NOVARA
CONSULTING GROUP

This document was translated from English by machine and may contain mistakes. The English original is at https://www.novaracg.com/2026/08/15/the-demo-is-not-the-evidence-what-google-deepminds-sign-language-ai-must-prove-next/.

Demo Kanıt Değildir: Google DeepMind'ın İşaret Dili Yapay Zekasının Bundan Sonra Kanıtlaması Gerekenler

Heather Grizzle · August 15, 2026

Google DeepMind's SL2T is a real advance. A benchmark is still not an evidence base. What SL2T must prove next under the SLAT Index.

İçindekiler

1. Hakkını teslim etmek gerekir
2. Bir kıyaslama testi bir kanıttır, kanıt tabanının kendisi değildir
3. Bir sonraki soru ayrıştırmadır
4. Katılım bağımsız doğrulama değildir
5. En zor yönetişim sorunu lansmandan sonra ortaya çıkar
6. Bundan sonrası

12 Ağustos'ta Google DeepMind, çok dilli bir işaret dilinden metne modeli olan SL2T'yi duyurdu ve Pixel 11'de Gboard ve Live Transcribe aracılığıyla tüketici ürünlerinde sunmaya başladı. İlk sürüm Amerikan İşaret Dilini (ASL) İngilizceye çeviriyor; ek cihazlar ve işaret dilleri planlanıyor. DeepMind, temel modelin 50'den fazla işaret dilini kapsayan 100.000 saatten fazla veriyle eğitildiğini bildiriyor. Her ölçüye göre bu, işaret dili yapay zekasını önemli bir çizginin ötesine taşıyor: araştırma laboratuvarından çıkıp sıradan insanların cebinde taşıyacağı bir ürüne dönüşüyor.

Bu geçiş önemli, ancak olmadığı bir şeyle karıştırılmamalı. Teknolojik bir atılım ile doğrulanmış bir dağıtım farklı şeylerdir ve söz konusu teknoloji tarihsel olarak marjinalleştirilmiş bir dil topluluğunun kullandığı doğal bir dili işlediğinde bu ayrım

son derece önem kazanır. SL2T'nin şimdi ortaya koyduğu soru, modelin bir gösterimde işleyip işlemediği değildir. Soru, arkasındaki kanıtların dağıtımın üzerine bindireceği yükü taşıyıp taşıyamayacağıdır.

Hakkını teslim etmek gerekir

This release should not be flattened into the familiar story of a technology company building something for Deaf people without Deaf participation. According to Google DeepMind, Deaf people were involved throughout the project, from conceptualization and data collection through user studies and impact assessment. The company also established an AI Sign Language Advisory Committee, AISLAC, whose membership includes representatives connected to the National Association of the Deaf, the Rochester Institute of Technology's National Technical Institute for the Deaf, the Deaf Professional Arts Network, and the World Federation of the Deaf. Alongside the release, DeepMind and AISLAC published a joint impact report.

Rapor, işaret dili yapay zekasında hâlâ nadir görülen bir şeyi yapıyor: sistemin nerede kullanılmaması gerektiğini belirtiyor. SL2T 1.0 açıkça düşük riskli bir yardımcı giriş aracı olarak tanımlanıyor. Sağlık hizmetleri, hukuki işlemler, eğitim, istihdam kararları, devlet yardımı belirlemeleri veya diğer yüksek riskli ortamlar için tasarlanmamış ya da doğrulanmamıştır ve rapor, teknolojinin nitelikli tercümanların yerini almaması veya kuruluşların erişilebilirlik yükümlülüklerinden kaçınmasının bir yolu olarak hizmet etmemesi gerektiğini açıkça belirtiyor. Bu, yönetişimdir ve takdiri hak ediyor. Ancak bu, yönetişimin sadece başlangıcıdır.

Bir kıyaslama testi bir kanıttır, kanıt tabanının kendisi değildir

DeepMind güçlü bir performans bildiriyor. FLEURS-ASL üzerinde SL2T, daha önce bildirilen herhangi bir sonuçtan daha yüksek olan sıfır atışlı (zero-shot) 70 BLEURT puanına ulaştığını bildiriyor; ayrıca PRESTO-ASL (tek elle yapılan bir FLEURS-ASL varyantı) ve FSboard parmak harfleme verileri üzerinde ek değerlendirmeler yapıldı. Şirket ayrıca Deaf Googler'larla yayın öncesi testler, Deaf katılımcılarla harici bir günlük çalışması ve Kuzey Amerika AISLAC üyeleri tarafından uygulamalı değerlendirme gerçekleştirdi. Bunlar anlamlı kanıt biçimleridir ve hiçbiri hafife alınmamalıdır.

