Yapay Zekâ ve Eniyilemede Açıklama Yöntemleri

Açıklanabilir Yapay Zekâ (AYZ) yöntemleri karmaşık makine öğrenimi modellerini şeffaflaştırmada önemli ilerlemeler kaydetmiş olsa da, kritik kararları yönlendiren matematiksel eniyileme problemleri çoğunlukla anlaşılmaz kara kutular olarak kalmaktadır. Bu konuşma, AYZ için geliştirilen tekniklerin eniyileme modellerine nasıl uyarlanabileceğini göstererek, son zamanlarda araştırmacıların ilgisini çeken açıklanabilir eniyileme alanını YAEM 2025 katılımcılarına tanıtmayı amaçlamaktadır.

Konuşmada birbirini tamamlayan iki yeni yaklaşım ele alınacaktır: "Matematiksel Eniyileme için Tutarlı Yerel Açıklamalar" ve "Doğrusal Programlama için Karşıt Olgusal Açıklamalar." Birinci yaklaşım, model parametrelerinin hem amaç fonksiyonunu hem de karar değişkenlerini nasıl etkilediğini paydaşların anlamasını sağlayarak, matematiksel yapıları koruyan eniyileme modelleri için açıklamalar sunmaktadır. İkinci yaklaşım ise "İstediğim sonuca ulaşmak için hangi minimum parametre değişiklikleri gerekir?" gibi paydaş sorularını yanıtlayan üç tür karşıt olgusal açıklama getirmektedir: zayıf, güçlü ve göreli.

Örnekler üzerinden, bu yöntemlerin şeffaflığı nasıl artırabileceği, paydaşlar arasındaki müzakereleri nasıl iyileştirebileceği ve karar vericilere ne tür içgörüler sağlayabileceği gösterilecektir. Konunun yeniliği ve henüz yanıtlanmamış birçok soru bulunması nedeniyle, konuşmanın son bölümü gelecekteki potansiyel araştırma alanlarına ayrılacaktır.