştiği anlamına gelir. Özniteliklere odaklanarak sistem, devasa bir sayı listesini konuşmayı tanımak için önemli olanı yakalayan daha küçük bir kümeye indirir.



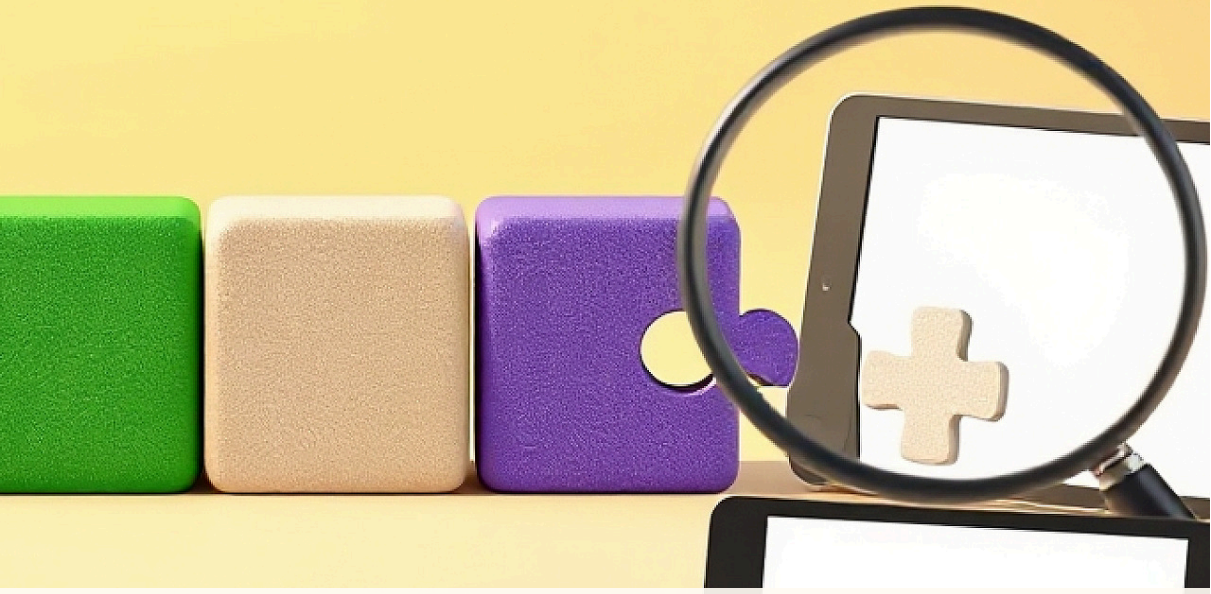
Sistem, özniteliklerden hangi küçük seslerin söylendiğini tahmin eder. Bu küçük ses birimleri, konuşulan sözcüklerin yapı taşlarıdır. Sistem öznitelikleri, birçok konuşma örneğinden öğrendiği örüntülerle karşılaştırır. Kayıttaki her an için en olası sesi tahmin eder. Bu tahmin edilen sesler henüz sözcük değildir, ama onlara doğru ilk adımdır.



Tahmin edilen sesler olası sözcüklerde birleştirilir. Sistem, saptadığı ses dizisine en iyi uyan sözcükleri listeler. Seslerin dilde tipik olarak sözcükleri nasıl oluşturduğuna dair bir model kullanır. Örneğin bir ses dizisi birkaç sözcüğe uyabilir, bu yüzden sistem adayları tutar. Bu sözcük tahmini, bir ses akışını bir dizi sözcük seçeneğine dönüştürür.



