GLM 4.5: Yapay Zekada Yeni Nesil Güç; Çinli yapay zeka şirketi Zipo AI tarafından tanıtılan, **tamamen açık kaynaklı ve ücretsiz** yeni nesil bir dil modelidir. Claude dahil olmak üzere çoğu güçlü yapay zeka modelini geride bırakmayı başarmıştır.

Bu makaleyi Spotify'da sesli olarak dinlemek için podcast'ine bu linkten ulaşabilirsiniz.

1.1 GLM 4.5: Yeni Nesil Özellikleri ve Yetenekleri

GLM 4.5, sadece sıradan bir dil modeli olmanın ötesinde, yeni nesil yapay zekayı temsil eden özellikler sunar. **Az parametreyle çok daha zeki modeller geliştirme** eğiliminin bir örneğidir, bu da sektörün evrildiği yeni yönü gösterir.

Model, özel görevler için eğitilmiş farklı **uzmanlara** sahiptir; kullanıcı ne istediğini yazdığında, sadece işine yarayacak uzmanları aktif eder, bu da **daha hızlı ve net sonuçlar** verir. Kullandığı yeni algoritmalar sayesinde birçok dil modelini geride bırakmıştır.

Bu makaleyi Youtube'da görüntülü olarak izlemek için videosuna bu linkten ulaşabilirsiniz.

1.2 GLM 4.5: Ücretsiz ve Güçlü Yapay Zeka

- **Tamamen açık kaynak ve ücretsizdir**.
- Kaynak kodları, eğitilmiş ağırlıkları ve entegrasyon araçları halka açıktır, bu da modelin istenildiği gibi kullanılabilmesini ve özel projelere entegre edilerek geliştirilebilmesini sağlar.
- Genel performansa bakıldığında, GLM 4.5'in **hem açık kaynak hem de kapalı kaynak modellerin çoğunu geride bıraktığı** görülmektedir. Özellikle kodlama performansında Kimy K2 ve Quantry Coder gibi güçlü modelleri geride bırakmış, Cloud modeliyle genel başarıda başa baş giderken, **araç çağırma görevlerinde tüm modellerin önüne geçmiştir**.
- **Temel Özellikler**

- **Parametre Yapısı:**
  - Ana sürüm 355 milyar parametreye sahiptir, ancak yalnızca 32 milyar parametresi aktif olarak çalışır.
  - İkinci versiyon (GLM 4.5) 106 milyar parametre içerir, ancak yalnızca 12 milyarlık kısmı aktif şekilde kullanılır.
  - Bu, **modelin tamamını değil, yalnızca küçük ve verimli bir kısmını çalıştırma** imkanı sunar, bu da **daha doğru, iyi yanıtlar, daha hızlı işlem süresi ve daha az bellek tüketimi** sağlar. Aktif parametrelerin sınırlı tutulması, kaynak tüketimi açısından büyük önem taşır.
- **Modlar Arası Geçiş:** Modeli ister **"düşünen"** ister **"düşünmeden yanıt veren"** modda kullanabilirsiniz.
  - **Düşünme Modu:** Karmaşık muhakeme gerektiren çok adımlı problemler için

modelin adım adım mantıksal akıl yürütme yapmasını sağlar.
    - **Düşünmeden Yanıt Verme Modu:** Basit bilgi soruları, çeviri veya hızlı yanıt gerektiren komutlar için devreye girer, gecikme süresi olmadan hızlı sonuç üretir.
- **Esneklik:** Bu esnek çalışma biçimi, GLM 4.5'i hem **reaktif hem de proaktif görevlerde** kullanılabilir hale getirir.
- **Bağlam Uzunluğu (Context Length):** Tam **128.000 KTEX** uzunluğuna sahiptir, bu da karmaşık görevlerde büyük miktarda veriyi tek seferde işleyebilmesini sağlar. Örneğin, bir ajan sistemi geçmişte yaptığı onlarca işlemi aynı anda analiz edebilir.
- **Tek Modelde Çeşitlilik:** Open AI modelleri gibi rakip modeller genellikle farklı yetenekleri ayrı ayrı modellerle çözerken, GLM 4.5 tüm bu farklı alanlardaki yetenekleri **tek bir modelde birleştirmeyi** başarmıştır.
- **Doğrudan Düzenleme:** Oluşturduğu projelerde **istenilen elementi doğrudan düzenleyebilme** yeteneği sunar.

