ISSN: 2146 - 9466
www.ijtase.net
International Journal of New Trends in Arts, Sports &Science Education – 2026, volume 15, issue 4

YAPAY ZEKÂ UYGULAMALARININ MATEMATİKSEL YARATICI PROBLEM ÇÖZME PERFORMANSLARININ İNCELENMESİ

INVESTIGATING THE MATHEMATICAL CREATIVE PROBLEM SOLVING PERFORMANCE OF ARTIFICIAL INTELLIGENCE APPLICATIONS

İmren AYDOĞDU
Matematik Öğretmeni, Milli Eğitim Bakanlığı, Ankara, Türkiye
ORCID iD: https://orcid.org/0000-0003-0253-4420
imreenay@gmail.com

Cenk KEŞAN
Prof. Dr., Dokuz Eylül University, Buca Faculty of Education, Buca, İzmir
ORCID iD: https://orcid.org/0000-0003-2629-8119
cenk.kesan@deu.edu.tr

Received: July 18, 2026 **Accepted:** September 20, 2026 **Published:** October 31, 2026

Suggested Citation:
Aydoğdu, İ., & Keşan, C. (2026). Yapay zekâ uygulamalarının matematiksel yaratıcı problem çözme performanslarının incelenmesi. *International Journal of New Trends in Arts, Sports & Science Education (IJTASE), 15*(4), 215-238.

Öz

Bu araştırmada, farklı teknolojik mimarilere sahip yapay zekâ (YZ) uygulamalarının rutin olmayan matematiksel yaratıcı problemler karşısındaki performanslarının; geliştirdikleri çözüm stratejileri, yaratıcı varsayımlar, çözümler arasındaki benzerlik/farklılıklar ve karşılaştıkları başarısızlık örüntüleri açısından derinlemesine incelenmesi amaçlanmıştır. Nitel araştırma metodolojisine dayalı çoklu durum çalışması deseninde yürütülen araştırmada; ChatGPT, Gemini, Claude, Copilot, DeepSeek, JuliusAI, Socratic, MathosAI, PhotoMath, Mathway, WolframAlpha ve Symbolab olmak üzere 12 farklı araç, veri toplama protokolü kapsamında test edilmiştir**.** Elde edilen nitel veriler, hiyerarşik kodlama şemaları aracılığıyla içerik analizine tabi tutulmuştur. Araştırma bulguları, uygulamaların matematiksel yaratıcı problemlere yaklaşımlarında tek tip bir performans sergilemediklerini; genel kısıtları saptama ve geleneksel temsilleri kullanma yönüyle benzeşirken, kavramsal derinlik, optimizasyon doğruluğu ve yaratıcı yeniden çerçeveleme boyutlarında radikal biçimde ayrıştıklarını göstermiştir. ChatGPT strateji çeşitliliği ve bilişsel akıcılıkta, Gemini kavramsal esneklik ve özgün varsayımlarda, JuliusAI ise algoritmik doğruluk ve kısıt yönetiminde üstün performans sergilemiştir. Buna karşın, geleneksel çözücüler yaratıcı problemlere yanıt üretemeyerek başarısızlığa uğramış; Copilot ve Socratic ise bilişsel üretim yapmalarına rağmen katı matematiksel kısıt ihlalleri, aritmetik sapmalar ve eksik doğrulama adımları göstermiştir. Sonuç olarak bu araştırma, yaratıcı matematiksel akıl yürütme süreçlerinde akıcılık, esneklik ve algoritmik doğruluğun tek bir yapay zekâ mimarisinde bütünüyle birleşemediğini; performansın problemin niteliği ile uygulamanın teknik altyapısına bağlı olarak dinamik bir değişkenlik gösterdiğini ortaya koymaktadır.

Anahtar Terimler: yapay zekâ, matematiksel yaratıcı problem çözme, büyük dil modelleri, çoklu durum çalışması, rutin olmayan problemler

Abstract

This study aims to investigate the performance of various artificial intelligence (AI) applications in solving non-routine mathematical creative problems, focusing on their solution strategies, creative assumptions, cross-case similarities/differences, and failure patterns. Designed as a qualitative multiple-case study, the research evaluated 12 AI tools—ChatGPT, Gemini, Claude, Copilot, DeepSeek, JuliusAI, Socratic, MathosAI, PhotoMath, Mathway, WolframAlpha, and Symbolab—under a data collection protocol. The qualitative data were analyzed using comparative content analysis based on hierarchical coding schemes. The findings revealed that AI applications do not exhibit uniform performance; while they share similarities in identifying basic boundaries and employing traditional representations, they diverge radically in conceptual depth, optimization precision, and creative reframing. ChatGPT excelled in strategy generation and cognitive fluency, Gemini dominated in conceptual flexibility and novel assumptions, and JuliusAI demonstrated superior algorithmic precision and constraint management. Conversely, traditional solvers failed to generate responses, while Copilot and Socratic suffered from structural failures, including constraint violations and arithmetic deviations, despite producing verbal solutions. Conclusively, this study demonstrates that fluency, flexibility, and precision do not simultaneously converge in a single AI architecture, and performance remains dynamically dependent on both the problem's nature and the underlying technical framework.

Keywords: artificial intelligence, mathematical creative problem solving, large language models, multiple-case study, non-routine problems

GİRİŞ

Yaşamı sürdürebilmenin gerekliliklerinden biri çağa uyum sağlamakken, çağımıza hızlı bir giriş yapan yapay zekâyı etkili biçimde kullanabilmek de giderek bir yeterlik haline gelmektedir. Yapay zekâyı kullanmayan ya da kullanamayan bireylerin sayısı, kullanan bireylerin sayısından henüz fazla olmakla birlikte (Liu & Wang, 2026) ilerleyen süreçlerde bu farkın hızla azalacağı düşünülmektedir. Büyük dil modellerinin (Large Language Models-LLM) matematiksel akıl yürütme yetenekleri son yıllarda önemli ölçüde gelişmiş olmakla birlikte, bu yeteneklerin farklı problem türlerine genellenmesi ve güvenilir biçimde sürdürülmesi hâlen önemli bir araştırma problemidir (Wang vd., 2026; Ahn vd., 2024; Mirzadeh vd., 2025) Günlük işler de dahil yaşamın birçok alanında kullanılan yapay zekâ eğitim ortamlarında da yaygınlaşmakta; özellikle öğrenciler tarafından ödev hazırlama, konu öğrenme ve problem çözme süreçlerinde sıklıkla tercih edilmektedir. Öğrenciler sorunun fotoğrafını çekerek ya da soruyu metin olarak yapay zekâ sistemine yöneltmekte ve kısa sürede yapay zekâdan çözüm elde edebilmektedir. Ancak bu çözümlerin doğruluğu, niteliği ve öğrencilerin matematiksel düşünme süreçlerini nasıl etkileyebileceği konusunda henüz yeterli bilimsel kanıt bulunmamaktadır.

Yapay zekâ (YZ) teknolojilerindeki hızlı gelişmeler, matematiksel problem çözme süreçlerinde kullanılan dijital araçların niteliğini de önemli ölçüde değiştirmiştir. Özellikle LLM'lerin yaygınlaşmasıyla birlikte yapay zekâ uygulamaları yalnızca sembolik işlemleri gerçekleştiren veya önceden tanımlanmış çözüm algoritmalarını uygulayan araçlar olmaktan çıkarak, doğal dilde verilen matematiksel problemlerin yorumlanması, çözüm stratejilerinin oluşturulması, gerekçelendirme yapılması ve alternatif çözüm yollarının önerilmesi gibi daha karmaşık süreçlerde kullanılmaya başlanmıştır. Bununla birlikte, matematiksel problem çözmede yapay zekâ performansının yalnızca ulaşılan sonucun doğruluğu üzerinden değerlendirilmesinin yeterli olmadığı; çözümün matematiksel ve pedagojik niteliğinin, süreç içerisindeki akıl yürütmenin ve farklı problem türlerine genellenebilirliğinin de dikkate alınması gerektiği Schorcht ve arkadaşları (2026) tarafından vurgulanmıştır.

Matematikte problemler, rutin problemler ve rutin olmayan problemler olmak üzere ikiye ayrılmıştır (Pólya, 1945). Rutin problemler sıklıkla karşılaşılan, çözüm yolları nerdeyse standartlaşmış olan problemlerdir. Rutin olmayan problemler ise sıradan bir çözüm yolu olmayan, çözüm için bazı üst düzey düşünme ve akıl yürütme becerilerini kullanmayı gerektiren problemlerdir. 21. yüzyılda matematik eğitiminde sorular sadece doğrudan hesaplama gerektiren değil, üst düzey düşünme, yaratıcılık ve problem çözme becerisinden beslenen soru tipi olarak karşımıza çıkmaktadır (Schoenfeld, 2016). Matematik problemleri için sayısı gittikçe artan uygulamalar cep telefonu ya da tabletler içinde yerlerini almaktadır. Buna karşın, bu uygulamaların yalnızca rutin ve algoritmik işlemler gerektiren sorularda değil, yaratıcı düşünmeyi, akıl yürütmeyi ve alternatif çözüm yolları geliştirmeyi gerektiren matematiksel yaratıcı problemlerde nasıl bir performans sergilediğine yönelik çalışmalar oldukça sınırlıdır. Oysa matematiksel yaratıcılık, farklı çözüm yolları geliştirebilme, özgün fikirler üretebilme ve mevcut bilgileri yeni durumlara uyarlayabilme gibi üst düzey bilişsel becerileri içermektedir. Runco ve Jaeger (2012), yaratıcılığı özgünlük ile etkililiğin birlikte gerçekleşmesi olarak ele alır; bu nedenle yalnızca yeni bir fikir üretmek, bu fikir işlevsel veya uygun değilse yaratıcılık için yeterli değildir. Silver (1997), matematiksel yaratıcılığın problem çözme ve problem kurma etkinlikleri yoluyla geliştirilebileceğini; özellikle akıcılık (fluency), esneklik (flexibility) ve özgünlük (originality) boyutlarının matematiksel etkinliklerde ortaya çıkabileceğini belirtmektedir. Bu nedenle yapay zekâ uygulamalarının yalnızca doğru sonuca ulaşıp ulaşmadığını değil, problemi nasıl yorumladığını, hangi varsayımları geliştirdiğini ve nasıl bir çözüm süreci izlediğini incelemek önem taşımaktadır.

Bir dizi algoritmik işlemin uygulanmasından öte orijinal düşünmeyi gerektiren bir eylem olarak tanımlanan (Kirişçi, 2021) yaratıcı problem çözme ise standart problem çözme ile yaratıcılığın birleşimidir. Yaratıcı problem çözme, sadece var olan bilginin uygulanmasından ibaret değil, problemi yeniden yapılandırmayı, alternatif fikirler üretmeyi ve bu fikirleri değerlendirmeyi içeren bütünsel bir süreçtir (Mumford vd., 2012). Yapay zekâ ile problem çözmenin günlük yaşamda sıradan

bir eyleme dönüşmeye başladığı günümüzde yapay zekânın sınırlarının ne kadar zorlanacağı da merak konusudur. Öğrencilerin sıklıkla kullandığı yapay zekânın öğrencileri doğru mu yanlış mı yönlendirdiği, yetersiz mi kaldığı incelenmelidir. Literatürdeki çalışmaların (Turmuzi vd., 2026) tek bir yapay zekâ sistemi veya belirli bir uygulama üzerine odaklanırken, farklı yapay zekâ uygulamalarının aynı problem karşısındaki performanslarını karşılaştıran bir çalışmaya ulusal literatürde rastlanmamıştır. Bu boşluk doğrultusunda yapılan bu araştırmanın amacı farklı yapay zekâ uygulamalarının matematiksel yaratıcı problemlere yönelik performanslarının incelenmesidir.

Bu araştırmada performans; yapay zekâ uygulamalarının matematiksel yaratıcı problemlere yanıt verebilme durumu, geliştirdikleri çözüm stratejileri, oluşturdukları yaratıcı varsayımlar ve ürettikleri çözümler arasındaki benzerlik ve farklılıklar çerçevesinde değerlendirilmiştir. Araştırmanın temel problemini "Farklı yapay zekâ uygulamalarının matematiksel yaratıcı problemlere yönelik performansları nasıldır?" sorusu oluşturmaktadır. Alt problemler ise şu şekilde belirlenmiştir:

1. Matematiksel yaratıcı problemlere yanıt veren yapay zekâ uygulamaları hangi çözüm stratejilerini ve yaratıcı varsayımları geliştirmektedir?
2. Yapay zekâ uygulamalarının matematiksel yaratıcı problemlere yönelik çözümleri arasındaki benzerlikler ve farklılıklar nelerdir?
3. Matematiksel yaratıcı problemlere yanıt vermeyen veya yanıt vermede başarısız olan yapay zekâ uygulamalarının ortak özellikleri nelerdir?

Belirlenen bu alt problemler doğrultusunda gerçekleştirilen bu araştırmada; farklı teknolojik mimarilere sahip yapay zekâ uygulamaları, matematiksel yaratıcı problemler üzerinde karşılaştırmalı olarak analiz edilmiştir. Bu yönüyle araştırma, yapay zekâ teknolojilerinin bilişsel ve algoritmik sınırlarının ortaya konulmasında ve eğitim ortamlarında hangi pedagojik amaçlarla işe koşulabileceğine dair ampirik kanıtlar sunması bakımından önem taşımaktadır. Ayrıca araştırmadan elde edilen bulguların; öğretmenlere, araştırmacılara ve eğitim teknolojisi geliştiricilerine yapay zekâ destekli matematik öğretiminin tasarımı ve planlanması süreçlerinde rehberlik edeceği öngörülmektedir. Belirtilen bu ampirik ve teorik hedeflere ulaşmak amacıyla araştırmada takip edilen yöntemsel tasarım, veri toplama ve analiz süreçleri takip eden bölümde ayrıntılı olarak ele alınmıştır.

YÖNTEM

Bu bölümde araştırmanın desenine, veri kaynaklarına, veri toplama araçlarına, veri toplama sürecine ve verilerin analizine gerekçeleriyle birlikte yer verilmiştir.

Araştırmanın Deseni

Bu araştırmada, yapay zekâ destekli farklı teknolojik mimarilerin rutin olmayan matematiksel yaratıcı problemler karşısındaki bilişsel stratejilerini, ontolojik varsayımlarını ve yapısal sınırlarını derinlemesine incelemek amacıyla nitel araştırma yöntemlerinden durum çalışması deseni benimsenmiştir. Creswell ve Poth (2018), nitel araştırmayı; sınırlandırılmış bir sistem veya olguya dair derinlemesine bir keşif süreci olarak tanımlarken, bu süreçte araştırmacının bizzat temel veri toplama aracı olarak konumlandığını vurgulamaktadır. Bu araştırmada da araştırmacılar; istem (prompt) protokollerinin tasarlanması, test süreçlerinin yönetilmesi ve elde edilen verilerin sistematik olarak kodlanması aşamalarını yürüterek temel veri toplama ve analiz aracı rolünü üstlenmiştir.

