

YAPAY ZEKA DESTEKLİ İLAÇ KEŞFİ:

Büyük Dil Modelleri ve HPC ile Trilyon Ölçeğinde Sanal Molekül Tarama

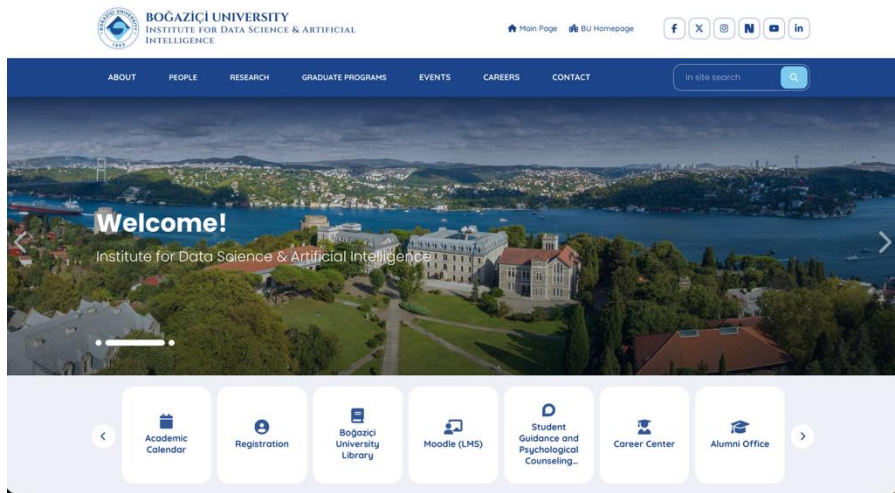
Şener Özönder

ArtificaX,
Kurucu

Veri Bilimi ve Yapay Zeka Enstitüsü,
Boğaziçi University, Faculty
Doç. Dr.



DSAI, Boğaziçi Üniversitesi



dsai.bogazici.edu.tr

BOĞAZIÇI UNIVERSITY
Institute for Data Science & Artificial Intelligence
dsai.bogazici.edu.tr

AI for Drug Discovery and Protein Design

DSAI 583, Spring 2026

Instructor: Assoc. Prof. Şener Özönder (sener.ozonder@bogazici.edu.tr)

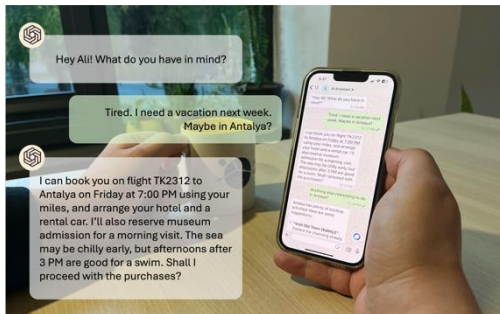
Course Plan:

Week	Subject
1	Small molecules, amino acids and proteins: An Introduction
2	Overview of classical machine learning: Data types, feature space, EDA, linear regression, evaluation metrics, train-val-test split, logistic regression, decision trees, random forests, XGBoost
3	Overview of classical machine learning: Dimensional reduction (PCA and t-SNE), clustering with kmeans, artificial neural networks
4	In silico protein-molecule docking and relevant tools
5	SMILES, molecular descriptors and fingerprints via rdkit
6	Supervised and unsupervised modeling examples and graph neural networks
	Midterm
7	Drug-like properties of molecules, ADMET/QSAR prediction with ML
8	Activity prediction with transformer-based language models: ChemBERTa
9	AlphaFold3: Protein folding (2024 Nobel Prize in Chemistry)
10	De novo protein design with protein language models
11	De novo drug design with generative models and Reinforcement Learning
12	Lead optimization with large language models
13	Current trends in AI for drug discovery
	Final exam

ArtificaX

- Büyük dil modelleri, görüntü işleme ve fiziksel simülasyonları birleştiren çok-kipli yapay zeka çözümleri geliştiriyoruz.

