

# **BÜYÜK DİL MODELLERİNDE ÖLÇEK MESELESİ VE ZUHUR ETME**

**NECMİ GÜRSAKAL  
(MUDANYA ÜNİVERSİTESİ)**

# Nedir ortaya çıkma, belirme veya zuhur etme?

“Kaostan doğan düzendir. Daha incelikli bir tanım ise, bir sistemin çeşitli varlıkları arasındaki etkileşimler yoluyla yeni, tutarlı yapıların bir araya geldiği kaostan doğan yüksek düzeyli karmaşıklıktır. Zuhur etme, bu etkileşimler bozulduğunda, sistemin farklılaşmasına ve nihayetinde yeni bir şeye dönüşmesine neden olduğunda meydana gelir”.

<https://peggyholman.com/papers/engaging-emergence/engaging-emergence-table-of-contents/part-i-the-nature-of-emergence/chapter-1-what-is-emergence/>



Ekonomiden iki örnek ile ölçek meselesini anlatmaya başlayalım. Marx, niceliksel farklılıkların niteliksel farklılıklara dönüştüğünü söylemişti, ancak 1920'lerde Paris'te iki edebiyatçının tartışması da konuyu çok güzel bir şekilde özetliyor:

“Scott Fitzgerald: Onlar bizden farklıydı.  
Hemingway: Evet, daha fazla paraları var”.

Birilerinin paraları artmış ve sonunda bu niceliksel fark öyle bir noktaya varmış ki, onlara “zengin” denilmeye başlanmış.

Beatles sahnede o kadar uzun süre çalmıştı ki, çaldıklarının niteliği değişmişti.



Başka bir örnek verelim, “Beatles, 1960 ile 1962'nin sonu arasında beş kez Hamburg'a seyahat etti. İlk yolculukta, gecede beş veya daha fazla saat olmak üzere 106 gece, ikinci yolculuklarında ise 92 kez çaldılar. Üçüncü yolculuklarında 48 kez sahneye çıkarak toplam 172 saat sahnede kaldılar. Kasım ve Aralık 1962'deki son iki Hamburg seyahatinde, 90 saatlik bir performans daha vardı. Sonuç olarak, bir buçuk yıldan biraz fazla bir süre içinde 270 gece performans sergilediler. 1964'te ilk başarı patlamaları sırasında, tahminen 1.200 kez canlı olarak çaldılar ki bu olağanüstü bir rakam. Bugün çoğu grup, tüm kariyerleri boyunca 1.200 kez çalmıyor. Beatles'ı farklı kılan Hamburg potasıdır”

Malcolm Gladwell on Joy, Gates and the Beatles -- and on stage, The Guardian, Nov 16, 2008,  
[https://www.theguardian.com/technology/blog/2008/nov/16/gladwell-outliers.](https://www.theguardian.com/technology/blog/2008/nov/16/gladwell-outliers)

“Biyologlar, fizikçiler, ekolojistler ve diğler bilim adamları, ‘zuhur eden’ (emerge) terimini, çok sayıda şey bir bütün gibi davrandığında ortaya çıkan kendi kendini organize eden, kolektif davranışları tanımlamak için kullanırlar. Cansız atomların birleşmesi canlı hücreleri meydana getirir; su molekülleri dalgalar oluşturur; sığırcık kuşlarının mırıltıları, değışen ama tanımlanabilir kalıplarla gökyüzünde süzölür; hücreler kasların hareket etmesini ve kalplerin atmasını sağlar. Kritik olarak, ortaya çıkan yetenekler, birçok bireysel parçayı içeren sistemlerde ortaya çıkar. Ancak araştırmacılar, son zamanlarda bu becerileri büyük dil modellerinde belgeleyebildiler, çünkü bu modeller çok büyük boyutlara ulaştı».