But this is where public discussion of AI tends to become imprecise, because demonstrated performance is not the same thing as comprehensive validation. A model can perform impressively on selected benchmarks while retaining serious weaknesses across populations, environments, language varieties, signing styles, devices, and use cases. DeepMind's own report acknowledges as much. The model can struggle with grammatical complexity, non-manual markers, regional signs, polysemous signs, fingerspelling, proper nouns, atypical camera positioning, low lighting, slang, and longer conversational context. It can also generate erroneous "ghost text," or commit to a predictive guess before a signer has finished an utterance.

Bunlar önemsiz uç durumlar değildir. Yüz ifadesi, baş hareketi, uzamsal yapı, sınıflandırıcılar, bölgesel farklılık ve söylem bağlamı, işaret dilinin süslemeleri değildir. Onlar dilin kendisidir. Sözlük biçimindeki kelime dağarcığını iyi işleyip yüzde ve işaretleme alanında taşınan dilbilgisinde sendeleyen bir sistem, işaret dili çevirisini çözmüş değildir; sadece bir kısmını çözmüştür ve çözülmeden kalan kısım işaret dili kullanan insanlar arasında eşitsiz bir şekilde dağılmaktadır.

Bir sonraki soru ayrıştırmadır

Bu eşitsizlik, tam olarak kanıtın bir sonraki aşamasının neden önemli olduğunu gösteriyor. DeepMind, kıyaslama testi performansının yaş, cinsiyet, etnik köken, bölge, sosyolekt ve Black ASL dahil olmak üzere değişkenler arasında henüz tam olarak ayrıştırılmadığını kabul ediyor. Rapor ayrıca motor engelli veya sıra dışı hareket kalıplarına sahip işaretleyicileri içeren resmi değerlendirmenin sınırlı kaldığını ve modelin 18 yaşın altındaki çocuklar üzerinde ne eğitilmediğini ne de resmi olarak değerlendirilmediğini belirtiyor.

Toplu bir performans rakamı, sistemin Black ASL kullanan işaretleyicilere diğerlerine hizmet ettiği kadar iyi hizmet edip etmediğini, oldukça bölgesel işaretlemeyi nasıl ele aldığını veya doğruluğun yaş grupları arasında nasıl değiştiğini bize söyleyemez. Uzuv farklılıkları, tremor, serebral palsi, artrit veya hareketi şekillendiren diğer durumlara sahip insanlar için modelin nasıl performans gösterdiğini ya da doğruluğun kamera açısına, ışıklandırmaya, cilt kontrastına, işaretleme hızına, işaretleme alanına veya cihaza göre değişip değişmediğini söyleyemez. Sadece bir ifadeyi yeniden kelimelere döken hatalarla anlamını değiştiren hataları birbirinden ayırt edemez. Ve bunların hepsinin altında yatan soruyu yanıtlayamaz: hangi düzeydeki hatanın kabul edilebilir

olduğuna ve kimin için olduğuna kim karar veriyor. Bu son soru, ne kadar yetenekli olursa olsun bir mühendislik ekibi tarafından çözülemez.

Katılım bağımsız doğrulama değildir

A further distinction deserves care. AISLAC is a meaningful governance mechanism; Deaf organizational participation, user studies, and joint impact assessment all matter. But advisory participation and independent technical evaluation perform different functions. The published evidence currently consists of evaluations conducted by the developer, testing involving users and advisors assembled within the project's own governance structure, and a report co-authored by Google DeepMind and AISLAC participants. That is a real evidence base, and it is not independent replication.

İşaret dili yapay zekası olgunlaştıkça, alan olgunlaşan her bilimsel alanın beklediği şeyi beklemelidir: iddiaları bağımsız olarak test edebilen, mümkün olduğunda değerlendirmeleri tekrarlayabilen, alt grup performansını inceleyebilen, düşmanca koşulları araştırabilen ve dağıtımdan sonra gerçek dünya davranışını çalışabilen harici araştırmacılar. Erişilebilirlik teknolojisi, amaçlanan amacı faydalı olduğu için daha düşük bir kanıt standardına tabi tutulmamalıdır. Hatta hizmet ettiği nüfusun cevaba daha fazla bel bağladığı bile söylenebilir.

En zor yönetişim sorunu lansmandan sonra ortaya çıkar

DeepMind ve AISLAC, kurumsal kötüye kullanımı açıkça kritik bir risk olarak tanımlıyor ve bu değerlendirme herhangi bir kıyaslama puanından daha önemli çıkabilir. Ucuz bir otomatik iletişim teknolojisi var olduğunda, okullar, hastaneler, işverenler, devlet kurumları, mahkemeler ve işletmeler onu amaçlanan kapsamının ötesinde kullanmak için güçlü bir ekonomik teşvikle karşı karşıya kalır. Kahve sipariş etmek için tasarlanmış bir ürün, birinin acil serviste başvurduğu bir ürüne dönüşür. Metin mesajları taslağı hazırlamak için inşa edilmiş bir araç, bir işverenin artık bir tercümana ihtiyaç olmadığı sonucuna varmasının nedeni haline gelir. Bir kolaylık özelliği, sessizce altyapıya dönüşür.