1.3 Uygulama Alanları ve Yetenekler

GLM 4.5, yalnızca sorulara yanıt vermenin ötesinde geniş bir uygulama yelpazesi sunar:

- **Oyun Geliştirme:** İki veya üç boyutlu oyunlar geliştirebilir.
- **Web Sitesi Geliştirme:**
    - Full-stack bir web sitesi kurup yayına alabilir.
    - Örneğin, kullanıcıların açıklama yazarak Cyberpunk temalı oyun kartları oluşturabilecekleri, bu kartları veri tabanında saklayabilecekleri ve görüntüleyebilecekleri **interaktif bir web sitesini sıfırdan inşa edebilir**.
    - Modern teknolojilerle (Next.js, TypeScript, Tailwind CSS gibi) birlikte çalışarak hem estetik hem de işlevsel projeler üretebilir.
    - Hatta Amazon'un e-ticaret sitesi gibi karmaşık sitelerin **full-stack taslaklarını** kısa sürede oluşturabilir (web arayüzünden gelen kısıtlı çalışma ortamı nedeniyle tüm işlevler çalışmasa da iyi bir başlangıç sunar).
- **Araştırma ve İçerik Oluşturma:**
    - Belirli bir konuda araştırma yapabilir.
    - Yapılan araştırmadan hareketle **dokümanlar, slaytlar veya tanıtım materyalleri** hazırlayabilir. Örneğin, Kodla YouTube kanalını araştırıp tüm içerik, link ve detayları analiz ederek kendi yorumlarını da ekleyerek **güzel bir slayt sunumu** oluşturabilir ve otomatik olarak yayınlayabilir.
    - Görsel içerikler ve **posterler** oluşturabilir ve bu posterleri düzenleme imkanı sunar.
- **Kodlama ve Mühendislik:** Kod tasarım ve içerik konularını kendi başına halledebilir. Ajan sistemleri, akıl yürütme görevleri ve kodlama gibi alanlarda oldukça güçlüdür.

1.4 Mimari ve Çalışma Mekanizması

GLM 4.5, geleneksel dil modellerinden farklılaşan yenilikçi bir mimariye ve çalışma

mekanizmasına sahiptir:

- **Uzman Odaklılık:** Modelin içinde özel görevler için eğitilmiş farklı **uzmanlar** bulunur. Bu sayede çok daha hızlı ve net sonuçlar verir.
- **MTP (Multi-Token Prediction) Tekniği:** Modelin aynı anda birden fazla token (örneğin dört token) üretebilmesini sağlar. Bu, çıktı üretimini ciddi oranda hızlandırır ve modelin **paralel tahmin** yapmasına olanak tanır. GLM 4.5 bu yetenek için özel bir MTP katmanı ile eğitilmiştir.
- **Spekülatif Kod Çözme (Speculative Decoding):** Küçük bir modelin çok hızlı bir şekilde 5-10 token üretmesi, büyük modelin ise bu tokenları analiz edip doğrulaması veya düzeltmesi prensibine dayanır. Bu sayede büyük model daha az çalışır, cevaplar çok daha hızlı gelir ve kaliteden ödün verilmez.
- **"Düşünme" Davranışının Taklidi:** Model aslında yine yalnızca bir sonraki tokenı tahmin eder. Ancak çok aşamalı akıl yürütme görevlerinde, daha uzun bağlamı dikkate alarak daha derin mantık zincirleri kurmaya çalışır. Bu yetenek için **ayrı verilerle özel olarak eğitilmiştir**, böylece kullanıcının isteğini analiz edip adım adım planlayarak uygulamaya geçer. Algoritma düşünmez, ancak **düşünme davranışını çok iyi taklit eder**.
- **MOI (Mixture of Experts) Mimarisi:** Kimik K2'da olduğu gibi her katmanda birçok uzmanın bulunduğu ancak **yalnızca bazılarının aktif hale geldiği** bu yapı sayesinde model hem hızlı hem de verimli çalışır.
- **Derinlik ve Genişlik Dengesi:** Modelde genişlik azaltılıp derinlik artırılmıştır, çünkü **derin modellerin daha iyi düşünebildiği** gözlemlenmiştir.
- **Attention Mekanizması:** Normal modellerden neredeyse **2,5 kat daha fazla attention** kullanır. Eğitim sürecinde oluşan dengesizlikleri engellemek için **QK Norma** adlı bir yöntemle attention dengesi sağlanır.
- **Takviyeli Öğrenme (Reinforcement Learning – RL) Mimarisi:** GLM 4.5, klasik RL mimarisinden farklı olarak **çok daha büyük ölçekli ve ayrıştırılmış bir sistem** kullanır.
    - **Görevin Dağıtımı:** Birden fazla router server üzerinden (GPU destekli) görev dağılımı yapılır, bu sunucular belirli bir görev için modelden veri üretir.
    - **Özelleştirilebilir Veri Üretimi:** "Custom Rollout Generation" bileşeni, çeşitli bağlamlarda ajan deneyimleri üretir ve bu veriler bir "Rollout Buffer"da depolanır.
    - **Megatron ile Eğitim:** Toplanan veriler, çok sayıda GPU üzerinde paralel çalışabilen "Megatron" eğitim altyapısına aktarılır ve modeli eğitir.
    - **Sürekli Gelişim Döngüsü:** Güncellenmiş model ağırlıkları tekrar rollout sunucularına gönderilir, bu sayede yeni rollout'lar daha iyi bir modelle yapılır ve model sürekli kendini geliştirir.
    - **Ayrık İşlem:** Veri üretimi ve eğitim işlemleri farklı donanımlarda yürütüldüğü için, veri üretiminin eğitim sürecini yavaşlatmasının önüne geçilir ve GPU'lar daima aktif kalır.
    - **Yatay Ölçeklenme:** Rollout sunucularını yatayda ölçekleyerek büyük dil modellerinde RL uygulamasının zorlukları aşılır.