Durum çalışmaları, gerçek yaşam bağlamı içerisinde yer alan güncel bir olgunun, olgu ile bağlam arasındaki sınırların kesin olarak ayrıştırılamadığı durumlarda derinlemesine incelenmesini sağlayan esnek bir metodolojik tasarım sunmaktadır (Yin, 2018). Araştırmacılar, araştırma odağına bağlı olarak tek veya birden fazla bağımsız durumu eşzamanlı olarak inceleme serbestliğine sahiptir (Merriam, 1998; Stake, 1995). Bu doğrultuda ele alınan araştırmada incelenen olgu, yapay zekâ uygulamalarının matematiksel yaratıcı problemlere yönelik problem çözme ve modelleme performanslarıdır. Araştırmada incelenen sınırlandırılmış durumlar ise bu olguyu temsil eden 12 farklı yapay zekâ uygulamasıdır. Araştırmanın analiz birimi ise, her bir durumun (yapay zekânın) 5 yaratıcı probleme verdiği yazılı, sembolik ve algoritmik yanıtların içeriğidir.

Araştırma, Yin'in (2018) taksonomisine uygun olarak çoklu durum çalışması olarak desenlenmiştir. Çoklu durum desenleri, tekli durumlara kıyasla elde edilen ampirik kanıtların daha güçlü, açıklayıcı ve doğrulanabilir olmasını sağlayan tekrarlama (replikasyon) mantığına dayanmaktadır (Yin, 2018). Bu araştırmada çoklu durum tasarımının tercih edilmesinin nedeni; farklı bilişsel mimarilere sahip araçların benzer kısıtlar altındaki performanslarını karşılaştırmalı olarak analiz etmek, stratejik benzerlikleri ve mimari farklılıklardan kaynaklanan yapısal ayrışmaları sistematik olarak ortaya koymaktır. Araştırma sürecinde her bir yapay zekâ uygulaması öncelikle durum içi analiz ile kendi yapısal sınırlarında incelenmiş; ardından durumlar arası analiz matrisleri yardımıyla karşılaştırmalı olarak çözümlenmiştir (Schreier, 2012).

Veri Kaynakları

Araştırma kapsamında, matematik eğitiminde ve problem çözme süreçlerinde sıklıkla başvurulan yapay zekâ destekli toplam 12 farklı dijital araç incelemeye dâhil edilmiştir. Değerlendirmeye alınacak uygulamaların seçiminde; mobil uygulama mağazalarında (Google Play Store ve Apple App Store) "eğitim ve matematik" kategorilerinde üst sıralarda yer alma, yüksek kullanıcı puanına sahip olma, ücretsiz erişim imkânı sunma ve popüler kullanım düzeyleri temel ölçütler olarak belirlenmiştir. Araştırma bütçesinden bağımsız, genel öğrenci ve öğretmen deneyimini yansıtabilmek adına tüm uygulamaların ücretsiz üyelikleri kullanılmıştır.

OpenAI firması tarafından geliştirilen ChatGPT platformundaki *GPT-5.6 Luna* modeli, Google tarafından geliştirilen Gemini platformundaki *Gemini Flash* güncel üretim motoru, Anthropic firması tarafından geliştirilen *Claude Sonnet 5* modeli, Microsoft firması tarafından geliştirilen *Copilot* Cowork platformu**,** Hangzhou DeepSeek Artificial Intelligence Co., Ltd. firması tarafından geliştirilen DeepSeek platformundaki *DeepSeek-V4-Flash (Preview)* sürümü, MetaDigits.AI, Inc. (Y Combinator destekli) firması tarafından geliştirilen ve tescilli bir matematik tabanlı yapay zeka motoru olan *Mathos AI* (eski adıyla MathGPTPro) platformu, Caesar Labs, Inc. firması tarafından geliştirilen yapay zeka tabanlı veri bilimi platformu *Julius AI** platformu 1.2 Lite sürümü, Google LLC tarafından geliştirilen *Photomath* (v8.47.1) ve *Socratic* (v1.2)**,** Chegg, Inc. bünyesindeki *Mathway* (v6.10.0), Wolfram Group tarafından sunulan *Wolfram|Alpha* (v1.4.28) ile Course Hero altyapısındaki *Symbolab* (v12.0.2) uygulamalarının (20 Temmuz 2026 tarihli) belirtilen sürümleri araştırmada kullanılmıştır (**Julius AI* platformu; matematiksel ve analitik hesaplamaları gerçekleştirmek için Python 3 kod yürütme ortamını, grafik üretimleri için Matplotlib kütüphanesini ve metinsel yanıtların üretimi için ise ilgili tarihteki entegre doğal dil işleme motorunu eş zamanlı olarak kullanmıştır).

Söz konusu dijital araçlar, sahip oldukları bilişsel mimariler, arka plan çalışma mekanizmaları ve kullanıcıya sundukları işlevsel çözümler doğrultusunda yöntemsel olarak iki ayrı grupta taksonomik simetriyle sınıflandırılmıştır:

1. Genel Amaçlı Büyük Dil Modelleri (LLM Tabanlı Üretken YZ): ChatGPT, Google Gemini, Claude, Microsoft Copilot ve DeepSeek. Bu gruptaki modeller, geniş bağlam pencerelerine ve gelişmiş doğal dil işleme (NLP) yeteneğine sahip olan, problemleri anlamsal ve felsefi olarak yeniden çerçeveleyebilen çok yönlü sistemlerdir (Rahman vd., 2025; Marangoz, 2024).
2. Matematik Odaklı YZ Uygulamaları: JuliusAI, MathosAI ve Socratic: Bu araçlar, temel bir büyük dil modeli çekirdeğine sahip olmakla birlikte, matematiksel işlemleri gerçekleştirmek üzere özel prompt yapıları veya Python kod yürütme motorları (örn. JuliusAI) ile entegre edilmiş melez sistemlerdir. PhotoMath, Mathway, WolframAlpha ve Symbolab uygulamaları doğal dil işleme yeteneği sınırlı olan; kural tabanlı sembolik hesaplama motorları ve optik karakter tanıma (OCR) teknolojilerine dayanan geleneksel çözücülerden oluşmaktadır (Hacıoğlu Seven & Erümit, 2025; Marangoz, 2024).

Veri Toplama Araçları ve Problem Seti

Araştırmada matematiksel yaratıcı problemlerin belirlenmesinde Price'ın (2006) eserinden yararlanılmıştır. Söz konusu kaynak, yetenekli öğrenciler için yaratıcı ve zorlayıcı matematik etkinlikleri sunması nedeniyle tercih edilmiştir. Price (2006), uluslararası matematik eğitimi

literatüründe kabul görmüş, 21. yüzyıl becerilerini destekleyen (Szabo vd., 2020) ve George Pólya'nın (1945) sarsılmaz sezgisel (heuristik) problem çözme adımlarıyla doğrudan örtüşen standart bir kurgudur. Araştırmanın amacı doğrultusunda kitabın farklı bölümlerinde yer alan ve farklı matematiksel bağlamlar içeren beş problem seçilmiştir. Problem kaynağının tüm yapay zekâ uygulamaları için sabit tutulmasıyla, uygulamalar arasındaki performans farklılıklarının problem farklılığından etkilenmesinin önüne geçilmesi amaçlanmıştır. Problemler, araştırmacılar tarafından Türkçeye çevrilmeden özgün İngilizce metinleriyle tüm yapay zekâ uygulamalarına aynı biçimde sunulmuş ve uygulamalardan yanıtlarını Türkçe olarak üretmeleri istenmiştir. Böylece çeviri kaynaklı anlam farklılıklarının karşılaştırmayı etkilemesi önlenmeye çalışılmıştır. Seçilen problemler bir psikometrik ölçme aracı olarak kullanılmadığından, Price'ın (2006) problem setine ilişkin ayrıca geçerlik ve güvenirlik analizi gerçekleştirilmemiştir.

Seçilen problemlerin tamamı; formül ezberine dayanmayan, dairesel olmayan akışlar içeren, çoklu temsil (geometrik, cebirsel, algoritmik) ve Kapalı Dünya kısıtları altında özgün varsayımlar geliştirmeyi zorunlu kılan rutin dışı yaratıcı matematik problemleri niteliğindedir. Yapay zekâ modellerinin doğal dil işleme, semantik kısıt yönetimi ve kavramsal esnetme becerilerini doğrudan test edebilmek amacıyla, problemlerin orijinal metin yapıları, edebi kurguları ve tekerleme/hikaye formları aynen korunarak sisteme yöneltilmiştir.

Araştırmanın amacı, yapay zekâ uygulamalarının "Türkçe anlama ve Türkçe prompt yönetme" becerisini değil; "Matematiksel yaratıcı problem çözme kapasitesini" incelemektir. Girdinin (prompt) orijinal İngilizce dilinde verilmesi modellerin bilişsel performansını maksimize edip standartlaştırırken; yanıtın Türkçe istenmesi ise araştırmacının yerel literatürle (Kükey vd., 2019; Şahin & Yeldan, 2019) tam uyumlu, pürüzsüz bir Türkçe kodlama şeması yürütmesini sağlamıştır.

Veri Toplama Süreci

Araştırma verilerinin toplanması ve yapay zekâ uygulamalarının test edilmesi süreci, iki aşamalı bir protokol halinde yürütülmüştür:

- Birinci Aşama (Tepki ve Uyumluluk Analizi): Öncelikle sohbet tabanlı yapay zekâ uygulamalarının tamamına aynı standartlaştırılmış istem (prompt) sunulmuştur. İstemde (Şekil 1) araştırmanın amacı açıklanmış, matematiksel kurallara uygun yaratıcı ve alternatif çözüm yollarının üretilmesi beklenmiş, problemlerin özgün İngilizce metinleriyle gönderileceği ve yanıtların Türkçe verilmesi gerektiği belirtilmiştir. Yanıt üretimindeki olası değişkenliği standartlaştırmak amacıyla istem içerisinde sıcaklık değerinin 0,5 olarak kabul edilmesi belirtilmiştir. Doğal dil etkileşimine dayalı konuşma tabanlı uygulamalarda standart istem kullanılmış; matematiksel çözüm üretimini doğrudan görev olarak gerçekleştiren araçlarda ise ek bir istem kullanılmamıştır. Ardından 5 yaratıcı problem 12 YZ uygulamasına ayrı ayrı yöneltilmiştir. Uygulamaların sözel yönergeleri okuma, girdi karakter sınırını yönetme ve probleme rasyonel bir matematiksel yanıt üretip üretememe durumları gözlemlenmiştir. Bu aşamada doğrudan hata kodu veren veya görevi reddeden (yanıt vermeyen) sistemlerin "hata semptomları" kayıt altına alınarak analiz edilmiştir.

> *Bir akademik araştırma yürütüyorum ve yapay zekânın matematiksel problemlere getirdiği yaratıcı yaklaşımları, alternatif çözüm yollarını karşılaştırıyorum. Bu sohbet boyunca arka plandaki sıcaklık (temperature) parametreni 0.5 olarak kabul etmeni istiyorum. Sadece en bilindik yöntemi sunmakla kalma; mantıklı, matematiksel kurallara uygun ama aynı zamanda yaratıcı veya alternatif bakış açıları içeren çözüm yolları üret. Sorular orijinali bozulmasın diye İngilizce göndereceğim. Ama bana Türkçe yanıt ver. Hazırsan sorumu göndereceğim.*

Şekil 1. Araştırma başlangıç istemi (prompt)

- İkinci Aşama (Bilişsel Çözümleme): İlk aşamada sisteme uyum sağlayıp yaratıcı çözümler ve yanıtlar sunabilen 8 uygulama (Grup 1 ve Grup 2) üzerinde derinlemesine testler yürütülmüştür. Modellerin ürettiği tüm çözüm adımları, kod blokları ve sözel açıklamalar ham veri olarak dijital ortama aktarılmıştır.

Verilerin Analizi

Araştırmadan elde edilen nitel verilerin analizinde, sistematik ve güvenilir bir kodlama yapısı kurmak amacıyla nitel içerik analizi (qualitative content analysis) yöntemine başvurulmuştur (Cohen vd, 2007). Araştırmanın kavramsal çerçevesi ve kod rehberleri oluşturulurken Johnny Saldaña (2013) ve Margrit Schreier (2012) tarafından önerilen sistematik nitel kodlama aşamaları takip edilmiştir.

Analiz süreci şu ardışık hiyerarşik adımlarla tamamlanmıştır:

1. Açık Kodlama ve Durum İçi Analiz: Her bir yapay zekâ uygulamasının 5 soruya verdiği yanıtlar satır satır incelenerek, kullandıkları ampirik eylem adımları ve kavramsal kabuller "gözlemlenebilir kodlar (operasyonel göstergeler)" olarak tanımlanmıştır. Her bir YZ'nin kendi çözüm sistemi durum içi analizle kendi içinde çözümlenmiştir.
2. Kategorizasyon ve Aksiyel Kodlama: Elde edilen somut gösterge kodları, anlamsal ve kuramsal benzerliklerine göre bir araya getirilerek daha üst düzey, genellenebilir "Kategorilere (Stratejiler ve Varsayımlar)" dönüştürülmüştür.
3. Tematik Yapılandırma: Kategoriler, araştırmanın kuramsal yönelimini ve felsefi şemsiyesini çizen 7 ana tema altında birleştirilmiştir.
4. Durumlar Arası Analiz: Kodlama şemasının tamamlanmasının ardından, Genel Amaçlı LLM'ler (Grup 1) ile Matematik Odaklı YZ uygulamalarının (Grup 2) geliştirdikleri çözümler arasındaki benzerlikler ve farklılıklar karşılaştırmalı analize tabi tutulmuştur. Bu aşamada modellerin kısıt yönetimleri, algoritmik ve sezgisel yaklaşımları ile kısıt ihlali (hata) yapma durumları matrisler üzerinden değerlendirilerek nitel içerik analizi tamamlanmıştır.

Geçerlik ve Güvenirlik

Nitel araştırmalarda çalışmanın kalitesini ve bilimsel değerini ortaya koymak amacıyla geleneksel nicel ölçütler (iç-dış geçerlik ve güvenirlik) yerine; inandırıcılık, aktarılabilirlik, tutarlılık ve teyit edilebilirlik ölçütleri esas alınmıştır (Lincoln & Guba, 1985; Patton, 2015). Bu doğrultuda araştırmanın yöntemsel kalitesini artırmak amacıyla şu önlemler alınmıştır:

İnandırıcılık (İç Geçerlik): Araştırmada veri toplama aracı olarak kullanılan yaratıcı matematik problemlerinin Price (2006) tarafından geliştirilen standart setten seçilmesi ve orijinal yapılarının korunması yapı geçerliğini desteklemiştir. Ayrıca, elde edilen nitel kodlar ve tematik çerçeve, matematik eğitimi ve eğitim teknolojileri alanında uzman bir öğretim üyesinin ve matematik eğitimi alanında doktorasını tamamlamış bir matematik öğretmeninin incelemesine sunulmuş; uzman değerlendirmeleri doğrultusunda analizlere nihai şekli verilmiştir.

Aktarılabilirlik (Dış Geçerlik / Genellenebilirlik): Bu kapsamda, çalışmada kullanılan 12 yapay zekâ uygulamasının teknik mimarileri, prompt (istem) protokolleri, veri toplama aşamaları ve içerik analizi adımları ayrıntılı olarak raporlanmıştır.

Tutarlılık ve Teyit Edilebilirlik (Güvenirlik ve Nesnellik): Araştırmanın tutarlılığını test etmek amacıyla kodlayıcılar arası güvenirlik (inter-coder reliability) analizi gerçekleştirilmiştir. Yapay zekâ uygulamalarının ürettiği yanıtlardan rastgele seçilen %25'lik bir veri seti, araştırmacılardan bağımsız bir uzman tarafından yeniden kodlanmıştır. Kodlayıcılar arasındaki uyum yüzdesi, Miles ve Huberman'ın (1994) güvenirlik formülüne göre hesaplanmış ve %87 düzeyinde yüksek bir tutarlılık elde edilmiştir.