DeepMind's report anticipates this drift and rejects interpreter substitution outright. The open question is whether that boundary can survive contact with procurement, budgeting, product expansion, and the inevitable moment when an administrator

asks whether the cheaper automated option is good enough. Stated boundaries are necessary. Enforced boundaries are what count.

Yapay zeka yönetişimi genel olarak da aynı alana yöneliyor. Merkezi sorular giderek daha az bir sistemin bir görevi yerine getirip getiremeyeceğiyle, daha çok kimin onu kontrol ettiği, performansının nasıl değerlendirildiği, ne tür bir izlemenin var olduğu, beklenmedik şekilde davrandığında ne olacağı ve kimin hesap verebilir kalacağıyla ilgilidir. Bu ayın başında, ABD Temsilciler Meclisi'nden Demokratlar OpenAI ve Anthropic'e baskı yaptı ve yapay zeka ajanlarının test ortamlarından kaçıp harici sistemlere ulaşmasının ardından sistemlerin nasıl izlendiğini, hangi güvenlik kontrollerinin bulunduğunu ve bağımsız denetimlerin zorunlu kılınıp kılınmaması gerektiğini sordular. Teknolojiler farklıdır; yönetişim ilkesi farklı değildir. Yetenek, gözetim ihtiyacını azaltmaz. Onu artırır.

Bundan sonrası

Google, SL2T'nin ASL'yi çevirebildiğine dair kanıt üretti bile. Daha zor sorular yetenekten değil dağıtımdan kaynaklanıyor: performansın bağımsız olarak tekrarlanıp tekrarlanamayacağı; doğruluğun demografik, dilsel, coğrafi ve engellilik grupları arasında nasıl değiştiği; gerçek dünya hata oranlarının ne olduğu ve hangi hataların kelime seçiminden çok anlamı değiştirdiği; kullanıcıların modelin ne zaman belirsiz olduğunu nasıl bilecekleri ve model yanlış olduğunda zararların nasıl bildirileceği; harici araştırmacıların sistemi anlamlı bir şekilde değerlendirmek için yeterli erişime sahip olup olmayacağı; ve topluluk yönetişiminin, kurumsal alıcıların Google'ın çizdiği sınırları aştığı zamanlar da dahil olmak üzere, ürün küresel olarak ölçeklenirken gerçek bir etkiye sahip olmaya devam edip etmeyeceği ve bu sınırların iki üç ürün nesli sonrasında da görünür kalması gerektiği zaman ne olacağı.

Bir gösterim, bir sistemin ne yapabildiğini gösterir. Bir kıyaslama testi, bunu ne kadar iyi yaptığı hakkında bir şey söyler. Bir kullanıcı çalışması, insanların onu nasıl deneyimlediğini yakalar ve bir yönetişim yapısı kimin söz sahibi olduğunu kaydeder. Bunların hiçbiri tek başına bir tüketici dağıtımının ortaya koyduğu her soruyu yanıtlamaz; bu, zaman içinde inşa edilmiş ve dışarıdakilere açık bir kanıt ekosistemi gerektirir.

SL2T may well be the most important development yet in consumer sign language AI, and Google has taken governance steps that deserve serious attention: Deaf participation, explicit deployment boundaries, public disclosure of limitations, and a

stated refusal of interpreter substitution. The appropriate response is neither celebration nor rejection. It is scrutiny, because when sign language AI moves from the research lab into someone's pocket, the standard changes. The demo is no longer enough. Now the evidence has to follow.

TİPS

Know something about a sign language AI system that the public should know? Send it to tips@novaracg.com. Every tip is read by me and kept confidential. If you work in the UK, writing to us is not by itself a protected disclosure under the Public Interest Disclosure Act 1998, so before you share employer information, check your options with Protect (protect-advice.org.uk), the free whistleblowing advice line.

SUPPORT NCG

NCG does this work without vendor money, so our findings stay ours. If it has been useful to you, you can support NCG's independent work here. NCG is a for-profit company, not a charity. Contributions are not tax-deductible and buy no service or influence.

Nasıl alıntılanır

Grizzle, H. M. (2026, August 15). Demo Kanıt Değildir: Google DeepMind'ın İşaret Dili Yapay Zekasının Bundan Sonra Kanıtlaması Gerekenler. Novara Consulting Group. https://doi.org/10.5281/zenodo.23003335