Ornes Stephen, The Unpredictable Abilities Emerging From Large AI Models, Quanta Magazine, March 16, 2023, <https://www.quantamagazine.org/the-unpredictable-abilities-emerging-from-large-ai-models-20230316/>

Büyük dil modellerinin ortaya çıkan yeteneklerinde, zuhur eden bir yeteneđi , küçük modellerde mevcut olmayan, ancak büyük modellerde mevcut olan bir yetenek olarak tanımlayabiliriz. Zuhur etme nadir olarak görülen bir olgu mudur, yoksa deđil midir? GPT-3, Chinchilla, PaLM gibi dil modellerinin ölçeklendirilmesiyle deneysel olarak keşfedilmiş zuhur etmede 100'den fazla ortaya çıkan yetenek örneđi olduđu gözlenmiş durumda.

Wei Jason, 137 emergent abilities of large language models, Nov 14, 2022,  
<https://www.jasonwei.net/blog/emergence>



Bir büyük dil modelinin performansı şu faktörlere bağlıdır:

- Model parametrelerinin sayısı  $N$ ,
- Veri kümesinin boyutu  $D$  ve
- Eğitim için kullanılan hesaplama miktarı  $C$ .

Kaplan Jared et al., Scaling Laws for Neural Language Models, arXiv:2001.08361v1 [cs.LG] 23 Jan 2020, s.3.

“Günümüzün yapay zekâ dil modellerinin garip bir özelliđi, genellikle yapımcılarının tahmin etmediđi şekillerde hareket etmeleri veya özel olarak programlanmadıkları becerileri edinmeleridir. Yapay zekâ arařtırmacıları bunlara "ortaya çıkan davranıřlar" diyor ve bunun pek çok örneđi var. Bir cümledeki bir sonraki kelimeyi tahmin etmek için eđitilmiş bir algoritma kendiliđinden kodlamayı öğrenebilir”.

Roose Kevin, GPT-4 Is Exciting and Scary, The New York Times, 15 Mart, 2023.

Dil modelleri, boyutları arttıkça benzer şekilde niteliksel olarak yeni davranışlar gösterirler: Örneğin, bilgisayar kodu yazımında, satrançta oynamada, tıbbi durumları teşhis etmede ve diller arası çeviri yapmada.

Biderman Stella, Raff Edward, Neural language models are effective plagiarists, 2022, <https://arxiv.org/abs/2201.07406>

Stöckl Andreas, Watching a language model learning chess. In Proceedings of the International Conference on Recent Advances in Natural Language Processing(RANLP2021),pp.1369–1379,Held Online,September 2021.INCOMALtd., <https://aclanthology.org/2021.ranlp-1.153>.

Rasmy Laila, Xiang Yang, Xie Ziqian, Tao Cui, Zhi Degui, Med-BERT:Pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction.npj Digital Medicine, 4 (1):86,2021.doi:10.1038/s41746-021-00455-y.<https://doi.org/10.1038/s41746-02100455-y>

Sutskever Ilya, Vinyals Oriol, Le.Quoc V., Sequence to sequence learning with neural networks,2014, <https://arxiv.org/abs/1409.3215>.

Bu yetenekler deneysel olarak gözlemlenmiştir, ancak yeni atılımların hangi ölçekte gerçekleşeceğini güvenilir bir şekilde tahmin edilemiyor.

Srivastava Aarohi et al., Beyond The Imitation Game: Quantifying and Extrapolating The Capabilities of Language Models, arXiv:2206.04615v2 [cs.CL] 10 Jun 2022.

Büyük dil modellerinin boyutu büyüdükçe, bu modeller sıfıra yakın performanstan bazen inanılmaz hızlı adımlarla ve öngörülemeyen ölçeklerde çok yüksek teknolojik bir performansa yükselirler. Benzer şekilde, tutarlı resimler çizemeyen büyüyen bir çocuğu düşünün. Büyüdükçe, beyninin boyutu sorunsuz bir şekilde artar, ince motor becerileri sorunsuz bir şekilde gelişir; ancak, belirli bir kritik yaşa geldiğinde, çizim yeteneğinde süreksiz bir “sıçrama” olur. Bu sıçrayış, çocuğun ince motor becerileri kademeli olarak gelişme göstermesine rağmen, aniden inanılmaz portreler çizmesini sağlar.