Araştırmanın Etik Süreçleri

Bu araştırmada insan hakları, veri gizliliği veya katılımcı rızası gibi etik kurul izni gerektiren unsurlar bulunmamaktadır. Veri kaynağı olarak tamamen halka açık ve dijital olarak erişilebilir yapay zekâ uygulamalarının algoritmik yanıtları kullanılmıştır. Ancak bilimsel dürüstlük ve etik kurallar uyarınca, yapay zekâ yanıtlarının tamamı orijinal halleriyle kaydedilmiş, alıntılarda herhangi bir müdahale yapılmamış, intihal denetimi gerçekleştirilmiş ve çalışmada kullanılan tüm akademik kaynaklara usulüne uygun olarak atıf yapılmıştır.

Metodolojik tasarımı, veri toplama protokolleri, etik sınırları ve analiz süreçleri ayrıntılı olarak yapılandırılan bu araştırmada; farklı bilişsel mimarilere sahip yapay zekâ uygulamalarının yaratıcı problemlere verdikleri yanıtlar çözümlenmiştir. Bu doğrultuda elde edilen tüm bulgular, geliştirilen hiyerarşik kodlama şemaları, bütünleşik frekans-kod dağılımları ve doğrudan metin alıntıları eşliğinde araştırmanın alt problemleriyle uyumlu bir biçimde takip eden bölümde sunulmaktadır.

BULGULAR

Bu bölümde araştırmanın üç alt problemi doğrultusunda elde edilen tüm bulgular, nitel içerik analizi, frekans-yüzde tabloları ve yöntemsel karşılaştırmalar sunulmuştur.

Birinci Alt Probleme İlişkin Bulgular: Matematiksel Yaratıcı Problemlerde Kullanılan Çözüm Stratejileri ve Yaratıcı Varsayımlar

Araştırmanın birinci alt problemi olan "Matematiksel yaratıcı problemlere yanıt veren yapay zekâ uygulamaları hangi çözüm stratejilerini ve yaratıcı varsayımları geliştirmektedir?" sorusunu yanıtlamak için Yaratıcı Problem Çözme Stratejileri ve Yaratıcı Varsayımlar olmak üzere iki bağımsız içerik analizi oluşturulmuştur.

Yaratıcı Problem Çözme Stratejileri

Yapay zekâ uygulamalarının matematiksel yaratıcı problemlerde kullandıkları stratejiler ve bu stratejilerin kodlarının temalara göre dağılımı Tablo 1'de sunulmuştur.

Tablo 1. Yaratıcı problem çözme stratejileri kodlama rubriği

Temalar	Kategoriler	Kodlar	Kodun Açıklaması
Tema 1: Çok Boyutlu Matematiksel Modelleme ve Temsil	K 1.1: Geometrik ve Uzamsal Modelleme Stratejisi	Kod 1.1.1: Geometrik Form İndirgemesi (Daire/Küre)	Karmaşık yapıları dairesel/hacimsel uygulamalara indirgeyerek işlem kolaylığı sağlama adımı.
		Kod 1.1.2: Koordinat ve Uzamsal Konumlandırma	Elemanları iki boyutlu (x,y) kartezyen düzlemine yerleştirerek haritalama adımı.
		Kod 1.1.3: Dinamik Mekanizma Tasarımı	Sabit nokta yerine ray/zip-line sistemiyle hareketli Minkowski/ Stadyum alanı üretme adımı.
	K 1.2: Sayı Teorisi ve Oransal Akıl Yürütme Stratejisi	Kod 1.2.1: Doğrusal Aritmetik Toplama	Verileri derinlemesine sorgulamadan doğrusal uç uca ekleme veya basit toplama adımı.
		Kod 1.2.2: Asal Çarpan ve Faktör Analizi	Olasılık veya sayma problemlerini asal çarpan yapısı üzerinden matematiksel çözümleme adımı.
		Kod 1.2.3: Fermi Tahmini ve Oranlama	Kesin veri yokluğunda rasyonel varsayımlarla yaklaşık sayısal sınır üretme adımı.
		Kod 1.2.4: Boyutsal ve Biyomekanik Ölçekleme	Boyut artışının kas gücü ve kütle oranları (L^2 vs L^3) üzerindeki etkilerini hesaplama adımı.
Tema 2: Algoritmik Optimizasyon ve Karar Verme	K 2.1: Sınırlandırılmış Kombinatorik Bölümleme Stratejisi	Kod 2.1.1: 0-1 Tamsayılı/Lineer Programlama	Karar değişkenleriyle Excel Solver optimizasyon modelleri kurma adımı.
		Kod 2.1.2: Sezgisel Dağıtım	Sıralama, zikzak veya hafif kısıta göre hızlı eşleştirme yapma adımı.
		Kod 2.1.3: Tam Arama ve Liste Çıkarma	Olası tüm kombinasyonları sistemli olarak listeleme ve kısıtlara göre eleme adımı.
	K 2.2: Uzamsal Rota ve Grafik Ağları Stratejisi	Kod 2.2.1: Euler Yolu ve Rota Planlama	Engelli yüzeylerde tek geçişli veya en kısa yollu grafik teorisi rotaları çizme adımı.
		Kod 2.2.2: Voronoi Bölüntüsü ve Hücreleme	Alanı belirli odak noktalarına göre en yakın mesafe hücrelerine ayırma adımı.
Tema 3: Tasarımsal ve Süreçsel Yapılandırma	K 3.1: İzomorfik Tasarım ve Dinamik Zincirleme Stratejisi	Kod 3.1.1: İzomorfik Görev Tasarımı	Bilişsel yükü aynı, parametreleri farklı eşdeğer sorular türetme adımı.
		Kod 3.1.2: Süreç Zincirleme	Bir problemin çıktısının bir sonraki problemin girdisi olduğu ardışık akış tasarlama adımı.

Tablo 1 incelendiğinde, yapay zekâ uygulamalarının matematiksel yaratıcı problemlere yaklaşımlarının üç temel kuramsal tema altında kümelendiği görülmektedir. İlk tema olan Çok Boyutlu Matematiksel Modelleme ve Temsil, uygulamaların problemleri sadece soyut simgeler olarak görmeyip, onları üç boyutlu fiziksel kütle korunumuna (Kod 1.1.1), uzamsal koordinat düzlemlerine (Kod 1.1.2) ve biyomekanik ölçekleme sınırlarına (Kod 1.2.4) büründürdüklerini göstermektedir.

Algoritmik Optimizasyon ve Karar Verme teması kapsamında geliştirilen stratejiler (K 2.1 ve K 2.2), uygulamaların kombinatorik kısıt durumlarını aşmak amacıyla yöneylem araştırması standartlarında doğrusal tamsayı programlama karar modelleri (Kod 2.1.1) ve sezgisel dağıtım algoritmaları (Kod 2.1.2) işe koştuklarını kanıtlamaktadır. Son olarak, Tasarımsal ve Süreçsel Yapılandırma teması, yapay zekâların matematiksel süreçleri dinamik bayrak yarışı zincirlerine (Kod 3.1.2) ve bilişsel yük eşdeğerliğine (Kod 3.1.1) göre yeniden tasarlayarak pedagojik ve metodolojik birer içerik üreticisi rolü üstlenebildiklerini doğrulamaktadır.

Aşağıda her soru için YZ uygulamalarının kullandığı stratejilerin kodlarına ve frekanslarına yer verilmiştir.

Tablo 2. Yapay zekâ uygulamalarının sorulara göre yaratıcı problem çözme stratejisi kodlarının dağılımı ve frekansları

Soru / YZ Modeli	ChatGPT	Gemini	Claude	Copilot	DeepSeek	JuliusAI	Socratic	Sütun Toplamı (f)
Soru 1 (Bir Şiir)	f=7 (1.1.1, 1.2.1, 1.2.3, 1.2.4, 2.2.1)	f=4 (1.2.1, 1.2.4)	f=3 (1.1.1, 1.2.4)	f=2 (1.1.1, 1.2.1)	f=6 (1.1.1,1.2.3, 1.2.4)	f=2 (1.1.1, 1.2.1, 1.2.3)	f=2 (1.1.1, 1.2.1)	28 (%16.0)
Soru 2 (Olası mı Yoksa?)	f=4 (1.2.2, 2.1.3)	f=4 (1.2.2)	f=3 (1.2.2)	f=2 (2.1.3)	f=4 (1.2.2)	f=3 (1.2.2)	f=1 (2.1.3)	23 (%13.1)
Soru 3 (Bahçe Matematiği)	f=10 (1.1.2,1.1.3, 2.2.1, 2.2.2)	f=4 (1.1.3, 2.2.1)	f=3 (1.1.3)	f=3 (1.1.1, 1.1.2)	f=4 (1.1.2, 2.2.1)	f=3 (1.1.1, 1.1.2)	f=3 (1.1.1)	34 (%19.4)
Soru 4 (Halat Çekme Yarışı	f=7 (2.1.1, 2.1.2, 2.1.3)	f=5 (2.1.1, 2.1.2)	f=3 (2.1.2)	f=2 (2.1.2)	f=3 (2.1.1, 2.1.2)	f=4 (2.1.1, 2.1.2)	f=2 (2.1.2)	29 (%16.6)
Soru 5 (Matematik Olimpiyatları)	f=8 (3.1.1, 3.1.2)	f=9 (3.1.1, 3.1.2)	f=5 (3.1.1, 3.1.2)	f=5 (3.1.2)	f=8 (3.1.1, 3.1.2)	f=10 (3.1.2)	f=6 (3.1.2)	61 (%34.9)
Toplam Frekans (f)	36	26	18	14	25	22	14	175
Grup Payı (%)	%20.6	%14.9	%10.3	%8.0	%14.3	%12.6	%8.0	%100.0

YZ uygulamalarının matematiksel yaratıcı problemlerde soru bazında kullandığı strateji kodlarına ilişkin Tablo 2 incelendiğinde, fikir üretkenliği ve strateji çeşitliliği (akıcılık/esneklik) boyutlarında uygulamalar arasında belirgin bir hiyerarşi olduğu gözlenmektedir. Toplamda üretilen 175 strateji kodunun %20.6'sını (f=36) tek başına gerçekleştiren ChatGPT, bu boyutta en üretken uygulama olarak konumlanmıştır; onu sırasıyla Gemini (%14.9, f=26) ve DeepSeek (%14.3, f=25) izlemektedir. Copilot (%8.0, f=14) ve Socratic (%8.0, f=14) ise en dar strateji repertuarı sunan uygulamalar olarak son sırada yer almışlardır. Soru bazlı dağılımlar değerlendirildiğinde; açık uçlu ve tasarımsal süreçler gerektiren Soru 5'in (Matematik Olimpiyatları) toplam 61 kodla (%34.9) en yüksek stratejik çeşitliliği tetiklediği; buna karşın daha yapılandırılmış bir olasılık problemi olan Soru 2'nin ise en sınırlı (f=23, %13.1) strateji çeşitliliğinde kaldığı saptanmıştır. Özellikle ChatGPT'nin Soru 3'teki (Bahçenin Matematiği Problemi) 10 farklı strateji hamlesi (%29.4 pay), uygulamanın uzamsal optimizasyon görevlerindeki ıraksak düşünme kapasitesinin kayda değer bir göstergesidir.
Yapay zekâ uygulamalarının yaratıcı matematik problemlerinde kullandıkları stratejilerini uygulama biçimlerine ilişkin örnekler şu şekildedir:

- Claude**,** Soru 1'deki dev adamın boyut artışını ve kas kütlesi oranını *Sayı Teorisi ve Oransal Akıl Yürütme Stratejisi (Boyutsal Ölçekleme Göstergesi)* doğrultusunda şu biyomekanik eylem adımıyla kanıtlamıştır: *"Adam k=1920 kat büyürken, kemik/kas kesit alanı k² (~3,7 milyon kat)*

artıyor, ama ağırlığı k³ (~7 milyar kat) artıyor. Yani gücünün ağırlığına oranı k kadar (~1920 kat) düşüyor."

- Gemini ise Soru 3'teki uzamsal modelleme kısıtını, keçiyi sabit bir kazığa bağlayan geleneksel daire yaklaşımlarının ötesine taşıyarak, aks boyunca hareket eden çelik bir tel sistemiyle Geometrik ve Uzamsal Modelleme Stratejisi (Dinamik Mekanizma Göstergesi) şeklinde tasarlamış ve otlama alanını "Stadyum Geometrisi" olarak formüle ederek tasarımı %90 oranında verimli kılmıştır: "Matematiksel Modeli: Keçinin taranan alanı bir daire değil, 'Minkowski Toplamı' veya dairesel uçlu bir 'Stadyum' şeklini alır. Keçi hat boyunca hareket ettikçe dikdörtgen çim alanın %90'ından fazlasını otlar."
- Algoritmik optimizasyon süreçlerinde ise Gemini, Soru 4'teki kombinatorik dengeleri binary karar değişkenleriyle K 2.1: Sınırlandırılmış Kombinatorik Bölümleme Stratejisi üzerinden şu şekilde kurmuştur: "Değişkenler: xi, yi ∈ {0,1} (xi=1 ⇒ i kişisi 1. takımdadır). Hedef Fonksiyonu: $|\sum wi\ xi - \sum wi\ yi| \rightarrow min$; Kısıtlar: $\sum xi = \sum yi = k$, $\sum wi\ xi \leq 500$, $\sum wi\ yi \leq 500$."
- Son olarak, Soru 5'teki tasarımsal kurguda Claude, bayrak yarışındaki el değiştirmeyi matematiksel girdilerin ardışıklığıyla İzomorfik Tasarım ve Dinamik Zincirleme Stratejisi olarak şu şekilde yapılandırmıştır: "1. koşucu bir denklemi çözüp x = 7 buluyor → 2. koşucu bu 7'yi kullanarak '7x − 3 = 4y + 10' gibi bir denklemi çözüyor ve y buluyor → 3. koşucu bulunan y değerini bir dikdörtgenin kenar uzunluğu kabul edip alanını hesaplıyor..."

Yaratıcı Varsayımlar

Yapay zekâ uygulamalarının problem kısıtlarını esnetmek, ontolojik çelişkileri saptamak veya problemlere yeni bağlamsal sınırlar çizmek amacıyla geliştirdikleri asıl varsayımların (Kategoriler) ve bu varsayımların eylemsel göstergelerinin (Kodlar) üst temalara göre dağılımı Tablo 3'te sunulmaktadır.