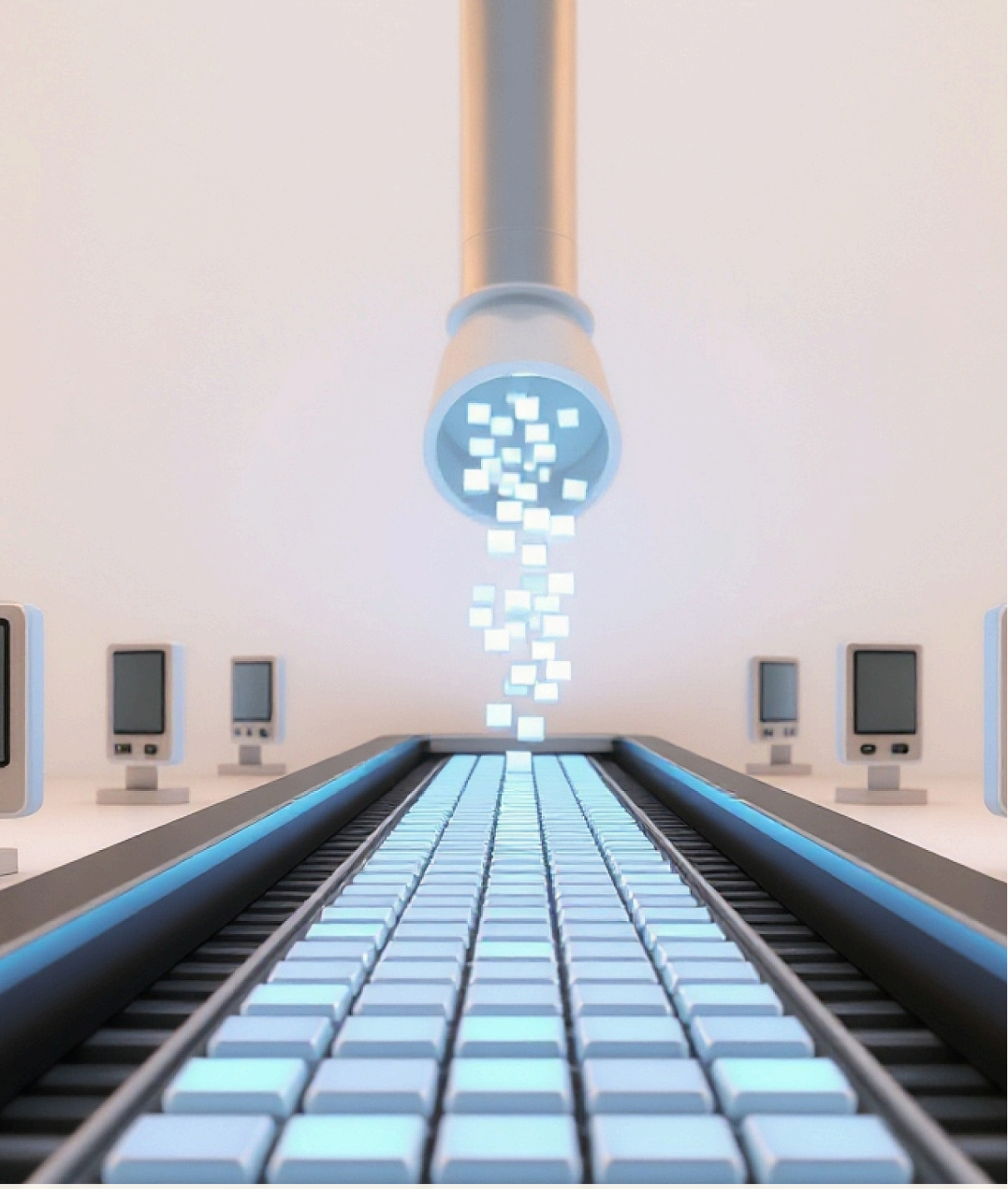
Benzer sesler birkaç farklı sözcüğe işaret edebilir. Sistem birden fazla aday tutar ve hangilerinin daha olası olduğunu tartar. Örneğin iki ve diğer anlamındaki iki sözcüğünün sesleri aynıdır, bu yüzden sistem başka ipuçlarından karar vermelidir. Hemen tek bir sözcüğe bağlanmaz. Bunun yerine birkaç seçeneği tutar ve cümleden daha fazla kanıt bekler.