<https://www.assemblyai.com/blog/emergent-abilities-of-large-language-models/>

Wei Jason, Scaling, emergence, and reasoning in large language models, Twitter, April, 2023



“Bu davranış tamamen yeniydi ve GPT-3'ün tasarımcılarını bile şaşırttı. Onlar GPT-3'ü bilgisayar kodu yazması için yapmamışlardı. Onu tek bir şey yapmak için inşa etmişlerdi: bir kelime dizisindeki bir sonraki kelimeyi tahmin etmek”.

Metz Cade, Meet GPT-3. It Has Learned to Code (and Blog and Argue), The New York Times, Nov. 24, 2020, <https://www.nytimes.com/2020/11/24/science/artificial-intelligence-ai-gpt3.html>

Google Research'te çalışan bilgisayar bilimcisi Ethan Dyer, "Sürprizler beklemeye alışık olsam da bu modellerin yapabildiklerine şaşırdım" diyor. Bu şaşırtıcı çünkü bu modeller, bir metin dizisini girdi olarak kabul edip sonra tamamen istatistiklere dayalı olarak tekrar tekrar tahmin gerçekleştiriyor. Bilgisayar bilimcileri, ölçeği büyütmenin bilinen görevlerdeki performansı artıracığını tahmin ettiler, ancak modellerin birdenbire bu kadar çok yeni, öngörülemeyen görevlerin üstesinden gelmesini beklemiyorlardı. Yeni analizler, bazı görevler ve bazı modeller için, modelin işlevselliğinin fırladığı bir karmaşıklık eşiği olduğunu öne sürüyor. Aynı zamanda bu modeller karanlık bir yüz de öneriyorlar: Karmaşıklık arttıkça, bazı modeller yanıtlarında yeni önyargılar ve yanlışlıklar ortaya koyuyor.

- Ornes Stephen, The Unpredictable Abilities Emerging From Large AI Models, Quanta Magazine, March 16, 2023, <https://www.quantamagazine.org/the-unpredictable-abilities-emerging-from-large-ai-models-20230316/>



# Şimdi de ađlara bakalım:

Ađlar büyüdükçe ađların büyümesinde de bir sıçrama gözlenerek dev bir bileşen oluşur. Bir düşünce ekolüne göre, bizler tek başına bireyler olarak değil, öncelikle ađlar ve sistemler içinde zekiyiz. Bilgiyi ađlarda tutuyoruz.

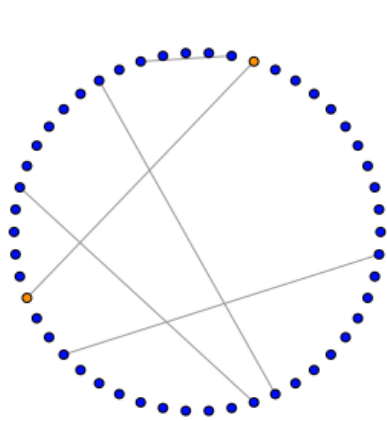
Siemens George, Has GPT-4 really passed the startling threshold of human-level artificial intelligence? Well, it depends, The Conversation, March 31, 2023, <https://theconversation.com/has-gpt-4-really-passed-the-startling-threshold-of-human-level-artificial-intelligence-well-it-depends-202856>

Ölçek meselesi ağlar için de geçerli. Bilindiği gibi bir çizge, bir dizi düğüm ve düğümler arasındaki bir dizi bağlantıdan oluşur. Bu noktadan hareketle, farklı adalardaki düğümler arasında hiçbir bağlantının bulunmadığı birkaç düğüm adasından oluşan bir çizgeye sahip olmak mümkündür. Ağ kuramında bu adalara “bileşenler” adı verilir. Örneğin bir sosyal ağda, gruplar arasında hiçbir arkadaşlığın olmadığı iki ayrı arkadaş grubu görebiliriz, bu durumda iki bileşenimiz vardır.