- **Karma Hassasiyet (Mixed Precision):** Veri üretimi sırasında daha hızlı FP8 formatı, model eğitimi sırasında daha dengeli ve doğruluğu yüksek BF16 formatı kullanılır, bu sayede hem hız hem doğruluk korunmuş olur.

Bu mimari, GLM 4.5'i sadece güçlü değil, aynı zamanda sürekli öğrenen, kendini güncelleyen ve uzun etkileşimli görevlerde bile **kararlı kalabilen bir yapıya** sahip kılar.

1.5 Eğitim Süreci

GLM 4.5'in eğitim süreci üç ana aşamadan oluşur:

1. **Ön Eğitim (Pre-training):** Modelin temeli atılır.
   - **Genel Dil Eğitimi:** 15.000 tokenlık veriler ve 4.000 bağlam uzunluğuyla, kitaplar ve makaleler gibi çok büyük metin koleksiyonları kullanılarak dilin yapısı öğrenilir.
   - **Kodlama ve Düşünme Eğitimi:** 7.000 token ve 4.000 bağlamla modelin mantıklı çıkarımlar yapması, kodlama ve problem çözme yetenekleri kazanması sağlanır.
2. **Ara Eğitim (Mid-training):** Modelin uzmanlaşmasını sağlar.
   - **Repo-level Kod Eğitimi:** 500.000 token ve 32.000 bağlamlı gerçek kod depoları üzerinden model eğitilir. Bu sayede model sadece fonksiyonları değil, bir projenin tüm yapısını, ilişkilerini ve bağımlılıklarını öğrenir.
   - **Sentetik Akıl Yürütme Verisi (Synthetic Reasoning Data):** 500.000 token ve 32.000 bağlamla yapay olarak üretilmiş mantık problemleriyle zincirleme düşünme, planlama ve çıkarım yapma yetenekleri geliştirilir.
   - **Uzun Bağlamlı Ajan Eğitimi (Long Context Agent training):** 100.000 token ve 128.000 bağlam kullanılarak modelin uzun görev zincirleri, etkileşimli ajan davranışları ve karmaşık görev planlamaları gibi alanlarda uzmanlaşması sağlanır.
3. **Son Eğitim (Post-training):** Bu bölüm iki ana adımdan oluşur: Uzman Eğitimi ve Birleşik Eğitim.
   - **Uzman Eğitimi:** Temel bir model alınarak üç farklı alanda uzmanlaşması sağlanır:
     - Mantıksal Akıl Yürütme görevleri (64.000 tokenlık uzun bağlamda tek aşamalı RL ile).
     - Ajan yetkinlikleri (yazılım mühendisliği, bilgi arama, araç kullanma).
     - Genel dil görevleri. Her bir uzmanlık alanı sonunda elde edilen modeller, "self distillation" adı verilen bir yöntemle tek bir birleşik modele aktarılır.
   - **Birleşik Eğitim:** Bu uzmanlıklar tek bir yapı altında entegre edilir. Birleşik model önce genel veri setleriyle görsel olarak ince ayarlarla eğitilir, ardından sırasıyla mantıksal akıl yürütme, ajan yetkinlikleri ve genel görevler aşamalarında RL ile pekiştirilir.
   - Bu süreç sayesinde GLM 4.5, genel görevleri başarmakla kalmayıp, zorlayıcı akıl yürütme süreçlerini sürdürebilen, kendi başına araçları kullanabilen ve karmaşık görevleri planlayarak çözüm üretebilen **güçlü ve yönlendirilmiş bir yapay**