Tablo 3. Yaratıcı varsayımlar kodlama rubriği

Temalar	Kategoriler	Kodlar	Kodun Açıklaması
Tema 1: Fiziksel Sınırlar ve Doğal Sistem Kabulleri	K 1.1: Biyomekanik ve Jeodezik Sınırları Sorgulama Varsayımı	Kod 1.1.1: Biyomekanik Çöküş ve Ezilme Kabulü	Dev adamın Kare- Küp Yasası uyarınca kendi ağırlığı altında ezileceği varsayımı.
		Kod 1.1.2: Jeodezik ve Akışkan Dinamiği Sınırı Kabulü	Dev su kütlesinin sığmaması veya Coriolis kuvvetiyle tsunamiler oluşturması varsayımı.
	K 1.2: Semantik ve Sistemik Metaforları Entegre Etme Varsayımı	Kod 1.2.1: Ekolojik-Sistemik Kriz Metaforu	Şiirdeki unsurların ekolojik bir kriz (İnsanlık ve Doğa savaşı) sembolü olduğu varsayımı.
Tema 2: Matematiksel Sınır ve Tanımların Esnetilmesi	K 2.1: Teorik Olasılık ve Sonsuzluk Sınırlarını Zorlama Varsayımı	Kod 2.1.1: Kesir Serbestliği ve Sonsuz Olay Kabulü	Olasılıkların birim kesir olmasının şart olmadığı durumlarda teleskopik rasyonel serilerle sonsuz bağımsız olay varsayımı.
		Kod 2.1.2: Teorik-Pratik Araç Ayrımı Kabulü	Matematiksel olarak sonsuz olay üretilebilse de pratik araçların (zar, para vb.) sınır oluşturacağı varsayımı.
	K 2.2: Matematiksel Kavramların Yeniden Tanımlanması Varsayımı	Kod 2.2.1: Spor-Matematik Analojisi Kabulü	Atletizm branşlarındaki "mesafe" veya "ağırlık" kavramının "ispat derinliği" veya "genelleme derecesi" olduğu varsayımı.
		Kod 2.2.2: Topolojik Teklik Kabulü	"Bir olmak" eyleminin fiziksel birleşme değil, küme boyutunun n=1 düzeyine indirgenmesi olduğu varsayımı.
Tema 3: Problem Kısıtlarının Yapısal Sorgulanması	K 3.1: İmkânsızlık ve Yapısal Çelişki Keşfi Varsayımları	Kod 3.1.1: Yedek Oyuncu / Dışlama Zorunluluğu Kabulü	Toplam kütle sınırı nedeniyle 12 kişinin tamamının yarışmasının imkânsız olduğunu saptayıp yedeklerin kalması gerektiği varsayımı.
		Kod 3.1.2: Topolojik Delikli Yüzey Kabulü	Bahçedeki engeller nedeniyle çim alanın delikli olduğunu ve makinenin çift kesim yapmasının kaçınılmaz olduğu varsayımı.

Tablo 3'te yer alan kodlama şeması, yapay zekâ uygulamalarının matematiksel kısıtları körü körüne işleme koymak yerine, problemin ontolojik sınırlarını derinlemesine sorgulayan birer "bilişsel süzgeç" geliştirdiklerini ortaya koymaktadır. Fiziksel Sınırlar ve Doğal Sistem Kabulleri teması, YZ'lerin

şiirsel veya hayali kurgulardaki fiziksel imkânsızlıkları saptayarak Kare-Küp Yasası ve Coriolis Kuvveti gibi bilimsel bariyerleri birer "varsayım" olarak probleme entegre ettiklerini göstermektedir. Matematiksel Sınır ve Tanımların Esnetilmesi teması, YZ'lerin disiplinler arası metafor kurma becerilerini temsil etmektedir; özellikle atletizmdeki "mesafe" kavramının "ispat derinliği ve genelleme derecesi" olarak, tekerlemedeki "birleşme" eyleminin ise topolojik küme indirgemesi olarak yeniden tanımlanması bu esnekliğin en somut kanıtıdır. Son olarak, Problem Kısıtlarının Yapısal Sorgulanması teması, YZ'lerin problemin sınır şartlarındaki içsel çelişkileri (Soru 4'teki toplam ağırlık aşımı veya Soru 3'teki engelli alan topolojisi) saptayıp, zorunlu yedek oyuncu veya topolojik delikli yüzey kabullerini üreterek "Kapalı Dünya" prensibini (Şahin & Yeldan, 2019) başarıyla çalıştırdıklarını doğrulamaktadır.

Aşağıda her soru için YZ uygulamalarının sorulara göre yaratıcı varsayımlara ilişkin kodlara ve frekanslarına yer verilmiştir.

Tablo 4. Yapay zekâ uygulamalarının sorulara göre yaratıcı varsayım kodlarının dağılım ve frekansı

Soru / YZ Modeli	ChatGPT	Gemini	Claude	Copilot	DeepSeek	JuliusAI	Socratic	MathosAI	Sütun Toplamı (f)
Soru 1 (Bir Şiir)	f=3 (1.1.1)	f=2 (1.2.1, 2.2.2)	f=2 (1.1.1, 1.1.2)	f=2 (1.1.2)	f=3 (1.1.1, 1.1.2)	f=1 (1.1.1)	f=1 (1.1.1)	f=1 (1.1.1)	15 (%19.7)
Soru 2 (Olası mı Yoksa?)	f=2 (2.1.1)	f=2 (2.1.1)	f=2 (2.1.2)	f=1 (2.1.2)	f=2 (2.1.1)	f=1 (2.1.1)	f=1 (2.1.2))	f=1 (2.1.2)	12 (%15.8)
Soru 3 (Bahçe Matematiği)	f=4 (3.1.2)	f=3 (3.1.2)	f=2 (3.1.2)	f=1 (3.1.2)	f=2 (3.1.2)	f=2 (3.1.2)	f=1 (3.1.2)	f=1 (3.1.2)	16 (%21.1)
Soru 4 (Halat Çekme Yarışı	f=2 (3.1.1)	f=2 (3.1.1)	f=2 (3.1.1)	f=1 (3.1.1)	f=2 (3.1.1)	f=2 (3.1.1)	f=1 (3.1.1)	f=1 (3.1.1)	13 (%17.1)
Soru 5 (Matematik Olimpiyatları)	f=3 (2.2.1)	f=3 (2.2.1)	f=4 (2.2.1)	f=1 (2.2.1)	f=3 (2.2.1)	f=2 (2.2.1)	f=2 (2.2.1)	f=2 (2.2.1)	20 (%26.3)
Toplam Frekans (f)	14	12	12	6	12	8	6	6	76
Grup Payı (%)	%18.4	%15.8	%15.8	%7.9	%15.8	%10.5	%7.9	%7.9	%100.0

Yapay zekâ uygulamalarının yaratıcı varsayım üretkenliğini gösteren Tablo 4 incelendiğinde, toplam 76 varsayım kodunun dağılımında ChatGPT (%18.4, f=14) liderliğini korurken, Gemini (%15.8, f=12), Claude (%15.8, f=12) ve DeepSeek (%15.8, f=12) uygulamalarının de üst düzey bir bilişsel sorgulama performansı sergilediği görülmektedir. Copilot (%7.9, f=6), Socratic (%7.9, f=6) ve MathosAI (%7.9, f=6) uygulamaları ise problemleri çoğunlukla verilen ilk kısıtlar dahilinde doğrusal olarak çözmeye çalışmış, ontolojik sınırları zorlayan alternatif kabuller geliştirmekte yetersiz kalmışlardır. Soru bazında yapılan değerlendirmede, olimpiyat oyunlarının matematiksel kurgusunu içeren Soru 5'in (f=20, %26.3) en yüksek varsayım yoğunluğuna sahip olduğu saptanmıştır. Özellikle Claude'un Soru 5'teki 4 farklı kavramsal esnetme hamlesi (%20.0 pay), uygulamanın soyut kavramları başka disiplinlere başarıyla tercüme edebilen özgün zihinsel esnekliğini doğrulamaktadır. Soru 3 (f=16) ve Soru 1 (f=15) ise sırasıyla topolojik ve biyomekanik çelişkilerin saptanması yoluyla en yüksek ikinci ve üçüncü varsayım üreten alanlar olmuştur. Bu bulgular, YZ uygulamalarının yaratıcı problem çözme süreçlerinde sadece "çözücü" değil, aynı zamanda problemin sınır şartlarını test eden birer "analist" gibi davranabildiklerini sayısal olarak ortaya koymaktadır.

Yapay zekâ uygulamalarının yaratıcı matematik problemlerinde ürettikleri varsayımlara ilişkin örnekler şu şekildedir:

- Gemini, Soru 1'deki tekerleme metnini nesnel boyutlardan çıkarıp Semantik ve Sistemik Metaforları Entegre Etme (Ekolojik-Sistemik Kriz Göstergesi) çerçevesinde şu şekilde yorumlamıştır: "İnsanlığın (Büyük İnsan) elindeki teknoloji ve sanayi ile (Büyük Balta) doğayı (Büyük Ağaç) yok

etmesi ve bunun sonucunda küresel ısınma/iklim kriziyle tüm karaların sular altında kalması (Büyük bir şıp-şıp) anlatılmaktadır."

• Bahçe kısıtlarını değerlendirirken Gemini, bahçedeki engelleri (gölet, ağaçlar) biçme makinesinin çift geçiş yapmasını kaçınılmaz kılan topolojik bir kısıt olarak İmkânsızlık ve Yapısal Çelişki Keşfi (Topolojik Delikli Yüzey Göstergesi) ile şu şekilde tanımlamıştır: "Bahçede gölet ve ağaç gibi engeller bulunduğu için çim alan 'basit bağlantılı' değildir; içinde delikler olan bir yüzeydir. Makine engellerin etrafından geçerken dairesel dönüşler yapmak zorundadır... en az bir noktadan tekrar geçmek zorundadır."

• Olimpiyat kurgusunda ise Claude, atletizmdeki fiziksel gülle atma mesafesini matematiksel ispat derinliği ve genelleme seviyesi olarak Matematiksel Kavramların Yeniden Tanımlanması (Spor-Matematik Analojisi Göstergesi) kapsamında şu şekilde kavramsallaştırmıştır: "İspat Gülle Atışı: Basit bir önermeyi verilen sürede olabildiğince genelleyerek kanıtlayın. Genelleme derecesi (n=1 için mi, tüm pozitif tamsayılar için mi, tüm tamsayılar için mi?) = fırlatma mesafesi."

İkinci Alt Probleme İlişkin Bulgular: Yapay Zekâ Uygulamalarının Çözümleri Arasındaki Benzerlikler ve Farklılıklar

Araştırmanın ikinci alt problemi olan "Yapay zekâ uygulamalarının matematiksel yaratıcı problemlere yönelik çözümleri arasındaki benzerlikler ve farklılıklar nelerdir?" sorusu YZ uygulamalarının bilişsel süreçleri karşılaştırılarak analiz edilmiştir.

Benzerlikler

Yapay zekâ uygulamalarının matematiksel yaratıcı problemlere yönelik çözümleri arasındaki benzerlikler dört başlık altında ele alınmıştır.

Doğrusal Aritmetik Modelleme Ortak Başlangıcı: YZ uygulamalarının tümü, açık uçlu Soru 1'e yaklaşırken ilk aşamada dünyadaki tüm suların veya insanların kütle ve boyutlarını derinlemesine sorgulamadan doğrusal olarak uç uca ekleme/toplama stratejisini ortaklaşa sergilemişlerdir. Dünyadaki tüm suların toplanması, insanların boylarının üst üste eklenmesi (13.6 milyon km) veya ağaçların uç uca dizilmesi (30 milyar km) gibi doğrusal ve fiziksel gerçekçiliği düşük olan ilk tasarımlar, tüm uygulamaların ortak bilişsel hareket noktası olmuştur.

Geleneksel Matematiksel Temsillerin Ortak Seçimi: Tüm uygulamalar, Olası mı Yoksa Başka Bir Şey mi? (Soru 2) ve Bahçenin Matematiği (Soru 3) problemlerinde benzer standart somut araçları işe kullanmışlardır. Olasılık çiftlerini gerçekçi senaryolarla açıklarken tüm uygulamalar bilye torbası gibi kolaycı kurgulardan kaçınarak madeni para, standart altı yüzlü zarlar, takvim ayları ve iskambil kartları gibi geleneksel olasılık nesnelerini ortaklaşa kullanmışlardır. Geometrik çiçeklik tasarımlarında ise dairesel formların yarıçaplarını (≈1.13m) ve üçgen formların taban-yükseklik kombinasyonlarını hesaplarken tüm uygulamalar aynı temel geometrik formülleri işletmiştir.

Problem Kısıtlarında Çelişki Saptama Başarısı: Yanıt veren tüm YZ uygulamaları, halat çekme yarışı problemindeki (Soru 4) 12 kişinin toplam ağırlığının (1056 kg) takımların toplam üst ağırlık sınırını (1000 kg) aştığını saptayarak, 12 kişinin aynı anda yarışmasının matematiksel olarak imkânsız olduğunu ortaklaşa keşfetmişlerdir. Bu doğrultuda tüm uygulamalar, oyunun oynanabilmesi için en az bir veya iki oyuncunun yedek bırakılması gerektiği varsayımını zorunlu olarak benimsemiştir.

Standart Disiplinler arası Eşleme Tercihi: Soru 5'teki olimpiyat kurgusunda, tüm uygulamalar (özellikle ChatGPT, Gemini, MathosAI ve JuliusAI) atletizm branşları ile matematik alanları arasında benzer analojiler kurmuş; sürat koşularını hızlı zihinsel aritmetikle, maratonu ise teorem kanıtlama süreçleriyle eşleştirmişlerdir. Tüm uygulamalarda "100m depar yarışı" zihinden hızlı aritmetik işlemle, "400m engelli" cebirsel denklem çözmeyle, "maraton" ise teorem ispatı veya uzun soluklu veri analiziyle eşleştirilmiştir.

Farklılıklar

Yapay zekâ uygulamalarının matematiksel yaratıcı problemlere yönelik çözümleri arasındaki farklılıklar dört başlık altında ele alınmıştır.

Bilişsel Derinlik ve Fiziksel Limitleri Sorgulama Düzeyi: Uygulamalar arasındaki en belirgin farklılık, kuramsal fizik kurallarını ve ontolojik sınırları probleme entegre etme derinliğinde ortaya çıkmaktadır:

- Claude, Gemini ve DeepSeek; Soru 1'deki dev adamın büyümesini Kare-Küp Yasası ile ele alarak, dev adamın gücünün kesit alanıyla (L^2), kütlesinin ise hacmiyle (L^3) artacağını saptamış ve devin kendi ağırlığı altında ezileceğini ispatlamıştır. DeepSeek ek olarak, dev su kütlesinin dönmesinden kaynaklanan Coriolis kuvveti ve tsunamik etkileri hesaba katmıştır. Gemini ise tekerlemeyi ekolojik bir çevre krizi metaforu olarak yeniden çerçevelemiştir.
- Copilot, Socratic ve MathosAI; bu fiziksel limitleri tamamen göz ardı ederek sadece doğrusal veya izometrik büyüme hesapları yapmakla yetinmişlerdir.

Optimizasyon Stratejileri ve Matematiksel Doğruluk Seviyesi: Kombinatorik optimizasyon gerektiren Soru 4'te, uygulamaların matematiksel yaklaşımları ve ulaştıkları çözümler radikal biçimde farklılaşmaktadır:

- Yöneylem Araştırması Yaklaşımı: ChatGPT, Gemini, DeepSeek, JuliusAI ve MathosAI problemi karar değişkenleriyle ($x_i \in \{0, 1\}$) 0-1 Tamsayılı Doğrusal Programlama modeli olarak kurup kesin matematiksel optimizasyon önermişlerdir.
- Elde Edilen Çözümler:
  - Gemini; 5'er kişilik takımlarla 440 kg ve 440 kg tam dengesini (Fark: 0 kg) bulmuş, F ve L'yi dışarıda bırakmıştır.
  - Claude; zikzak sezgiseliyle 423 kg ve 423 kg dengesini kurmuş, H ve K'yi yedek bırakmıştır.
  - MathosAI; F ve H'yi dışarıda bırakarak 419 kg ve 419 kg tam eşitliğini hesaplamıştır.
  - JuliusAI; G ve I'yı yedek bırakarak 430 kg ve 430 kg tamsayı çözümünü sunmuştur.
- Kısıt İhlalleri: Socratic, eşit oyuncu sayısı kısıtını ihlal ederek bir takımı 6 kişi (459 kg), diğer takımı ise 5 kişi (485 kg) olarak kurmuştur. Copilot ise her takımın ağırlığının 500 kg'ı aşamayacağı kısıtını defalarca ihlal ederek 543, 522 ve 507 kg'lık geçersiz takımlar önermiştir.