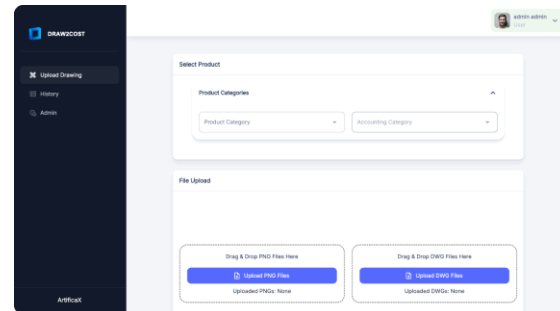
AXAgents



RapidDrug

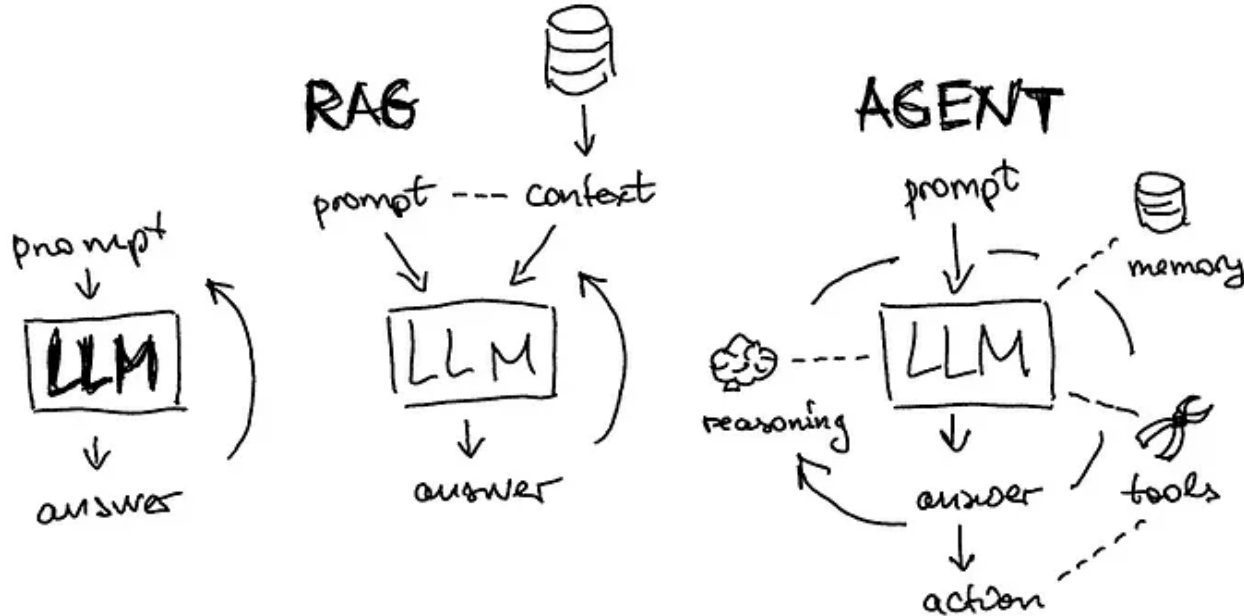


Draw2Cost



Yapay Zeka Ajanları

- Araç kullanan yapay zeka: AI Agents
 - Araçlar: Excel, API/MCP, Gmail, Google Drive, Notion, Youtube, Stripe, vb.