zeka modeli haline gelmiştir.

1.6 Pratik Deneyimler ve Sonuçlar

Yapılan pratik testlerde GLM 4.5'in iddialı yetenekleri gözlemlenmiştir:

- **Web Sitesi Oluşturma:** Kullanıcının isteğini analiz ederek yüzlerce satır kodu çok hızlı bir şekilde yazıp **otomatik olarak projeyi oluşturmuş ve yayına almış**; oluşturulan web sitesi oldukça kaliteli ve mobilde de tamamen uyumlu bulunmuştur. Ayrıca, oluşturulan elementlerin **doğrudan düzenlenebilmesi** büyük bir kolaylık sağlamıştır.
- **Performans Karşılaştırması:** Daha önce kullanılan tüm açık kaynak modellerle kıyaslandığında **çok daha fazla özelliğe sahip ve performans açısından çok daha iyi** olduğu belirtilmiştir.
- **E-ticaret Sitesi Taslağı:** Kısa sürede tam bir full-stack proje taslağı (örneğin Amazon'un e-ticaret sitesi için) oluşturabilmiştir; her ne kadar web arayüzündeki kısıtlı ortamda tüm işlevler çalışmasa da, iyi bir başlangıç sunabildiği ve kendi bilgisayarına kurulduğunda çok daha ileri düzey projeler çıkarılabileceği düşünülmüştür.
- **Slayt ve Afiş Oluşturma:** Kodla YouTube kanalının araştırılması ve sonuçlarla bir sunum hazırlanması istendiğinde, model tüm verileri analiz edip **kendi yorumlarını da ekleyerek çok güzel bir slayt oluşturmuş ve otomatik olarak yayınlamıştır**. Benzer şekilde, önceki bilgilerle çok iyi bir **afiş tasarlayabilmiş ve düzenleme imkanı** sunmuştur.
- **Genel İzlenim:** Modelin **gerçekten etkileyici** olduğu, hatta son zamanlarda kullanılan en iyi modellerden biri olabileceği ifade edilmiştir. Benchmark testlerinde belirtildiği kadar başarılı olduğu ve özellikle **düşünme modunun oldukça başarılı çalıştığı** gözlemlenmiştir.

1.7 GLM 4.5'i Rakipleriyle Karşılaştırılması

GLM 4.5'i rakipleriyle karşılaştıran ve üstünlüklerini gösteren bir tablo aşağıdadır:

Özellik / Kategori	GLM 4.5	Rakipler (genel)	GLM 4.5 Üstünlükleri
Açık Kaynak ve Maliyet	**Tamamen açık kaynak ve tamamen ücretsizdir**. Kaynak kodları, eğitilmiş ağırlıkları ve entegrasyon araçları halka açıktır.	Claude dahil olmak üzere çoğu güçlü model kapalı kaynaklı olabilir veya ücretli abonelik gerektirebilir.	GLM 4.5'in **tamamen açık kaynak ve ücretsiz** olması, modelin özgürce kullanılmasına, istenildiği gibi geliştirilmesine ve özel projelere entegre edilmesine olanak tanır.