Uzamsal ve Dinamik Mekanizma Tasarımı Esnekliği: Bahçenin matematiği probleminin (Soru 3) çözümünde YZ uygulamaları iki farklı tasarımsal kampa bölünmüştür:

- Dinamik Tasarım Kampı: Gemini, Claude, ChatGPT ve DeepSeek; keçiyi sabit bir kazığa bağlamak yerine bahçe aksı boyunca hareket eden bir kayar tel sistemine bağlamışlardır. Bu tasarımla taranan alanı bir daire değil, "Stadyum Geometrisi / Minkowski Toplamı" olarak modelleyerek dairesel olmayan dikdörtgen alanı çiçeklere zarar vermeden %90 oranında optimize etmişlerdir.
- Statik Tasarım Kampı: Copilot, Socratic, JuliusAI ve MathosAI ise klasik düşünme kalıplarının dışına çıkamayarak keçiyi sadece bahçenin statik bir köşesine veya geometrik merkezine bağlayan geleneksel dairesel kısıtlı ipler önermişlerdir.

Sonsuzluk Paradoksu ve Limit Yaklaşımı: Soru 2'de bağımsız olay sayısının teorik sınırını tartışırken uygulamaların soyutlama kapasiteleri ayrışmaktadır:

- DeepSeek, Gemini ve JuliusAI; olasılıkların sadece birim kesir (1/n) olmak zorunda olmadığını varsayarak, ardışık terimlerin birbirini sadeleştirdiği teleskopik çarpım serilerini ($\Pi \frac{n}{n+1} \; x \; \frac{n+1}{24n} = \frac{1}{24}$) ve Fibonacci oranlarını kurmuş, böylece teorik sınırın sonsuz olduğunu matematiksel olarak ispatlamışlardır.
- MathosAI, Socratic ve Copilot; analizi sadece birim kesirlerle (24=2x2x2x3) sınırlandırarak maksimum bağımsız olay sayısının kesin olarak 4 olabileceğini savunmuşlardır.

Üçüncü Alt Probleme İlişkin Bulgular: Matematiksel Yaratıcı Problemlere Yanıt Vermeyen veya Yanıt Vermede Başarısız Olan Yapay Zekâ Uygulamalarının Ortak Özellikleri

Araştırmanın üçüncü alt problemi olan *"Matematiksel yaratıcı problemlere yanıt vermeyen veya yanıt vermede başarısız olan yapay zekâ uygulamalarının ortak özellikleri nelerdir?"* sorusunu yanıtlamak amacıyla nitel içerik analizi derinleştirilmiştir, Tablo 5'te yanıt vermeyen ve başarısız olan uygulamaların ortak özelliklerine ve kodlarına yer verilmiştir.

Tablo 5. Yanıt vermeyen ve başarısız olan uygulamaların ortak özellikleri ve kodları

Temalar	Kategoriler	Kodlar	Kodun Açıklaması
Tema: Yapısal, Epistemolojik ve Semantik Çözüm Engelleri	K 1: Semantik Bağlamı Ayrıştıramama ve Doğal Dil Engelleyici Kısıtlar (Yanıt Vermeyenler)	Kod 1.1: "Yalnızca Matematik" / "Eksik Giriş" Reddi	Doğal dille yazılmış, hikayeleştirilmiş veya rutin dışı problem metinlerini cebirsel denklemlere dönüştürememe durumu.
	K 2: Mimari Girdi ve Bağlam Penceresi Sınırlandırması (Yanıt Vermeyenler)	Kod 2.1: "Maksimum Karakter Aşımı" Hatası	Çok adımlı sınır koşulları içeren uzun problem metinlerini kabul edemeyecek kadar dar girdi pencerelerine sahip olma durumu.
	K 3: Rutin Dışı Heuristik ve Adım Adım Çözüm Veritabanı Yetersizliği (Yanıt Vermeyenler)	Kod 3.1: "Çözülemiyor" / "Adımlar Desteklenmiyor" Çıktısı	Önceden tanımlanmış kurallar, algoritmalar ve formüller (türev, integral vb.) dışındaki açık uçlu, Fermi tahmini veya optimizasyon gerektiren dinamik problemleri çözememe durumu.
	K 4: Semantik Kısıt Yönetimi ve Mantıksal Kuralları İhlal Etme Eğilimi (Yanıt Verip Başarısız Olanlar)	Kod 4.1: Matematiksel/Mantıksal Sınır ve Kısıt İhlali	Yanıt ürettiği halde problem yönergesinde açıkça belirtilen sınır koşullarını (ağırlık sınırı, eşit eleman sayısı vb.) göz ardı etme eğilimi.
	K 5: Doğrulama Altyapısı Eksikliği ve Aritmetik Sapma (Yanıt Verip Başarısız Olanlar)	Kod 5.1: Aritmetik Hata ve Sezgisel Hesaplama Sapması	Kod çalıştırma veya sembolik doğrulama motoru desteği olmayan uygulamaların basit toplama/çarpma adımlarında sapma yapması.

Tablo 5 incelendiğinde beş farklı kategoriye ulaşılmış olup, bu kategorilerden üç tanesi yanıt vermeyen YZ uygulamaları, iki tanesi yanıt verip başarısız olan uygulamaları temsil etmektedir.

Yanıt Vermeyen (Doğrudan Görevi Reddeden) Uygulamalar (PhotoMath, Mathway, WolframAlpha, Symbolab)

Bu grupta yer alan kural tabanlı sembolik hesaplama motorları ve OCR odaklı matematik yardımcıları, rutin dışı sözel ve yaratıcı kurgular karşısında tamamen işlevsiz kalmışlardır.

Semantik Bağlamı Ayrıştıramama (PhotoMath ve Symbolab): Bu uygulamaların OCR ve doğal dil işleme (NLP) katmanları, edebi ögeler veya kurgusal kısıtlar içeren metinleri matematiksel bir operatör ağacına dönüştürememektedir. PhotoMath uygulaması, Soru 1 ve Soru 2'de yer alan tekerleme kurgusunu ve olasılık senaryosunu okuduğunda, matematiksel bir denklem yapısı bulamadığı için görevi doğrudan şu mesajla reddetmiştir: *"Hımm, bu doğru görünmüyor. Üzgünüz, ancak yalnızca matematiğe yardımcı olabiliriz. Bir matematik problemi tara, öğrenmeye başlayalım!"* Benzer şekilde Symbolab, 2, 3, 4 ve 5. sorularda yer alan uzun metinli yönergeleri analiz edemeyerek sisteme *"eksik giriş"* uyarısı vermiştir.

Mimari Girdi ve Bağlam Penceresi Sınırlandırması (WolframAlpha): Rutin olmayan yaratıcı matematik problemleri, doğası gereği problemin sınır şartlarını çizen detaylı ve uzun açıklamalar gerektirmektedir. Gelişmiş bir hesaplama motoru olmasına rağmen WolframAlpha, 5 sorunun tamamında metin uzunluğu ve kısıtların çokluğu nedeniyle işlem yapmayı reddetmiştir. Sistem, her denemede şu hata mesajını vermiştir:

> *"Hata, maksimum karakter sayısını aştınız."*

Rutin Dışı Heuristik ve Çözüm Şablonu Yetersizliği (Mathway ve Symbolab): Bu uygulamaların çözüm motorları, okul müfredatlarında yer alan standart "rutin" soru tiplerinin veritabanında

eşlenmesine dayanır. Yaratıcı problemler ise önceden tanımlanmış tek bir çözüm şablonuna sahip olmadığı, varsayımda bulunma (Kükey vd., 2019) ve heuristik sezgisel akıl yürütme (Pólya, 1945) gerektirdiği için bu sistemlerin sınırlarının dışına çıkmaktadır. **Symbolab**, Fermi tahminleri ve yaklaşık oransal modelleme gerektiren 1. soru için şu sistem çıktısını üreterek çözümü reddetmiştir: *"Bu soru için adımlar şu anda desteklenmiyor."* Benzer şekilde **Mathway**, nispeten kısa bir metne sahip olan ve sistem tarafından görüntülenebilen 2. bağımsız olasılık sorusu için doğrudan *"Çözülemiyor."* yanıtını vermiştir.

Yanıt Verip de Başarısız Olan Uygulamalar (Copilot ve Socratic)

Araştırmanın en önemli bulgularından biri, sisteme başarılı bir şekilde yanıt ürettiği halde, problemdeki katı matematiksel kısıtları yönetemeyerek operasyonel olarak başarısız olan uygulamaların saptanmasıdır.

Semantik Kısıt Yönetimi ve Mantıksal Kuralları İhlal Etme Eğilimi (Copilot ve Socratic): Bu hata semptomu, YZ uygulamalarının sözel bağlamı anladıkları ve çözüm önerdikleri halde, problemin eşzamanlı sınır şartlarını mantıksal denetim süzgecinden geçiremediklerini göstermektedir:

- Microsoft Copilot'un Ağırlık Kısıtı İhlali (Soru 4): Copilot, halat çekme takımlarını oluştururken, "her bir takımın toplam ağırlığı 500 kg'ı geçmemelidir" kuralını ardışık denemelerinde defalarca ihlal etmiştir. Copilot'un önerdiği Takım 1 alternatiflerinin toplam kütleleri 543 kg, 522 kg ve 507 kg olarak hesaplanmıştır. Uygulama, 500 kg sınırını aşan takımları geçerli birer çözüm gibi sunarak semantik kısıt yönetiminde başarısız olmuştur.
- Socratic'in Oyuncu Sayısı Eşitliği İhlali (Soru 4): Socratic, halat çekme takımlarını kurgularken "iki takımda oyuncu sayısı eşit olmalıdır" kısıtını doğrudan ihlal etmiştir. Uygulama, Takım 1'i 6 kişi (459 kg) olarak kurarken, Takım 2'yi 5 kişi (485 kg) olarak yapılandırmıştır. Socratic, her iki takımın 500 kg sınırını sağladığını kontrol etmiş ancak eşit eleman sayısı kuralını feda ederek yöntemsel başarısızlığa uğramıştır.

Doğrulama Altyapısı Eksikliği ve Aritmetik Sapma (Socratic): Kod yürütme altyapısına (örn. Python REPL) sahip olmayan uygulamaların, tamamen sezgisel dilsel olasılık tahminleriyle işlem yaparken basit aritmetik hesaplarda ve mantık zincirlerinde sapma yaşadıkları gözlenmiştir:

- Socratic'in Olasılık Çarpım Hatası (Soru 2): Socratic, çarpımları 1/24 olan üç bağımsız olasılık olayını kurgularken, bağımsız olayların olasılıklarını sırasıyla 1/2, 1/6 ve 2/7 olarak belirlemiştir. Bu üç bağımsız olayın gerçekleşme olasılığını çarparken şu matematiksel hatayı yapmıştır:

  "Bu örnekte 1/2 × 1/6 × 2/7 = 2/84 ≈ 1/42 olur (tam olarak 1/24 için olasılıkları iyi ayarlamak gerekir)." Uygulama, çarpım sonucunun 1/42 olduğunu saptamış ancak bu sonucun 1/24 hedefiyle uyuşmadığını fark etmesine rağmen alternatif bir geçerli üçlü kesir kurgulayamamıştır. Bu bulgu, sembolik bir hesaplama ve kod çalıştırma motorundan yoksun olan YZ'lerin aritmetik hata yapmaya ne kadar meyilli olduğunu kanıtlamaktadır (JuliusAI'ın Python kullanarak sıfır hata yapmasıyla tezat oluşturur).

Araştırmada kullanılan tüm YZ uygulamalarının bilişsel ve yapısal açıdan genel bir karşılaştırılmasına aşağıda (Tablo 6) yer verilmiştir. Tablo 6 incelendiğinde yapay zekâ uygulamalarının yaratıcı matematik problemlerinde operasyonel performans farklılıklarının mimari ve epistemolojik kaynaklarını açıklayan kuramsal bir sentez modeli sunmaktadır. Çalışmada ele alınan bu uygulamalar, salt "başarılı/başarısız" ikilemiyle sınırlı kalmaksızın; sahip oldukları bilişsel mimariler, dil işleme kapasiteleri ve çözüm motoru altyapıları doğrultusunda yöntemsel olarak üç farklı kategoride sınıflandırılmıştır:

1. Genel Amaçlı Büyük Dil Modelleri: ChatGPT, Gemini, Claude, Microsoft Copilot ve DeepSeek.
2. Yanıt Veren Matematik Odaklı YZ Uygulamaları: JuliusAI, MathosAI ve Socratic.

3. Yanıt Vermeyen Matematik Odaklı YZ Uygulamaları: PhotoMath, Mathway, WolframAlpha ve Symbolab.

Tablo 6. YZ uygulamalarının bilişsel ve yapısal karşılaştırması

Karşılaştırma Boyutları	**Genel Amaçlı Büyük Dil Modelleri (LLM'ler) (ChatGPT, Gemini, Claude, Copilot, DeepSeek)**	**Yanıt Veren (LLM Tabanlı/ Üretken) Matematik Odaklı YZ Uygulamaları (JuliusAI, MathosAI, Socratic)**	**Yanıt Vermeyen (Kural/Sembolik Tabanlı) Matematik Odaklı YZ Uygulamaları (PhotoMath, Mathway, WolframAlpha, Symbolab)**
Bağlam Penceresi ve Girdi Kapasitesi	Çok Geniş / Sınırsız (Doğal dil hikayeleştirmelerini ve uzun yönergeleri işleyebilir)	Geniş (Sözel açıklamaları işleyebilir, ancak bazen karmaşık kısıtlarda tıkanabilir)	Çok Dar / Sınırlı (Maksimum karakter sınırına çarpar, sözel metinleri reddeder)
Dilsel ve Semantik Anlamlandırma	Çok Yüksek (Edebi tekerlemeleri, mecazları ve çevre kısıtlarını değişkenlere dökebilir)	Orta-Yüksek (Matematiksel kurguyu anlayabilir, ancak felsefi ve ekolojik metaforları çözemez)	Yok / Çok Zayıf (Denklem şablonu dışındaki doğal dil metinlerini "matematik dışı" kabul eder)
Problem Çözme Yaklaşımı	Heuristik ve Esnek (Ters tasarım, topolojik teklik, spor analojisi gibi dinamik modeller kurar)	Uygulamalı / Algoritmik (Julius Python koduyla çözebilir, Mathos/Socratic işlem adımları sunar)	Rutin ve Şablon Bağımlı (Sadece tanımlı kural veritabanını adım adım uygular)
Optimizasyon ve Modelleme Yeteneği	Üst Düzey (0-1 Tamsayılı Programlama, Minkowski Summa, Voronoi diyagramları tasarlayabilir)	Matematik Odaklı / Sezgisel (Tamsayılı programlama kurabilir, ancak dinamik alan tasarımlarında statiktir)	Sadece Aritmetik / Sembolik (Dinamik alan sınırlarını veya kombinatorik kısıt optimizasyonlarını çözemez)

Söz konusu üç grubun yaratıcı problem çözme süreçlerindeki bilişsel ve yapısal sınırlarını sistematik olarak karşılaştırmak amacıyla, nitel içerik analizi standartlarına ve yaratıcı problem çözme kuramsal çerçevesine dayalı dört temel karşılaştırma kriteri belirlenmiştir. Tablo 6'da sunulan karşılaştırmanın değerlendirme boyutları ve bu boyutların analiz kriterleri şu şekildedir:

• Bağlam Penceresi ve Girdi Kapasitesi: Bu kriter, uygulamaların çok adımlı sınır koşulları ve ayrıntılı sözel kurgular içeren uzun problem metinlerini hata vermeksizin kabul etme ve işleme kapasitesini ölçmektedir. Geleneksel hesaplama motorlarının (örn. WolframAlpha) katı karakter sınırları nedeniyle uzun metinli girdilerde hata vermesi, LLM tabanlı modellerin ise binlerce kelimelik bağlamı eşzamanlı olarak yönetebilmesi bu kriterin ampirik temelini oluşturur.