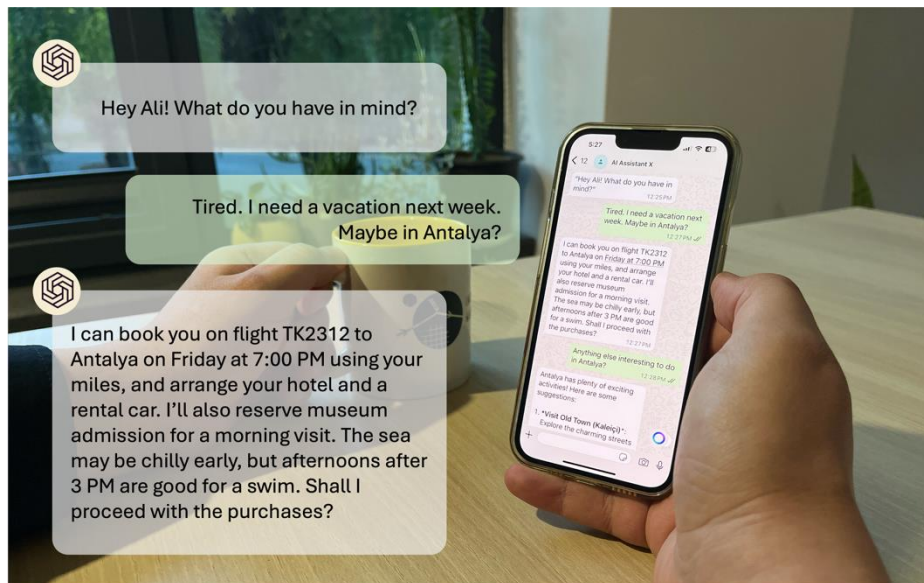


FFPlus & EuroHPC (2025)

We are proudly supported by EuroHPC and FFPlus for our project:
On Device AI Agents: Revolutionizing Real-time API Interactions with
Privacy Preserving Intelligence



ARTIFICAX
TECHNOLOGIES



FFPlus & EuroHPC (2025)

- Yapay zeka ajanlarını eğitmek için,
 - açık kaynaklı LLM'ler ile araç ve API kullanımını simüle eden çok sentetik veri üretilmelidir,
 - LLM bu veri kümesi ile uyarlanmalıdır (fine-tune),
 - bu LLM'leri mobil cihazlarda çalıştırabilmek için damıtma (distillation) yapılmalıdır.
- **LLAMA 405b** ve **QWEN 235b** gibi açık ağırlıklı LLM'leri çıkarım modunda bile çalıştırmak için, düğümlerin InfiniBand (200 Gbps) ağı üzerinden birbirine bağlı olduğu en az 20 adet H100 GPU'ya ihtiyaç vardır. Full fine tune için ise anlık 100'den daha fazla GPU gerekir.
- For this project, BSC Mare Nostrum 5 ACC kümesinde EuroHPC'den 128.000 H100 GPU-saat hesaplama kotası aldık.

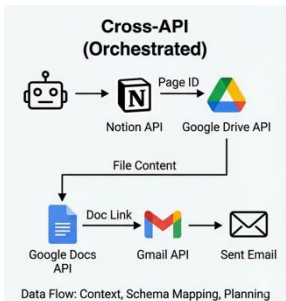


FFPlus & EuroHPC (2026)

- **EuroHPC:** Cross-API Agents for Reliable End-to-End Workflow Automation

- 200.000 H100 GPU-saat (~ €300k)

- **FFplus:** CROSS-AI-AGENTS: Cross-API Agents for Reliable End-to-End Workflow Automation with Reinforcement Learning



FORTISSIMO
PLUS

ARTIFICAX
TECHNOLOGIES



Yapay Zeka ile İlaç Keşfi

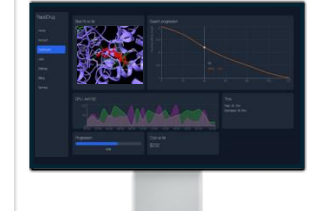
- *MProDockAI: SARS-CoV-2 ana proteazının potansiyel inhibitörlerinin kenetleme, moleküler dinamik ve SMILES transformer yapay zeka modeli aracılığıyla araştırılması*

(EuroHPC, 2025-2026)



- *Yapay Zeka Destekli İlaç Keşfi Platformu Geliştirilmesi*

(TÜBİTAK TEYDEB, 2025-2027)