Parametre Kullanımı ve Verimlilik	Ana sürüm 355 milyar toplam parametreye sahipken **yalnızca 32 milyar parametresi aktif olarak çalışır**. İkinci versiyon (GLM 4.5 Lite) 106 milyar parametre içerir ancak **yalnızca 12 milyarlık kısmı aktif şekilde kullanılır**. Modelin içinde özel görevler için eğitilmiş "uzmanlar" bulunur ve yalnızca işe yarayacak uzmanlar aktif edilir.	Büyük modeller genellikle tüm parametre setini çalıştırır, bu da gereksiz hesaplamalara yol açabilir.	**Çok daha az aktif parametreyle yüksek performans** elde edebilir. Bu sayede **çok daha hızlı ve çok daha net sonuçlar verirken, daha az bellek tüketir**. Bu yaklaşım, daha az veriyle daha zeki modeller geliştirme sektör trendini yansıtır.
Mimari ve Çalışma Modları	**MOI (Mixture of Experts) mimarisi** kullanır; her katmanda birçok uzman bulunur ancak yalnızca bazıları aktif hale gelir. Genişliği azaltılıp derinliği artırılmıştır. **Normal modellerden neredeyse 2.5 kat daha fazla attention mekanizması** kullanır. **"Düşünme modu" (karmaşık muhakeme için adım adım mantıksal akıl yürütme)** ve **"düşünmeden yanıt verme modu" (basit sorular için hızlı yanıt)** arasında geçiş yapabilir.	Diğer dil modelleri genellikle tek bir akıl yürütme veya yanıt verme mekanizması kullanır. Kimik K2 de MOI mimarisi kullanır.	MOI mimarisi sayesinde **hem hızlı hem de verimli** çalışır. Derin mimarisiyle daha iyi düşünme yeteneği gösterir. Esnek çalışma biçimi, GLM 4.5'i **hem reaktif hem de proaktif görevlerde kullanılabilir** hale getirir.

Genel Performans ve Yetenekler	**Genel performansta hem açık kaynak hem de kapalı kaynak modellerin çoğunu geride bırakmıştır**. **Claude modeliyle genel başarı bakımından kafa kafaya ilerlese de, araç çağırma görevlerinde tüm modellerin önüne geçmiştir**.	OpenAI modelleri gibi rakipler genellikle farklı yetenekleri (örn. kodlama, matematik) ayrı ayrı modellerle çözüyordu. Bazı rakiplerin benchmark testlerindeki başarıları gerçek kullanımda abartılı bulunabilir.	**Tüm farklı alanlardaki yetenekleri tek bir modelde birleştirmeyi başarmıştır**. **128.000 KTEXT uzunluğunda bağlama sahip olması**, modelin karmaşık görevlerde büyük miktarda veriyi tek seferde işlemesine olanak tanır. Web araştırmaları ve çok katmanlı kodlama senaryolarında **pek çok modelden daha iyi sonuçlar** üretir.
Kodlama ve Proje Geliştirme	Kodlama performansı **oldukça iddialıdır**, Kimy K2 ve Quantry Coder gibi güçlü modelleri geride bırakmıştır. **Full-stack proje geliştirme** (frontend'den backend'e, tasarımdan kullanıcı deneyimine kadar her aşama) yeteneğine sahiptir. 2D/3D oyunlar, web siteleri oluşturabilir; araştırma yapıp doküman, slayt ve tanıtım materyalleri hazırlayabilir. Modern teknolojilerle (Next.js, TypeScript, Tailwind CSS) çalışır.	Rakip modeller genellikle ayrı ayrı kodlama veya tasarım yetenekleri sunar. ChatGPT gibi modellere entegre slayt oluşturma özellikleri mevcuttur.	**Tek bir modelde tam teşekküllü full-stack proje geliştirme yeteneği** sunar. **Daha az aktif parametreyle güçlü kodlama performansı** sergiler. **Oluşturulan projeleri otomatik olarak yayına alabilir ve düzenlenebilir** olması büyük bir avantajdır.