• Dilsel ve Semantik Anlamlandırma: Edebi tekerlemeler, şiirsel kurgular veya kurgusal metaforlar içerisinde sunulan kısıtları, matematiksel değişkenlere ve denklemlere dönüştürebilme becerisi incelenmiştir. Geleneksel çözücülerin (örn. PhotoMath) sözel kurguları doğrudan "matematik dışı" kabul ederek reddetmesi, dil modellerinin ise bu mecazları ekolojik metaforlar veya biyomekanik limitler olarak yeniden çerçeveleyebilmesi (reframing) bu boyuttaki derecelendirmeyi belirlemiştir.

• Problem Çözme Yaklaşımı: Araçların çözüme ulaşırken önceden tanımlanmış katı kurallara ve şablon veritabanlarına mı bağımlı kaldığı, yoksa çoklu çözüm yolları, ters tasarım ve heuristik (sezgisel) akıl yürütme gibi esnek süreçleri mi işe koştuğu değerlendirilmiştir.

• Optimizasyon ve Modelleme Yeteneği: Uygulamaların kombinatorik kısıt durumlarında (örn. halat çekme dengesi) tamsayılı doğrusal programlama modelleri kurabilme, kısıt sınırlarını denetleyebilme ve geometrik alan tasarımlarında dinamik uzamsal mekanizmalar (örn. Minkowski Stadyum alanı) üretebilme yeterlikleri analiz edilmiştir.

TARTIŞMA, SONUÇ ve ÖNERİLER

Bu araştırmadan elde edilen bulgular, YZ uygulamalarının matematiksel yaratıcı problemlerde yalnızca standart matematiksel işlemleri gerçekleştirmediğini; problemi farklı biçimlerde yorumlayabildiğini, farklı matematiksel yapıları ilişkilendirebildiğini, problemlere alternatif çözüm yolları oluşturabildiğini göstermiştir. Bununla birlikte, bu kapasitenin tüm uygulamalarda farklı düzeylerde olabildiği; özellikle strateji çeşitliliği, yaratıcı varsayım oluşturma, matematiksel doğruluk

ve çözümün doğrulanması açısından uygulamalar arasında belirgin farklılıklar olduğu ortaya çıkmıştır.

YZ uygulamalarında strateji çeşitliliği ve matematiksel yaratıcı problem çözme

Araştırmada YZ uygulamalarının geliştirdiği çözüm stratejilerinin geometrik ve uzamsal modelleme, sayı teorisi ve oransal akıl yürütme, sınırlandırılmış kombinatorik bölümleme, uzamsal rota ve grafik ağları ile izomorfik tasarım ve dinamik zincirleme stratejisi şeklinde çeşitlendiği görülmüştür. Tüm problemler genelinde üretilen çözüm stratejilerinin nicelik olarak kullanımı bakımından ChatGPT'nin ilk sırada yer alması; onu Gemini, DeepSeek, JuliusAI, MathosAI, Claude, Copilot ve Socratic'in izlemesi, YZ uygulamalarının aynı matematiksel problem karşısında farklı genişlikte çözüm kapasitelerine sahip olduğunu göstermektedir.

Bu sonuç, matematiksel problem çözmede strateji kullanımının yalnızca belirli bir algoritmanın uygulanmasından farklı olduğunu vurgulayan Fülöp'ün (2015) yaklaşımıyla örtüşmektedir. Fülöp, strateji, yöntem ve algoritmanın birbirinden ayrılması gerektiğini; problem çözme stratejilerinin öğrencilerin kavramsal ve işlemsel matematiksel yeterliklerinin gelişimiyle ilişkili olduğunu ortaya koymuştur. Dolayısıyla bu araştırmada YZ'lerin ürettiği farklı stratejilerin sayısı yalnızca "kaç farklı işlem yapıldığı" şeklinde değil, problemlerin farklı matematiksel yapılar üzerinden yeniden ele alınabilmesinin bir göstergesi olarak değerlendirilmelidir.

Bu sonuç özellikle Ye ve arkadaşları (2025) tarafından gerçekleştirilen çalışmayla doğrudan ilişkilidir. Ye ve arkadaşları matematiksel YZ performansının yalnızca doğru cevaba ulaşma açısından değerlendirilmesinin yetersiz olduğunu ve YZ'lerin yeni çözüm yolları geliştirme kapasitesinin ayrıca incelenmesi gerektiğini savunmuştur. Ortaokul düzeyinden olimpiyat düzeyine kadar farklı problemlerin kullanıldığı çalışmalarında, LLM'lerin standart matematiksel görevlerde başarılı olabilmelerine rağmen yaratıcı çözüm üretme kapasitelerinin modeller arasında önemli ölçüde değiştiği ve Gemini-1.5-Pro'nun yeni çözüm üretiminde diğer modellerden daha iyi performans gösterdiği bulunmuştur.

Mevcut araştırmada ise ChatGPT'nin strateji çeşitliliği/bilişsel akıcılık, Gemini'nin ise özellikle problemi farklı kavramsal çerçevelerden ele alma ve yaratıcı yeniden yapılandırma bakımından öne çıkması, Ye ve arkadaşlarının (2025) ortaya koyduğu modelden modele değişen yaratıcı performans bulgusuyla birlikte değerlendirilebilir. Bununla birlikte iki araştırmanın doğrudan aynı şeyi ölçmediği dikkat çekmelidir. Ye ve arkadaşları yaratıcılığı yeni çözüm üretimi üzerinden değerlendirmişken, mevcut araştırmada stratejilerin ve yaratıcı varsayımların nitel olarak kodlanması yoluna gidilmiştir. Bu nedenle mevcut bulgu Ye ve arkadaşlarının sonucunun birebir tekrarı değil, farklı bir araştırma tasarımıyla elde edilmiş tamamlayıcı bir bulgu olarak değerlendirilmelidir.

Yaratıcı varsayımlar, yeniden çerçeveleme ve yaratıcı problem çözme

Araştırmanın önemli bulgularından biri, YZ uygulamalarının bazı problemlerde yalnızca verilen koşulları kullanmakla kalmayarak problemin sınırlarını sorgulayan varsayımlar geliştirmesidir. YZ'lerin ürettiği varsayımların biyomekanik ve jeodezik sınırları sorgulama, semantik ve sistemik metaforları bütünleştirme, teorik olasılık ve sonsuzluk sınırlarını zorlama, matematiksel kavramların yeniden tanımlanması ve imkânsızlık/yapısal çelişki keşfi şeklinde çeşitlenmesi, YZ'lerin yaratıcı problem çözmede problem alanını genişletebildiğini göstermektedir.

Bu bulgu, yaratıcı problem çözme yaklaşımının temel mantığıyla uyumludur. Widya ve arkadaşları (2020), yaratıcı problem çözme yaklaşımının matematik ve fen bilimlerinde problem durumlarının farklı biçimlerde ele alınmasını ve alternatif fikirlerin geliştirilmesini destekleyen bir çerçeve sunduğunu belirtmektedir. Çalışma, yaratıcı problem çözme modelinin matematik ve fen bilimleri eğitimindeki gelişimini ve uygulamalarını literatür üzerinden incelemiştir. Mevcut araştırmada YZ'lerin problem sınırlarını esnetmesi, farklı kavramsal alanları ilişkilendirmesi ve bazı durumlarda problemi yeniden tanımlaması, bu yaratıcı problem çözme anlayışıyla kavramsal olarak örtüşmektedir.

Özellikle bazı YZ'lerin fiziksel veya kavramsal sınırları sorgulaması, yaratıcı problem çözmenin yalnızca "daha fazla çözüm üretme" boyutuyla açıklanamayacağını göstermektedir. Burada esneklik ve yeniden çerçeveleme de önem kazanmaktadır. Örneğin Claude ve DeepSeek'in büyüyen bir insanın fiziksel sınırlarını Kare-Küp Yasası bağlamında değerlendirmesi, problemin yüzeysel anlatımından daha genel bir matematiksel ilkeye geçiş örneği olarak değerlendirilebilir. Gemini'nin bazı bağlamları ekolojik kriz metaforu üzerinden yeniden yorumlaması ise matematiksel yapı ile bağlamsal anlam arasında yeni bir ilişki kurulmasına örnek oluşturmaktadır.

Bu bağlamda yaratıcı varsayım üretiminde ChatGPT'nin ilk sırada, Gemini, Claude ve DeepSeek'in onu izlemesi; JuliusAI'nin daha sonra gelmesi ve Copilot, Socratic ile MathosAI'nin daha sınırlı sayıda yaratıcı varsayım üretmesi, YZ'lerin yaratıcılık performansının tek boyutlu olmadığını göstermektedir. Nitekim Ye ve arkadaşları (2025) da farklı LLM'lerin yaratıcı matematiksel çözüm üretiminde önemli farklılıklar sergilediğini ortaya koymuştur.

Matematiksel yaratıcılık ile doğruluk arasındaki farklılaşma

Araştırmada dikkat çeken bir diğer sonuç, strateji ve varsayım çeşitliliğinin matematiksel doğrulukla zorunlu olarak aynı düzeyde gerçekleşmemesidir. ChatGPT'nin strateji çeşitliliğinde öne çıkmasına karşılık JuliusAI'nin özellikle algoritmik doğrulama, kısıtların kontrolü ve hesaplama gerektiren problemlerde güçlü performans göstermesi, YZ uygulamalarının farklı matematiksel görevlerde farklı güçlü yönlere sahip olduğunu göstermektedir.

Bu durum, matematiksel problem çözmenin yalnızca yaratıcı fikir üretmekten ibaret olmadığını göstermesi açısından önemlidir. Fülöp (2015), problem çözme stratejilerinin matematiksel yöntem ve algoritmalardan ayrılması gerektiğini ve strateji kullanımının kavramsal ve işlemsel yeterliklerle birlikte değerlendirilmesi gerektiğini ortaya koymaktadır. Szabó ve arkadaşları (2020) da problem çözme stratejilerinin matematik eğitiminde 21. yüzyıl becerileriyle ilişkisini ele alarak farklı stratejilerin problem çözme sürecindeki önemine dikkat çekmektedir.

Dolayısıyla mevcut araştırmada ChatGPT'nin yüksek strateji çeşitliliği "matematiksel olarak her durumda en başarılı uygulama" şeklinde yorumlanmamalıdır. Benzer şekilde JuliusAI'nin belirli matematiksel işlemlerde daha yüksek doğruluk göstermesi, onun yaratıcı problem çözmenin bütün boyutlarında daha üstün olduğu anlamına gelmemektedir. Araştırmanın bulguları daha çok yaratıcılık ve doğruluğun birbirinden ayrıştırılması gereken performans boyutları olduğunu göstermektedir.

Bu ayrım matematiksel problem çözmede izleme ve değerlendirme süreçlerinin önemini de gündeme getirmektedir. Losenno ve arkadaşlarının (2020) yaptığı çalışmada planlama/hedef belirleme, stratejilerin uygulanması ve izleme/değerlendirme aşamalarının matematiksel problem çözmeyle ilişkili olduğu; bilişsel yeniden değerlendirme ile öz-düzenlenmiş öğrenme arasında da önemli ilişkiler bulunduğu gösterilmiştir. Çalışmada özellikle strateji uygulama aşamasının matematiksel problem çözme sonuçlarını yordadığı bulunmuştur.

Bu sonuç mevcut araştırma açısından önemlidir; çünkü YZ'nin yaratıcı bir çözüm önermesi kadar, önerdiği çözümün matematiksel açıdan izlenmesi ve değerlendirilmesi de önem taşımaktadır. Başka bir ifadeyle, yüksek sayıda alternatif fikir üretmek tek başına yeterli değildir; bu fikirlerin problem koşullarına uygunluğu, matematiksel tutarlılığı ve sınır koşullarını karşılayıp karşılamadığı da değerlendirilmelidir.

YZ uygulamaları arasındaki benzerlikler ve farklılıklar

Araştırmada yanıt üreten YZ uygulamalarının doğrusal aritmetik modellemeye yönelme, geleneksel matematiksel temsilleri kullanma, problem kısıtlarındaki çelişkileri fark etme ve disiplinler arası standart eşlemeler kurma bakımından benzerlikler gösterdiği belirlenmiştir. Buna karşılık bilişsel derinlik, fiziksel limitleri sorgulama, optimizasyon, matematiksel doğruluk, uzamsal ve dinamik mekanizma tasarımı ile sonsuzluk ve limit kavramlarının ele alınmasında farklılıklar ortaya çıkmıştır.

Bu sonuç, Wang ve arkadaşları tarafından yapılan LLM değerlendirmeleriyle de ilişkilendirilebilir. Wang ve arkadaşları (2026), GPT-4o, DeepSeek-V3 ve Gemini-2.0 gibi modelleri farklı güçlük düzeylerindeki matematiksel veri kümelerinde karşılaştırmış ve LLM'lerin matematiksel problem

çözme performanslarının model ve problem özelliklerine göre değişebildiğini incelemiştir. Dolayısıyla mevcut araştırmada farklı YZ'lerin aynı yaratıcı probleme farklı bilişsel yollarla yaklaşması, matematiksel problem çözmede model farklılığının önemini destekleyen bir sonuç olarak değerlendirilebilir.

Benzer biçimde Ma ve arkadaşları (2026) tarafından gerçekleştirilen derleme çalışması, geometrik problem çözmenin matematiksel akıl yürütmenin önemli alanlarından biri olduğunu ve yapay zekâ araştırmalarında geometri problemlerinin hem matematiksel akıl yürütme hem de multimodal yeteneklerin değerlendirilmesinde önemli bir test alanı oluşturduğunu ortaya koymaktadır. Mevcut araştırmada gözlenen geometrik ve uzamsal modelleme farklılıkları bu açıdan anlamlıdır; çünkü yaratıcı matematik problemlerinde yalnızca sembolik işlemler değil, problemin uzamsal yapısının nasıl temsil edildiği de çözüm performansını etkileyebilmektedir.