Bir cümledeki çevre sözcükler doğru sözcüğü seçmeye yardımcı olur. Örneğin bağlam, kulağa neredeyse aynı gelen iki sözcüğü, örneğin yaz ve yaz sözcüklerini ayırt edebilir. Önceki sözcükler kâğıda sözcük koymakla ilgiliyse sistem yazmayı tercih eder. Cümle bağlamı, seçenekleri daraltan bir kılavuz gibi işler. Sistem karar vermek için yalnızca sesten fazlasını böyle kullanır.



Gürültüde heceler kaybolsa bile sistem en olası cümleyi tamamlamaya çalışır. Boşluğun çevresindeki sözcüklerle en iyi uyanı kullanır. Örneğin bir sözcüğün bir kısmı öksürükle örtülürse sistem yine de cümlenin geri kalanından amaçlanan sözcüğü tahmin edebilir. Bu tamamlama, cümlelerin genellikle nasıl ilerlediğine dair öğrenilmiş örüntülere dayanır. Yazılı metnin okunabilir kalmasına yardımcı olur.



Sözcükler seçildikten sonra sistem bunları metin olarak yazar. Son çıktı, söylenenin yazılı sürümüdür. Metin, sistem karar verdikçe sözcük sözcük sırayla görünür. Ekranda gördüğün şey bu yazılı metindir. Ses dalgasından metne kadar tüm süreç, konuşmanın o parçası için artık tamamlanmıştır.



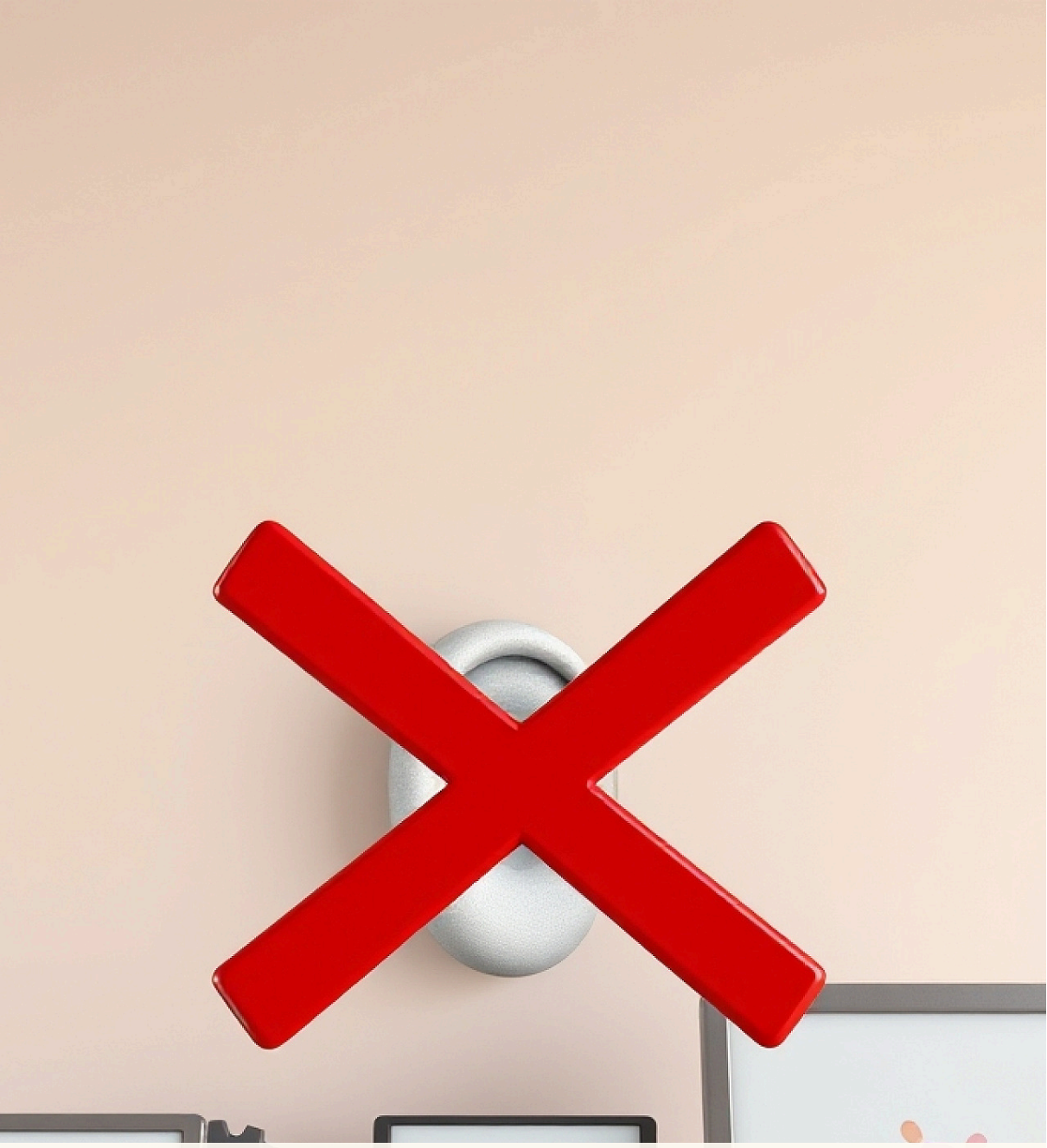
Farklı aksanlar ve ses kaliteleri tanımayı zorlaştırabilir. Sistemin aynı sözcükleri birçok söyleyiş biçimini ele alması gerekebilir. Bir aksan seslerin nasıl üretildiğini değiştirir ve ses kalitesi perde ve pürüzlülük gibi şeyleri içerir. Sistem birçok örnekten öğrenir, ama alışılmadık aksanlar yine de hatalara yol açabilir. Doğruluk, sistemin benzer sesler üzerinde ne kadar iyi eğitildiğine bağlıdır.