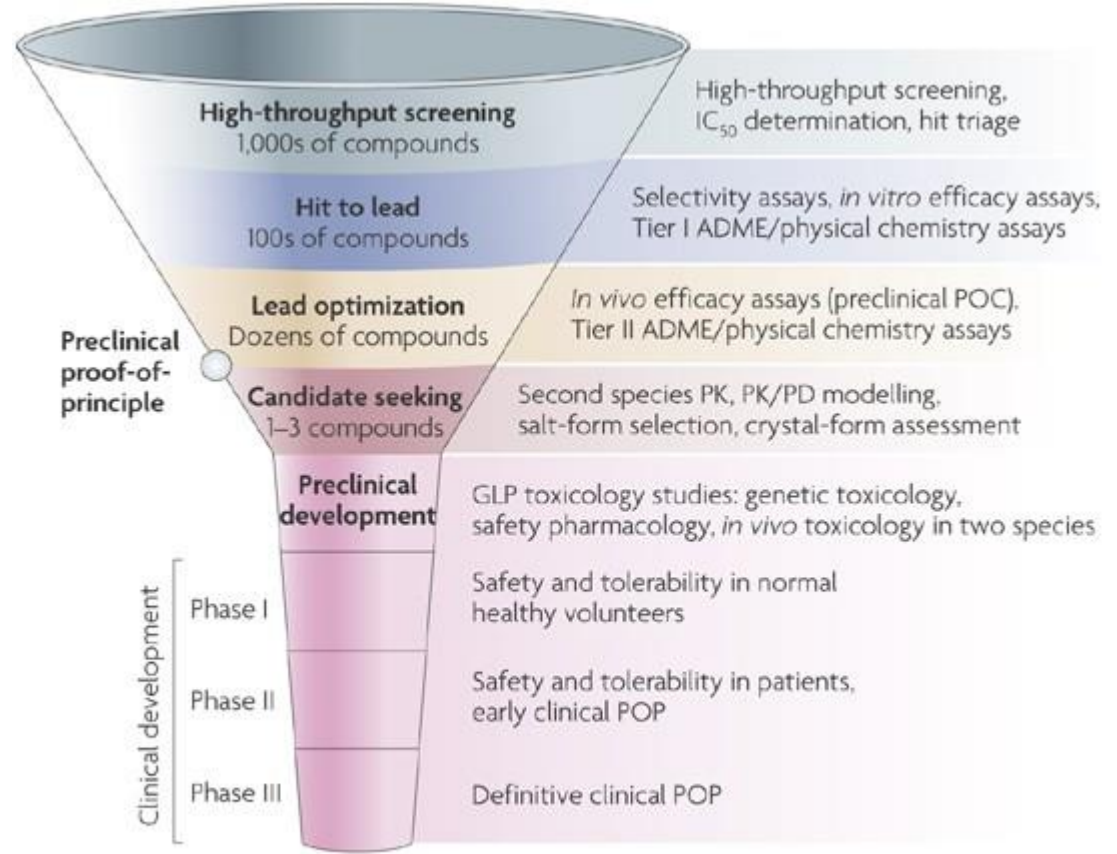
- *EUPilot*

(EU Project, 2024)