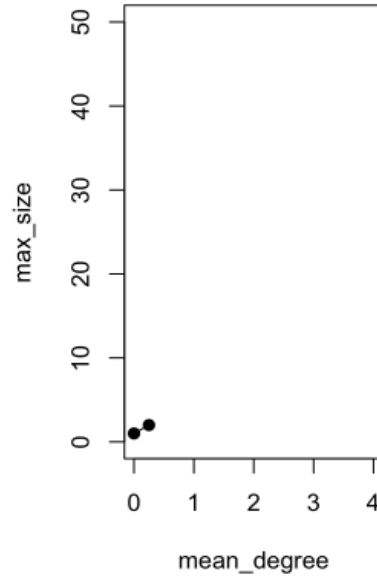
Birçok ağda, ağ büyüdükçe bileşenler kademeli olarak birbirine bağlanır. Bir noktada düğümlerin önemli bir kısmının tek bir dev bileşende birbirine bağlandığını söyleyebiliriz. Bir bileşenin ne zaman dev bileşen haline geldiğine dair kesin bir tanım yoktur, bu ağın türüne bağlıdır. Basitçe, şekilde de görüldüğü gibi, düğümlerin çoğunluğu birbirine bağlandığında (yarısından fazlası) bir bileşenin dev bileşen haline geldiğini söyleyebiliriz.

<https://www.futurelearn.com/info/courses/social-media/0/steps/16048>

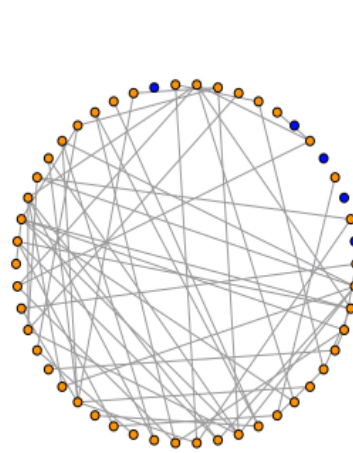
$p^*(n-1)=0.25$ , max\_size=2



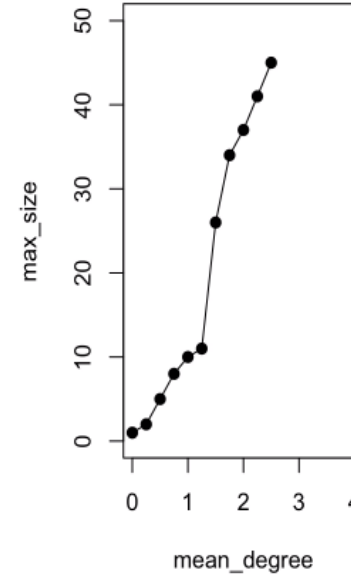
Summary



$p^*(n-1)=2.50$ , max\_size=45



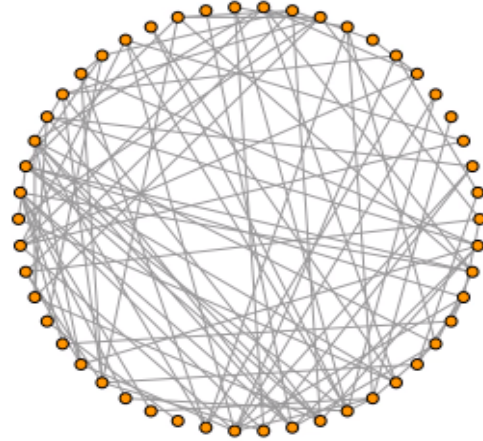
Summary



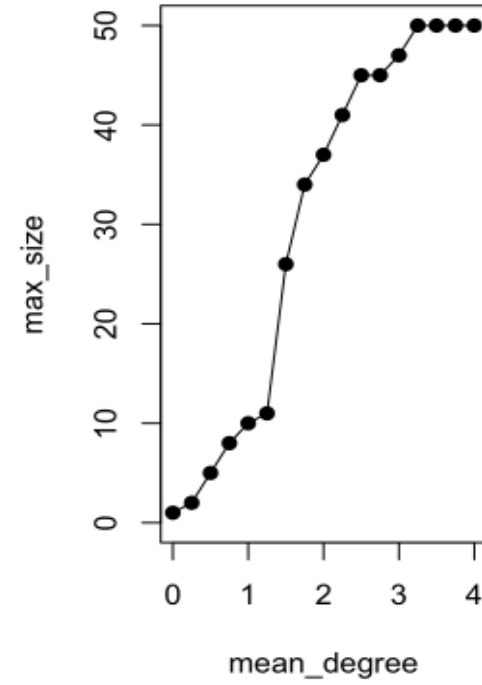
Dev bileşen oluşumu. Bu durumda, en büyük bileşen  $p(n-1) < 0,75$  olduğunda nispeten küçük görünür ve 1 ila 3 aralığında hızla artar.