Birkaç kişi aynı anda konuştuğunda sesler karışır ve doğruluk düşebilir. Üst üste binen sesler sistem için zor bir durumdur çünkü farklı konuşmacılardan gelen öznitelikler birleşir. Sistem bir kişinin sözcüklerini diğerinden ayırmakta zorlanabilir. Bu tür durumlarda yazılı metin hatalar veya eksik sözcükler içerebilir. Bu, net ve tek konuşmacılı konuşmanın neden en iyi çalıştığının bir nedenidir.



Arka plan gürültüsü konuşmanın bazı kısımlarını gizleyebilir. Sistem, net duyamadığı kısmı doldurmak için bağlama güvenir. Örneğin geçen bir kamyon bir heceyi bastırırsa sistem yine de sözcüğü cümleden tahmin edebilir. Gürültü öznitelikleri daha az net yapar, bu yüzden sistem daha çok çalışmak zorundadır. Bağlam ve sözcük tahmini yardımcı olur, ama bazı hatalar kalabilir.



Yaygın bir yanlış anlama, bilgisayarın gerçekten bir insan gibi duyduğunu düşünmektir. Aslında sayılar ve örüntüler üzerinde çalışır ve hata yapabilir. Anlamı senin anladığın gibi anlamaz; özniteliklerden ve bağlamdan olası sözcükleri tahmin eder. Bunu bilmek çıktısını dikkatle yorumlamana yardımcı olur. Önemli metinleri her zaman kontrol et, çünkü sistem kusursuz değildir.



Ses dalgaları sayılara, sonra özniteliklere, sonra seslere, sonra sözcüklere dönüşür ve bağlam seçim yapmaya yardımcı olur. Aksanların, ses kalitesinin, gürültünün ve üst üste konuşanların doğruluğu etkilediğini unutma. Sistem gerçekten duymak yerine tahmin eder, bu yüzden hatalar olabilir. Bu yol, yapay zekânın konuşmayı nasıl yazıya dönüştürdüğünü açıklar. Konuşma tanımayı kullanırken bu noktaları aklında tut.