Eğitim Süreci ve İnovasyonlar	Eğitim süreci **üç ana aşamadan (pre-training, mid-training, post-training) oluşur**. **Reinforcement Learning (RL) için büyük ölçekli ve ayrıştırılmış bir sistem kurgulanmıştır**; veri üretimi ve eğitim farklı donanımlarda yürütülerek GPU'ların daima aktif kalması sağlanır. **MTP (Multi-Token Prediction)** ile aynı anda birden fazla token üretebilir, çıktıyı hızlandırır. **Speculative Decoding** ile küçük bir modelin taslak üretip büyük modelin kontrol etmesiyle hız ve doğruluk sağlanır. Post-training aşamasında **uzman eğitimi** (mantıksal akıl yürütme, ajan yetkinlikleri, genel dil görevleri) ve **birleşik eğitim** (self-distillation ile uzmanlıkların entegrasyonu) uygulanır.	Klasik RL sistemlerinde veri üretimi eğitim sürecini yavaşlatabilir. Çoğu dil modeli ardışık olarak tek token üretir.	**RL'deki ölçeklenebilirlik ve verimlilik sorunlarını çözen ayrıştırılmış mimarisi** sayesinde sürekli öğrenen ve uzun etkileşimli görevlerde kararlı kalabilen bir yapıya sahiptir. **MTP ve Speculative Decoding** teknikleriyle **çıktı üretimini ciddi oranda hızlandırır** ve kaliteden ödün vermez. **Uzmanlaşmış ve birleşik eğitim süreci**, modelin karmaşık görevleri planlama, araçları kullanma ve çözüm üretme yeteneklerini geliştirir.
Gerçek Dünya Uygulama Başarısı	Benchmark testlerinde bahsedilen başarısının **gerçek kullanımda da doğrulandığı** gözlemlenmiştir. Web arayüzünde hızlı ve kaliteli web siteleri oluşturup yayına alabilmiş, mobil uyumluluğu ve doğrudan düzenleme özelliği göstermiştir.	Bazı rakiplerin (örn. Quent Coder) benchmark testlerindeki yüksek puanlarına rağmen, gerçek kullanımda bu kadar başarılı sonuçlar vermediği iddia edilmiştir.	GLM 4.5'in **test başarısının gerçek dünya uygulamalarında da istikrarlı bir şekilde performansa yansıdığı** teyit edilmiştir. **Doğrudan yayına alma, farklı cihazlarda test etme ve elementleri düzenleme gibi pratik yetenekler** sunarak kullanım kolaylığı sağlar.

Bu tablo, GLM 4.5'in genel dil modelleri, kodlama yetenekleri, mimarisi ve eğitim felsefesi açısından rakiplerine kıyasla sunduğu önemli avantajları özetlemektedir.

1.8 GLM 4.5 Mimarisi

GLM 4.5, Çinli yapay zeka şirketi Zhipu AI tarafından tanıtılan ve yapay zeka dünyasında **yeni nesil bir güç** olarak konumlandırılan, **tamamen açık kaynaklı ve ücretsiz** bir dil modelidir. Bu model, Claude gibi güçlü kapalı kaynak modeller de dahil olmak üzere birçok rakibini geride bırakarak dikkat çekmiştir.

GLM 4.5'i yapay zekada yeni nesil bir güç yapan temel özellikler ve yenilikler şunlardır:

- **Verimli Parametre Kullanımı ve MOI Mimarisi**:
  - GLM 4.5, toplamda çok sayıda parametreye (ana sürüm 355 milyar, Lite sürümü 106 milyar) sahip olsa da **yalnızca çok daha küçük bir kısmını aktif olarak kullanır** (ana sürümde 32 milyar, Lite sürümde 12 milyar). Bu sayede, **çok daha az bellek tüketirken, çok daha hızlı ve net sonuçlar** elde edebilir.
  - Model, her katmanda birçok uzman barındıran **Mixture of Experts (MoE) mimarisini** kullanır ve yalnızca göreve uygun uzmanları aktifleştirir. Bu yapı, modelin hem hızlı hem de verimli çalışmasını sağlar.
  - Bu yaklaşım, sektörün daha fazla parametre eklemek yerine **daha az veriyle daha zeki modeller geliştirme** eğilimini yansıtır ve GLM 4.5'i bu yeni neslin temsilcisi yapar.
- **Esnek Çalışma Modları**:
  - GLM 4.5, karmaşık muhakeme gerektiren çok adımlı problemler için **"düşünme modu"**nu, basit bilgi soruları veya hızlı yanıtlar için ise **"düşünmeden yanıt verme modu"**nu kullanabilir.
  - Bu esneklik, modelin **hem reaktif hem de proaktif görevlerde** etkin bir şekilde kullanılabilmesini sağlar. Model, düşünme modunda adım adım mantıksal akıl yürütme yeteneği sergiler.
- **Üstün Genel Performans ve Kapsamlı Yetenekler**:
  - GLM 4.5, **hem açık kaynak hem de kapalı kaynak modellerin çoğunu genel performansta geride bırakmıştır**. Özellikle **araç çağırma görevlerinde tüm modellerin önüne geçmiştir**, Claude modeliyle genel başarıda kafa kafaya ilerlese de.
  - OpenAI modelleri gibi rakiplerin aksine, GLM 4.5 **tüm farklı alanlardaki yetenekleri (kodlama, matematik, muhakeme vb.) tek bir modelde birleştirmeyi başarmıştır**.
  - **128.000 KTEXT uzunluğunda bağlama** sahip olması, modelin karmaşık görevlerde büyük miktarda veriyi tek seferde işlemesine olanak tanır. Web araştırmaları ve çok katmanlı kodlama senaryolarında pek çok modelden daha iyi sonuçlar üretebilir.
- **Full-Stack Proje Geliştirme Yeteneği**:
  - GLM 4.5, **oldukça iddialı bir kodlama performansına** sahiptir ve yapılan karşılaştırmalarda Kimy K2 ve Quantry Coder gibi güçlü modelleri geride bırakmıştır.
  - Modelin en dikkat çekici farklarından biri, **full-stack proje geliştirme**

yeteneğidir. Frontend'den backend'e, tasarımdan kullanıcı deneyimine kadar bir projenin tüm aşamalarını sıfırdan oluşturabilir.
    - **2D/3D oyunlar, web siteleri oluşturabilir**, belirli bir konuda araştırma yapıp doküman, slayt veya tanıtım materyalleri hazırlayabilir.
    - Modern teknolojilerle (Next.js, TypeScript, Tailwind CSS) uyumlu çalışır ve oluşturduğu projeleri **otomatik olarak yayına alabilir ve düzenlenebilir** kılar.
- **Gelişmiş Mimari ve Eğitim İnovasyonları**:
    - **Multi-Token Prediction (MTP)** ve **Speculative Decoding** gibi teknikler sayesinde, model aynı anda birden fazla token üreterek **çıktı üretimini ciddi oranda hızlandırır** ve kaliteden ödün vermez.
    - Eğitim süreci, **üç ana aşamadan (ön eğitim, ara eğitim, son eğitim)** oluşur. Bu aşamalar, modelin genel dil ve kod bilgisinden mantıksal muhakemeye, ardından görev bazlı düşünce yapısına kadar uzmanlaşmasını sağlar.
    - Reinforcement Learning (RL) için **büyük ölçekli ve ayrıştırılmış bir sistem** kurgulanmıştır. Veri üretimi ve eğitim işlemlerinin farklı donanımlarda yürütülmesi, GPU'ların daima aktif kalmasını sağlar ve eğitim sürecinin yavaşlamasını engeller. Bu, büyük dil modellerinde RL uygulamasını ölçeklenebilir ve verimli hale getiren önemli bir yeniliktir.
    - Son eğitim aşamasında **uzman eğitimi** (mantıksal akıl yürütme, ajan yetkinlikleri, genel dil görevleri) ve **birleşik eğitim** (self-distillation ile uzmanlıkların entegrasyonu) uygulanır. Bu sayede model, karmaşık görevleri planlama, araçları kullanma ve çözüm üretme yeteneklerini geliştirir.

Özetle, GLM 4.5, **açık kaynaklı ve ücretsiz oluşu, yenilikçi mimarisi (MoE, MTP, Speculative Decoding), verimli parametre kullanımı, esnek çalışma modları, kapsamlı yetenekleri (özellikle full-stack proje geliştirme) ve optimize edilmiş eğitim süreci** sayesinde yapay zeka alanında yeni nesil bir güç olarak öne çıkmaktadır. Benchmark testlerindeki başarısı, gerçek dünya uygulamalarında da istikrarlı bir şekilde doğrulanmıştır.