<https://brianzhang01.github.io/2018/07/random-graphs-and-giant-components/>

$p^*(n-1)=4.00$ , max\_size=50



Summary



“Nasıl çalıştıklarına dair kuralları yazmak mümkün olsa da, büyük dil modellerinin çıktıları tamamen öngörülebilir değildir. Bu son derece büyük abaküslerin, onları yapan insanları bile şaşırtacak şekilde, daha küçük olanların yapamayacağı şeyleri yapabildiği ortaya çıkmıştır. OpenAI'da araştırmacı olan Jason Wei, çeşitli farklı büyük dil modellerinde 137 adet ‘ortaya çıkan’ yetenek saymıştır. Ortaya çıkan yetenekler sihirli değildir - hepsi büyük dil modellerinin eğitim verilerinde (ya da onlara verilen istemlerde) bir şekilde temsil edilirler, ancak bu modeller boyutlarında belirli, çok büyük bir eşiği geçene kadar belirgin hale gelmezler. Bir boyutta, bir büyük dil modeli, Almanca'da cinsiyet içeren cümlelerin nasıl yazılacağını, bunu rassal bir şekilde yapmaktan daha iyisini bilmez. Ancak modeli biraz daha büyüttüğünüzde, birdenbire yeni bir yetenek ortaya çıkmaktadır».

Large, creative AI models will transform lives and labour markets, Economist, April 22, 2023,  
<https://www.economist.com/interactive/science-and-technology/2023/04/22/large-creative-ai-models-will-transform-how-we-live-and-work>

Bir söyleşide Sam Altman şunları söylüyor:

“Altman: Bir dizi şey. Tek bir tanesine işaret etmek zor ama GPT-2'yi yaptığımızda, ki bu hâlâ pek çok açıdan zayıftı, ölçeklendirme yasalarına bakıp neler olacağını görebiliyordunuz. ‘Hımm. Bu çok, çok ileri gidebilir’ dedim. Bu konuda çok heyecanlandım. Bu konuda aşırı heyecanlı olmayı hiç bırakmadım.

Swisher: Ölçeklendirilmiş bir şey gördünüz mü ya da ...

Altman: Daha fazla hesaplama ve daha fazla veri ile sistemi ne kadar iyi hale getirebileceğimize dair verilere bakıyorduk”.

Swisher Kara, Sam Altman on What Makes Him ‘Super Nervous’ About AI, *Intelligencer*, March 23, 2023, <https://nymag.com/intelligencer/2023/03/on-with-kara-swisher-sam-altman-on-the-ai-revolution.html>

Ancak Altman, GPT-4'ün 100 trilyon parametrelili olduğu söylentisinin nereden çıktığını bilmediğini ve bunun "tamamen saçmalık" olduğunu söylüyor, ayrıca GPT-5'i eğitmediklerini belirterek, "Yapmıyoruz ve bir süre daha yapmayacağız" diyerek GPT-4 konusunda her türlü güvenlik sorununa sahip olduğunu düşündüğü başka şeyler yapacaklarını ve güvenlik konularında artan bir titizlik göstereceklerini ekliyor.