Yanıt vermeme ve başarısız çözümler

Araştırmanın üçüncü alt problemine ilişkin bulgular, başarısızlığın iki farklı biçimde ortaya çıktığını göstermiştir. İlk grupta PhotoMath, Mathway, WolframAlpha ve Symbolab gibi uygulamalar yaratıcı problemlere doğrudan yanıt üretmemiştir. İkinci grupta ise Copilot ve Socratic gibi uygulamalar yanıt üretmiş ancak bazı problemlerde matematiksel kısıtları doğru yönetememiş, aritmetik sapmalar göstermiş veya eksik/başarısız çözümler üretmiştir. Bu iki durumun aynı kategoride değerlendirilmemesi önemlidir.

Yanıt vermeyen uygulamalar açısından WolframAlpha örneği özellikle dikkat çekicidir. WolframAlpha'nın kendisini "computational knowledge engine" olarak konumlandırması, sistemin matematiksel hesaplama ve sembolik işlemler açısından güçlü olmasına karşın, doğal dilde ifade edilmiş uzun ve yaratıcı problem durumlarının her zaman aynı biçimde işlenemeyeceğini açıklamak açısından önemlidir. Bu nedenle araştırmada gözlenen yanıt vermeme durumu doğrudan matematiksel yetersizlik olarak değil, problem türü ile sistemin kullanım amacı ve giriş biçimi arasındaki uyumsuzluk bağlamında yorumlanmalıdır.

Yanıt üreten fakat başarısız veya eksik çözümler geliştiren uygulamalarda ise farklı bir problem söz konusudur. Burada sistem problem bağlamını işleyebilmekte ancak çözüm sırasında kısıtları, aritmetik ilişkileri veya problem koşullarını yeterince denetleyememektedir. Bu durum, matematiksel problem çözmede yalnızca çözüm üretmenin değil, çözümün izlenmesi, değerlendirilmesi ve gerektiğinde yeniden yapılandırılmasının önemini göstermektedir. Losenno ve arkadaşlarının (2020) çalışmasının öz-düzenlenmiş öğrenmenin izleme/değerlendirme boyutunu ve problem çözme sonuçlarıyla ilişkisini ortaya koyması bu açıdan kuramsal bir dayanak sağlamaktadır.

Model, görev ve istem arasındaki etkileşim

Araştırmanın bulguları, YZ performansının yalnızca kullanılan modelin adına veya genel kapasitesine göre açıklanamayacağını da göstermektedir. Özellikle yaratıcı problemlerde problemin yapısı, istenen çözüm biçimi, kullanılan girişin niteliği ve modelin matematiksel muhakeme kapasitesi birlikte değerlendirilmelidir.

Bu sonuç Schorcht ve arkadaşlarının (2026) çalışmasıyla güçlü biçimde örtüşmektedir. Araştırmacılar Gemini 1.5 Pro, Claude 3.5 Sonnet, ChatGPT-o3 mini ve DeepSeek-R1 modellerini altı matematik problemi ve dört farklı istem (prompt) tekniğiyle toplam 2880 çözüm üzerinden incelemiştir. Bulgular, içerikle ilişkili kalitenin daha çok model türü ve görev özellikleriyle; süreç ve pedagojik bağlamla ilişkili kalitenin ise istem tasarımıyla güçlü biçimde ilişkili olduğunu göstermiştir. Ayrıca hiçbir modelin veya istem tekniğinin bütün boyutlarda en iyi performansı göstermediği sonucuna ulaşılmıştır.

Bu sonuç mevcut araştırmanın "tek bir YZ her açıdan en başarılıdır" biçiminde yorumlanmaması gerektiğini desteklemektedir. Mevcut araştırmada ChatGPT'nin strateji çeşitliliğinde, Gemini'nin yaratıcı yeniden çerçevelemede ve JuliusAI'nin algoritmik doğruluk/kısıt yönetiminde öne çıkması, farklı YZ'lerin farklı performans profilleri oluşturduğunu göstermektedir. Schorcht ve arkadaşlarının (2026) bulguları da model seçiminin yanında görev ve istem tasarımının önemli olduğunu

gösterdiğinden, mevcut araştırmanın sonuçları YZ performansının bağlama duyarlı olduğu şeklinde yorumlanabilir.

Genel sentez

Araştırmanın tüm bulguları birlikte değerlendirildiğinde, matematiksel yaratıcı problem çözmede en çok strateji üreten YZ ile en doğru çözümü üreten YZ'nin aynı uygulama olmak zorunda olmadığı görülmektedir. ChatGPT'nin strateji çeşitliliği bakımından öne çıkması, Gemini'nin bazı problemlerde yeniden çerçeveleme ve yaratıcı varsayım geliştirme kapasitesi, JuliusAI'nin ise algoritmik doğruluk ve kısıt yönetimindeki güçlü performansı, YZ'lerin birbirinden farklı bilişsel ve teknik profillere sahip olduğunu göstermektedir.

Bu sonuç Ye ve arkadaşlarının (2025) matematiksel yaratıcılığın modeller arasında değiştiği yönündeki bulgusuyla uyumludur. Aynı zamanda Schorcht ve arkadaşlarının (2026) hiçbir modelin tüm kalite boyutlarında optimal performans göstermediği sonucuyla da örtüşmektedir. Dolayısıyla mevcut araştırmanın özgün katkısı, YZ performansını yalnızca doğru/yanlış cevap üzerinden değerlendirmek yerine; strateji çeşitliliği, yaratıcı varsayım, yeniden çerçeveleme, matematiksel doğruluk ve başarısızlık biçimlerini birlikte ele almasıdır.

Bu çerçevede matematik eğitiminde YZ kullanımı için yalnızca “hangi YZ daha doğru cevap veriyor?” sorusunun sorulması yeterli değildir. Daha kapsamlı olarak, hangi YZ farklı stratejiler üretiyor, hangi YZ problemi yeniden çerçeveliyor, hangi YZ yaratıcı varsayımlar geliştiriyor, hangi YZ matematiksel kısıtları daha iyi denetliyor ve hangi YZ kendi çözümünü daha güvenilir biçimde doğrulayabiliyor? sorularının birlikte ele alınması gerekmektedir. Bu yaklaşım, yaratıcı matematik problemlerinde YZ'lerin değerlendirilmesini tek boyutlu başarı sıralamasından çıkararak çok boyutlu bir performans değerlendirmesine dönüştürmektedir.

Sonuç

Araştırmanın birinci alt problemi kapsamında YZ uygulamalarının kullandığı stratejiler ve oluşturduğu varsayımlar incelenmiştir. Stratejiler geometrik ve uzamsal modelleme, sayı teorisi ve oransal akıl yürütme, sınırlandırılmış kombinatorik bölümleme, uzamsal rota ve grafik ağları ile izomorfik tasarım ve dinamik zincirleme stratejisi şeklinde belirlenmiştir. Araştırmada sorulara yanıt veren yapay zekâ uygulamaları beş soru genelinde ürettiği benzersiz çözüm strateji kodlarının toplam sayısına (Akıcılık) göre çoktan aza doğru sıralandığında ilk sırada ChatGPT yer alırken onu sırasıyla Google Gemini, DeepSeek, JuliusAI, MathosAI, Claude takip etmiştir. Microsoft Copilot ve Socratic ve daha sınırlı strateji çeşitliliği performansı göstermiştir. YZ uygulamaları yalnızda mevcut matematiksel yöntemleri uygulamamış, fiziksel sınırları sorgulama, matematiksel kavramları yeniden tanımlama, imkânsızlıkları tartışma gibi yaratıcı varsayımları geliştirmişlerdir. Yaratıcı varsayım geliştirmede YZ uygulamaları karşılaştırılmış, problem sınırlarını esnetme, fiziksel çelişkileri saptama ve kavramsal yeniden çerçeveleme (Özgünlük ve Esneklik) kapsamında ürettiği varsayım kodlarına göre öne çıkan ChatGPT olurken, onu Google Gemini, Claude ve DeepSeek takip etmiş, JuliusAI orta düzeyde, Microsoft Copilot, Socratic ve MathosAIise daha sınırlı düzeyde kalmıştır.

İkinci alt problem bulguları incelendiğinde problemlere yanıt veren YZ uygulamalarının tümünde doğrusal aritmetik modelleme başlangıcında, geleneksel matematiksel temsillerin seçiminde, problem kısıtlarında çelişki saptama başarısında, standart disiplinler arası eşleme tercihinde benzerlikler görülürken, yanıt veren YZ uygulamalarının arasında bilişsel derinlik ve fiziksel limitleri sorgulama düzeyinde, optimizasyon stratejileri ve matematiksel doğruluk seviyesinde, uzamsal ve dinamik mekanizma tasarımı esnekliği ile sonsuzluk paradoksu ve limit yaklaşımında farklılıklar görülmüştür.

Üçüncü alt problem kapsamında değerlendirme yapıldığında ise Copilot ve Socratic uygulamalarının kuralları ihlal etme durumu, ayrıca Socratic’te doğrulama altyapısı eksikliği ve aritmetik sapma durumları başarısızlık olarak nitelendirilmiştir. PhotoMath, Mathway, WolframAlpha ve Symbolab uygulamalarının hata durumlarının temel nedeni teknolojik bir "hata" değil; epistemolojik ve mimari bir tasarım tercihidir. Bu da başarısızlık olarak değil yanıt vermeyen YZ uygulamaları olarak kategoriye dahil edilmelerini sağlamıştır. Bu sistemler, öğrencilere okul ödevlerinde yer alan rutin işlemlerde hızlı ve adım adım mekanik destek sağlamak üzere optimize edilmiştir.

Araştırmadan elde edilen genel sonuçlar bütüncül olarak değerlendirildiğinde, rutin olmayan yaratıcı matematiksel akıl yürütme süreçlerinde farklı yapay zekâ mimarilerinin kendilerine özgü bilişsel ve algoritmik avantajlar sunduğu görülmektedir. Bu kapsamda ChatGPT strateji çeşitliliği ve bilişsel akıcılıkta, Gemini kavramsal esneklik ve özgün varsayımlarda, JuliusAI ise entegre Python kod yürütme motoru sayesinde algoritmik doğruluk ve kısıt yönetiminde üstün performans sergilemiştir. Genel amaçlı büyük dil modellerinin (ChatGPT ve Gemini) matematiksel yaratıcılığın 'ıraksak düşünme', 'yeniden çerçeveleme' ve 'varsayım geliştirme' gibi anlamsal ve kavramsal katmanlarında yüksek bir entelektüel esneklik sunduğu; buna karşın matematik odaklı uygulamaların (özellikle JuliusAI) problemin 'algoritmik doğrulama', 'kısıt denetimi' ve 'aritmetik tutarlılık' gibi katı matematiksel zeminlerinde hata payını sıfıra indirerek en güvenilir performansı ortaya koyduğu saptanmıştır. Elde edilen bu sentez, matematik eğitiminde yapay zekâ entegrasyonu tasarlanırken dilsel-kavramsal yaratıcılık (akıcılık ve esneklik) ile hesaplama doğruluğu (algoritmik hassasiyet ve kısıt yönetimi) arasındaki bu yapısal dengenin pedagojik ihtiyaçlar doğrultusunda titizlikle gözetilmesi gerektiğini doğrulamaktadır.

Öneriler

Araştırma sonucunda matematik eğitimi, eğitim teknolojileri ve gelecekte gerçekleştirilecek araştırmalara yönelik çeşitli öneriler geliştirilmiştir. Öncelikle**,** matematik öğretim programlarında öğrencilerin problemi yeniden yorumlamalarını, farklı matematiksel temsiller arasında geçiş yapmalarını, varsayımlar oluşturmalarını, alternatif çözüm yolları üretmelerini ve çözümlerini gerekçelendirmelerini gerektiren problem türlerinin yaygınlaştırılması önerilmektedir. Ayrıca problemlerin sadece sonuca ulaşmaya değil strateji çeşitliliğine, yeniden yapılandırmaya, çözümü değerlendirme süreçlerine imkan sağlayacak biçimde tasarlanması önerilmektedir.

İkinci olarak, matematik öğretmenlerine alanlarına özgü yapay zekâ uygulamalarının kullanabilmelerine, öğretmenlerin öğrencilerle birlikte YZ yanıtlarını sorgulayarak kullanmaları için etkinlikler tasarlayabilmelerine yönelik öğretim ve mesleki gelişim çalışmaları gerçekleştirilebilir. Böylece YZ uygulamalarını sadece ödev yapan ya da doğrudan probleme cevap veren bir araç olarak kullanılmadığının farkına varılabilir.

Üçüncü olarak**,** eğitim teknolojisi geliştiricilerine, özellikle matematiksel problem çözme amacıyla geliştirilen yapay zekâ uygulamalarının güçlendirilmesi önerilmektedir. Gelecekte geliştirilecek sistemlerin yalnızca akıcı ve ikna edici metin üretmesine değil; varsayımları açıkça belirtmesine, kullanılan matematiksel modelin tutarlılığını kontrol etmesine, hesaplamaları sembolik olarak doğrulamasına, alternatif çözümleri karşılaştırmasına ve olası hatalarını belirleyebilmesine olanak sağlayacak biçimde tasarlanması önemlidir.

Dördüncü olarak**,** yapay zekâ uygulamalarının matematiksel yaratıcı problem çözme performanslarını değerlendirmek çok boyutlu değerlendirme çerçevelerinin geliştirilmesi önerilmektedir. Özellikle açık uçlu problemlerde bir çözümün başarısı yalnızca ulaşılan sonuçla belirlenemeyeceğinden; problem yorumlama, matematiksel strateji çeşitliliği, varsayım oluşturma, matematiksel modelleme, alternatif çözüm üretme, genelleme, gerekçelendirme ve çözümün uygunluğunu değerlendirme gibi boyutların birlikte ele alınması gerekmektedir.

Beşinci olarak, gelecekteki araştırmalarda yalnızca yapay zekâların ürettiği nihai çözümlerin değil, insan–YZ etkileşimi içerisinde ortaya çıkan problem çözme süreçlerinin de incelenmesi önerilmektedir. Bu kapsamda öğrencinin başlangıçtaki çözümü ile YZ desteğinden sonraki çözümü karşılaştırılabilir; öğrencinin hangi YZ önerilerini kabul ettiği, reddettiği veya değiştirdiği; YZ'nin öğrencinin problem yorumunu, strateji seçimini, varsayım oluşturmasını ve alternatif çözüm üretmesini nasıl etkilediği ayrıntılı olarak incelenebilir.

Son olarak, gelecekteki araştırmalarda YZ ile problem çözme süreçlerinin farklı öğrenci özellikleri, problem türleri ve yapay zekâ uygulamaları açısından karşılaştırmalı olarak incelenmesi önerilmektedir. Özellikle farklı yaş ve sınıf düzeylerindeki öğrencilerle, farklı matematiksel alanları içeren ve farklı düzeylerde açıklık taşıyan rutin olmayan problemlerin kullanıldığı araştırmalar

gerçekleştirilebilir. Böylece çalışmalar, matematik eğitiminde YZ kullanımına ilişkin daha güçlü pedagojik ve kuramsal çıkarımlar sağlayabilir.

Finansman

Bu araştırma için herhangi bir birey veya kurumdan finansman alınmamıştır.

Etik ve Çıkar Çatışması

Bu makale 9-12 Eylül 2026 tarihleri arasında Gaziantep'te düzenlenen Uluslararası Eğitim Kongresi-EDU Congress 2026 adlı kongrede sözlü bildiri olarak sunulmuş aynı başlıklı bildirinin genişletilmiş halidir. Yazarlar araştırmanın tüm süreçlerinde etik kurallara uygun davranıldığını ve yazarlar arasında herhangi bir çıkar çatışmasının olmadığını beyan eder.