Stern Jacob, GPT-4 Might Just Be a Bloated, Pointless Mess, The Atlantic, March 6, 2023,

<https://www.theatlantic.com/technology/archive/2023/03/openai-gpt-4-parameters-power-debate/673290/>

Miller Ron, Sam Altman: Size of LLM's won't matter as much moving forward , TechCrunch, April 14, 2023,

[https://techcrunch.com/2023/04/14/sam-altman-size-of-llms-wont-matter-as-much-moving-forward/?guccounter=1&guce\\_referrer=aHR0cHM6Ly93d3cuZ29vZ2xllmNvbS8&guce\\_referrer\\_sig=AQAAAlrrlv2tIT5GFudwU7mlqmkNObr0-INjUI8CGSRX8aTMkLoeGxcl4K5-nt2GMwiarL79mSz47PIJgBblaGIU0unNXr5fWIUVUoZs6G2xIHjhsDMsBqzjos-AylCV5\\_Y3IEioTUxfNNpWsBeVXTaUgfuH1gjplY6qD\\_WXYiL\\_Dg1P](https://techcrunch.com/2023/04/14/sam-altman-size-of-llms-wont-matter-as-much-moving-forward/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xllmNvbS8&guce_referrer_sig=AQAAAlrrlv2tIT5GFudwU7mlqmkNObr0-INjUI8CGSRX8aTMkLoeGxcl4K5-nt2GMwiarL79mSz47PIJgBblaGIU0unNXr5fWIUVUoZs6G2xIHjhsDMsBqzjos-AylCV5_Y3IEioTUxfNNpWsBeVXTaUgfuH1gjplY6qD_WXYiL_Dg1P)

Altman, büyük dil modellerinde büyüklüğünün sınırlarına yaklaştığımızı inanıyor, "Bence bu dev modellerin olduğu dönemin sonuna geldik ve onları başka şekillerde daha iyi hale getireceğiz" diyor.

Miller Ron, Sam Altman: Size of LLM's won't matter as much moving forward , TechCrunch, April 14, 2023, [https://techcrunch.com/2023/04/14/sam-altman-size-of-llms-wont-matter-as-much-moving-forward/?guccounter=1&guce\\_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce\\_referrer\\_sig=AQAAAIrrlv2tIT5GFudwU7mlqmkNObr0-INjUI8CGSRX8aTMkLoeGxcl4K5-nt2GMwiarL79mSz47PIJgBblaGIU0unNXr5fWIUVUoZs6G2xIHjhsDMsBqzjos-AylCV5\\_Y3IEioTUxfNNpWsBeVXTaUgfuH1gjplY6qD\\_WXYiL\\_Dg1P](https://techcrunch.com/2023/04/14/sam-altman-size-of-llms-wont-matter-as-much-moving-forward/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAIrrlv2tIT5GFudwU7mlqmkNObr0-INjUI8CGSRX8aTMkLoeGxcl4K5-nt2GMwiarL79mSz47PIJgBblaGIU0unNXr5fWIUVUoZs6G2xIHjhsDMsBqzjos-AylCV5_Y3IEioTUxfNNpWsBeVXTaUgfuH1gjplY6qD_WXYiL_Dg1P))



Birileri 2022 yılında, “ölçekten elde edilen gelişmelerin henüz sabitlenmediğini gösteriyor” sonucuna varmıştı.

Chowdhery Aakanksha et al., PaLM: Scaling Language Modeling with Pathways, arXiv:2204.02311v5 [cs.CL] 5 Oct 2022, s.50.

Oysa 2023 yılında Altman, GPT-4'ün OpenAI'in modelleri büyütme ve onları daha fazla veri ile besleme stratejisinden ortaya çıkan son büyük gelişme olabileceğini belirtiyor. GPT-4'ü açıklayan makalede OpenAI, tahminlerinin model boyutunu büyütmenin getirisinin azaldığını gösterdiğini söylüyor. Altman, şirketin birkaç veri merkezi inşa edebileceği ve bunları ne kadar hızlı inşa edebileceği konusunda da fiziksel sınırlar olduğunu ekliyor.

Knight Will, OpenAI's CEO Says the Age of Giant AI Models Is Already Over, Wired, April 17, 2023, <https://www.wired.com/story/openai-ceo-sam-altman-the-age-of-giant-ai-models-is-already-over/amp>

MIT'de düzenlenen bir etkinlikte dinleyicilere "Bence devasa, devasa modellerin olacağı dönemin sonuna geldik, onları başka yollarla daha iyi hale getireceğiz" dedi.

Teşekkürler