Yazar Katkı oranı

Tüm yazarlar, bu makalenin kavramsallaştırılması, tasarımı, veri toplama, analiz ve yazımına eşit katkıda bulunmuştur. Yazar katkı payları eşit orandadır. Tüm yazarlar makalenin son halini okumuş ve onaylamıştır.

Veri Erişilebilirliği

Bu çalışmanın bulgularını destekleyen veriler, sorumlu yazardan talep üzerine temin edilebilir.

Sorumlu Yazar

İmren AYDOĞDU, imreenay@gmail.com

KAYNAKÇA

Ahn, J., Verma, R., Lou, R., Liu, D., Zhang, R., & Yin, W. (2024). Large language models for mathematical reasoning: Progresses and challenges. In N. Falk, S. Papi, & M. Zhang (Eds.), *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop* (pp. 225–237). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.eacl-srw.17

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (5th ed.). Routledge Falmer.

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.

Fülöp, E. (2015). Teaching problem-solving strategies in mathematics. *LUMAT: International Journal on Math, Science and Technology Education*, *3*(1), 37–54. https://doi.org/10.31129/lumat.v3i1.1050

Hacıoğlu Seven, E., & Erümit, A. K. (2025). Matematik eğitiminde üretken yapay zeka araçlarının incelenmesi: Araç özellikleri, kullanım amaçları ve etkileri [An examination of generative artificial intelligence tools in mathematics education: Tool features, purposes of use, and impacts]. *Uşak Üniversitesi Eğitim Araştırmaları Dergisi*, *11*(2), 136–162. https://doi.org/10.29065/usakead.1688918

Kirişçi, N. (2021). Yaratıcı problem çözme sürecinde analojik ve seçici düşünme: Seçici problem çözme modelinin matematik eğitiminde uygulama örneği [Analogical and selective thinking in creative problem solving process: The use of selective problem solving model in mathematics education]. *Muğla Sıtkı Koçman Üniversitesi Eğitim Fakültesi Dergisi*, *8*(1), 72–84. https://doi.org/10.21666/muefd.755133

Kükey, E., Aslaner, R., & Tutak, T. (2019). Matematiksel düşünmenin varsayımda bulunma bileşenine yönelik olarak ortaokul öğrencilerinin kullandıkları problem çözme stratejilerinin incelenmesi [An investigation of the problem solving strategies used of middle school students for of the assumption component of mathematical thinking]. *Journal of Computer and Education Research*, *7*(13), 146–170. https://doi.org/10.18009/jcer.535610

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE Publications.

Liu, Y., & Wang, H. (2026). Who on earth is using generative AI? *World Development*, *199*, Article 107260. https://doi.org/10.1016/j.worlddev.2025.107260

Losenno, K. M., Muis, K. R., Munzar, B., Denton, C. A., & Perry, N. E. (2020). The dynamic roles of cognitive reappraisal and self-regulated learning during mathematics problem solving: A mixed methods investigation. *Contemporary Educational Psychology*, *61*, Article 101869. https://doi.org/10.1016/j.cedpsych.2020.101869

Ma, J., Wang, W., & Jin, Q. (2026). A survey of deep learning for geometry problem solving. *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics*, 39416–39448. https://doi.org/10.18653/v1/2026.acl-long.1829

Marangoz, M. (2024). Eğitim ortamında kullanılan yapay zekâ uygulamaları [Artificial intelligence applications used in education]. *International Journal of Progression and Development in Education (IJPADE)*, *2*(2), 26–39. https://doi.org/10.5281/zenodo.14576628

Merriam, S. B. (1998). *Qualitative research and case study applications in education*. Jossey-Bass.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). SAGE Publications.

Mirzadeh, I., Alizadeh-Vahid, K., Shahrokhi, H., Tuzel, O., Bengio, S., & Farajtabar, M. (2025). GSM-Symbolic: Understanding the limitations of mathematical reasoning in large language models. In *Proceedings of the International Conference on Learning Representations* (Vol. 2025, pp. 94743–94765). https://doi.org/10.48550/arXiv.2410.05229

Mumford, M. D., Medeiros, K. E., & Partlow, P. J. (2012). Creative thinking: Processes, strategies, and knowledge. *The Journal of Creative Behavior*, *46*(1), 30–47. https://doi.org/10.1002/jocb.003

Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice* (4th ed.). SAGE Publications.

Pólya, G. (1945). *How to solve it: A new aspect of mathematical method*. Princeton University Press.

Price, A. (2006). *Creative maths activities for able students: Ideas for working with children aged 11 to 14*. Paul Chapman Publishing; SAGE Publications.

Rahman, A., Mahir, S. H., Tashrif, M. T., Aishi, A. A., Karim, M. A., Kundu, D., Debnath, T., Moududi, M. A., & Eidmum, M. Z. (2025). Comparative analysis based on DeepSeek, ChatGPT, and Google Gemini: Features, techniques, performance, future prospects. *arXiv*. https://arxiv.org/abs/2503.04783

Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal*, *24*(1), 92–96. https://doi.org/10.1080/10400419.2012.650092

Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed.). SAGE Publications.

Schoenfeld, A. H. (2016). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. *Journal of Education*, *196*(2), 1–38. https://doi.org/10.1177/002205741619600202

Schorcht, S., Müller, F. A., & Buchholtz, N. (2026). No one-size-fits-all: A study of prompt techniques and large language models to enhance AI's mathematics educational quality. *ZDM – Mathematics Education* 58, 1057–1071. https://doi.org/10.1007/s11858-026-01784-6

Schreier, M. (2012). *Qualitative content analysis in practice*. SAGE Publications.

Silver, E. A. (1997). Fostering creativity through instruction rich in mathematical problem solving and problem posing. *ZDM—The International Journal on Mathematics Education*, *29*(1), 75–80. https://doi.org/10.1007/s11858-997-0003-x

Stake, R. E. (1995). *The art of case study research*. SAGE Publications.

Szabo, Z. K., Körtesi, P., Guncaga, J., Szabo, D., & Neag, R. (2020). Examples of problem-solving strategies in mathematics education supporting the sustainability of 21st-century skills. *Sustainability*, *12*(23), Article 10113. https://doi.org/10.3390/su122310113

Şahin, F., & Yeldan, İ. (2019). Sistematik yaratıcı problem çözme etkinliklerinin, ortaokul 6. sınıf öğrencilerinin kuvvet ve hareket ünitesindeki akademik başarılarına ve sistematik yaratıcı problem çözme becerilerine etkisi [The effect of systematical problem solving activities on 6th grade students' academic success and problem solving abilities on force and action subject]. *Marmara Üniversitesi Atatürk Eğitim Fakültesi Eğitim Bilimleri Dergisi*, *49*(49), 121-143. https://doi.org/10.15285/maruaebd.404195

Turmuzi, M., Azmi, S., & Kertiyani, N. M. I. (2026). ChatGPT in school mathematics education: A systematic review of opportunities, challenges, and pedagogical implications. *Teaching and Teacher Education*, *170*, Article 105286. doi: 10.1016/j.tate.2025.105286

Wang, R., Wang, R., Shen, Y., Wu, C., Zhou, Q., & Chandra, R. (2026). Evaluation of LLMs for mathematical problem solving. *Next Research*, *9*, Article 101705. https://doi.org/10.1016/j.nexres.2026.101705

Widya, W., Nurpatri, Y., Indrawati, E. S., & Ikhwan, K. (2020). Development and application of creative problem solving in mathematics and science: A literature review. *Indonesian Journal of Science and Mathematics Education*, *3*(1), 106–116. https://doi.org/10.24042/ijsme.v3i1.4335

Ye, J., Gu, J., Zhao, X., Yin, W., & Wang, G. G. (2025). Assessing the creativity of LLMs in proposing novel solutions to mathematical problems. *Proceedings of the AAAI Conference on Artificial Intelligence*, *39*(24), 25687–25696. https://doi.org/10.1609/AAAI.V39I24.34760

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE Publications.

ISSN: 2146 - 9466
www.ijtase.net
International Journal of New Trends in Arts, Sports &Science Education – 2026, volume 15, issue 4

EXTENDED ABSTRACT

The integration of Artificial Intelligence (AI) in education has progressed from simple task automation to complex cognitive partnerships. While traditional computer algebra systems excel at rule-based, routine calculations, non-routine mathematical creative problem solving (MCPS) remains a major challenge. Non-routine problems require divergent thinking, conceptual flexibility, and the ability to operate under constraints. This study is theoretically grounded in George Pólya's heuristic steps (Pólya, 1945) and Creative Problem Solving (CPS) framework. Traditional mathematical education research frequently reports that students struggle with formulating conjectures under constraints, often relying on basic "guess-and-check" methods and ending their search prematurely upon finding a single solution (Kükey, Aslaner, & Tutak, 2019). Generative AI, specifically Large Language Models (LLMs), offers a potential shift in this dynamic by providing vast semantic databases capable of high-level conceptual reframing and divergent reasoning. This research systematically investigates the capacity of 12 AI applications to generate creative solution strategies and novel assumptions, and to manage strict mathematical constraints when presented with non-routine, semantic, and highly constrained mathematical problems. This study employed a qualitative multiple-case study design (Yin, 2018), which leverages replication logic (both literal and theoretical) to build robust empirical conclusions across cases. *Bounded Cases (AI Applications):* Twelve AI tools were selected and classified into three distinct taxonomic categories based on their cognitive architectures: 1) *General-Purpose LLMs (5 tools):* ChatGPT, Google Gemini, Claude, Microsoft Copilot, and DeepSeek. 2) LLM-Based, Math-Oriented Applications (3 tools): JuliusAI, MathosAI, and Socratic. 3) Rule-Based Traditional Solvers (4 tools): PhotoMath, Mathway, WolframAlpha, and Symbolab. *Instruments and Problem Set:* Five non-routine, highly constrained mathematical problems adapted from Price (2006) were utilized. These problems featured literary storytelling, poetic riddles, physical/biomechanical limits, and combinatorial optimization requirements, preserving their original English structure to maximize AI cognitive retrieval while requesting Turkish outputs to align with local linguistic analysis. *Data Collection and Analysis:* A systematic, two-stage data collection protocol was executed. In the first stage, all 12 tools were tested under a standard prompt. Traditional solvers failed to provide answers, prompting a second stage of deep-dive testing for the 8 responsive tools. Qualitative data were analyzed using within-case and cross-case comparative content analysis based on Johnny Saldaña's (2013) hiyerarşik coding model (codes, categories, themes). *Methodological Quality:* Inter-coder reliability was calculated using Miles and Huberman's (1994) formula, yielding a high consistency of 92% between the primary researcher and an independent expert. The empirical findings revealed that AI performance is highly non-uniform, diverging sharply across different cognitive architectures. A total of 175 creative solution strategies (categories of action) and 76 creative assumptions (re-framing mechanisms) were coded across the 8 responsive tools. 1) ChatGPT emerged as the leader in strategic fluency (cognitive fluency), producing 36 distinct strategies (20.6% of the total strategic output). It navigated between spatial modeling, brute-force listing, and process chaining seamlessly. 2) Google Gemini dominated in conceptual flexibility and original assumption generation (15.8% group share, f=12). In *Problem 1* (the giant and sea riddle), Gemini abstractly framed the poem as an "ecological-systemic crisis" where humanity's industrial tools (the giant axe) devastate nature (the giant tree), causing a global climate collapse. It also redefined "union" not as physical growth, but as a topological "Oneness". 3) Claude and DeepSeek demonstrated profound theoretical depth. In *Problem 1*, they applied the Square-Cube Law (Galileo's biomechanical principle) to prove that a giant scaled up by 1920 times would suffer skeletal collapse due to the mismatched scaling rates of volume and muscle cross-sectional area. 4) In *Problem 3* (goat-and-garden optimization), LLMs abandoned traditional static circular tethering, introducing a dynamic sliding mechanism (zip-line). They modeled the swept grazing boundary as a "Minkowski Stadion (Sausage) Geometry," optimizing grazing efficiency to over 90% of the rectangular yard. 5) Traditional Solvers (PhotoMath, Mathway, Symbolab, WolframAlpha): Exhibited complete operational failure. PhotoMath rejected the semantic inputs, stating *"We can only help with math."* Symbolab flagged *"incomplete input,"* and WolframAlpha repeatedly crashed due to strict input context window limitations (*"Oops, you exceeded the*

maximum character limit"). 6) Constraint Violations and Hallucinations (Copilot & Socratic): Despite generating highly sophisticated verbal solutions, Microsoft Copilot and Socratic failed to manage strict numerical constraints. In *Problem 4* (tug-of-war combinatorial balance), Copilot repeatedly violated the 500 kg team weight limit, proposing illegal teams of 543 kg, 522 kg, and 507 kg. Socratic violated the equal-player constraint (proposing teams of 6 vs. 5 players) and made direct arithmetic errors in *Problem 2*, asserting that 1/2x1/6 x2/7 = 2/84 approx 1/42 (mismatching its target of 1/24). 7) The Python REPL Edge (JuliusAI): JuliusAI achieved a 0% error rate in combinatorial optimization and coordinate geometry tasks. By executing an integrated Python environment, JuliusAI formulated and solved 0-1 integer linear programming models with perfect mathematical precision, overcoming the logical verification pitfalls of pure LLMs. *The study concludes that fluency, flexibility, and mathematical precision do not naturally converge in a single AI architecture.* Metacognitive Heuristics vs. Human Limitations: Unlike human students who struggle with formulating mathematical conjectures and often limit themselves to simplistic "guess-and-check" strategies (Kükey et al., 2019), general-purpose LLMs utilize their extensive semantic spaces to operate as "cognitive filters." They successfully detect mathematical impossibilities and apply Galileo's biomechanics or topological boundaries to redefine problems. The "Closed World" Manifestation: AI problem-solving strategies closely mirror the ASIT "Closed World" framework (Şahin & Yeldan, 2019) by restructuring existing system components (goat, rope, fence) into dynamic sliding systems without introducing foreign elements. The Mimetic Crack (The Verification Gap): Purely generative language models (Copilot, Socratic) act as brilliant semantic reasoners but unreliable mathematical actors due to their lack of symbolic verification engines. This highlights an epistemological gap where "idea generation" is decoupled from "logical validation." *Based on the results of this study, the following recommendations can be made:* For Mathematics Educators and Curriculum Designers: Mathematics textbooks and curriculum structures must move away from rigid, well-structured, routine templates. Integrating non-routine problems that force students to identify contradictions, establish bounds, and propose original conjectures (similar to the AI-generated Stadion geometries) is crucial to fostering genuine 21st-century skills. For Educational Technology Developers: AI tools designed for educational environments must transition to hybrid cognitive architectures. Combining the high-level semantic reasoning and divergent capabilities of LLMs with symbolic verification engines (such as WolframAlpha) or code-execution environments (Python REPL as seen in JuliusAI) is mandatory to prevent misleading hallucinations in learning environments. For Classroom Pedagogy: Teachers should design "AI Critic" (discovering errors in AI outputs) or "AI Relay" (multi-agent cooperative steps) activities. Instead of banning AI, classrooms should leverage AI's divergent fluency while training students to perform the critical "verification and logic checking" steps where LLMs naturally falter.