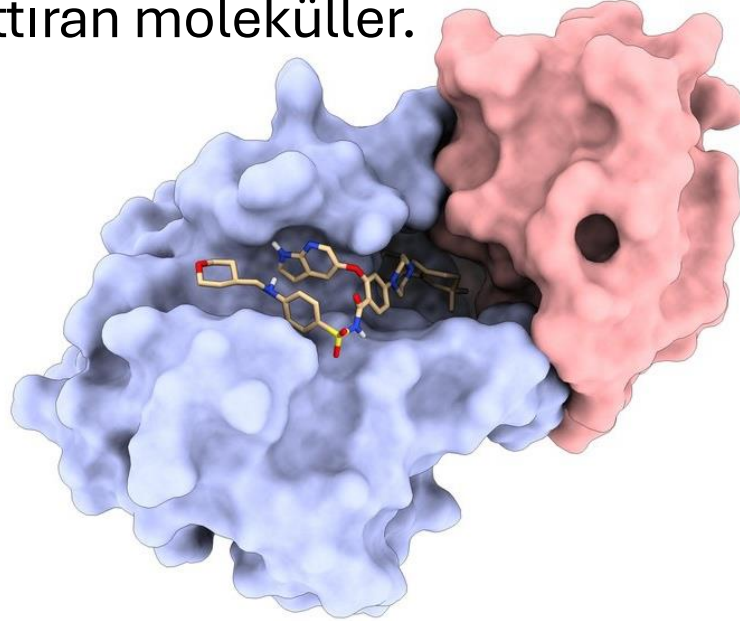
İLAÇ KEŞFİ

- 10-15 yıl
- ~\$2.5b Ar-Ge
- 10% başarı oranı
- Yapay zeka ile high-throughput sanal tarama sürecini ile 2 yıldan 1 aya indiriyoruz.
 - 100 kat daha fazla hit
 - Trilyon seviyesinde kütüphane tarama.

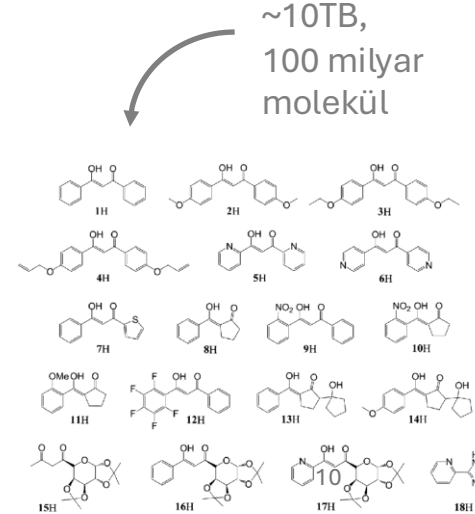


İlaç Keşfi

- İlaç: Protein üstündeki oyuğa uyan ve proteinin fonksiyonunu durduran veya arttıran moleküller.

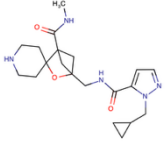


- Sorun: ZINC-22 kütüphanesinde 100.000.000.000 molekül!



Maliyet

- Deneyler pahalı, öncesinde bilgisayarda kenetleme simülasyonları gerekiyor.
- 100 milyar protein-molekül simülasyonu 300 yıl sürüyor.



Z3414521516

Product class Screening Compounds
MW 387.48

Purity 90%

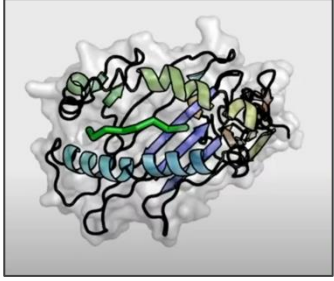
10 mg - \$167.5 - In stock

1

Add to cart



High throughput sanal tarama



Kenetleme simülasyonu



2 dk/kenetleme



0.1 sn/kenetleme

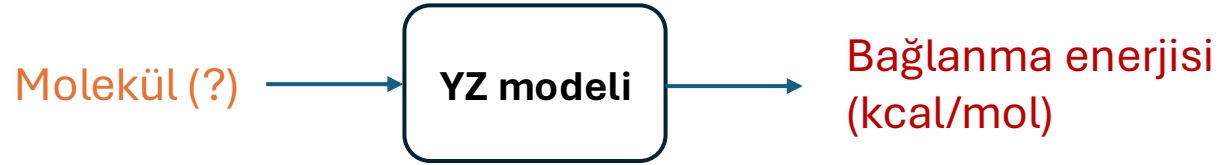


Türkiye'de HPC'ler: ~400 GPUs
(+ BSC ortaklığı).

100 milyar molekül 10 GPU ile
32 yılda kenetleme
simülasyonu ile taranabilir.

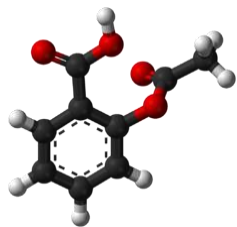
- Bu pratik değil! Yapay zekanın yardımı gerekiyor.

İlaç Keşfi: Yapay Zeka



SMILES büyük dil modelleri

- Moleküllerin SMILES dizgeleri ile kimyasal/biyolojik özellikleri arasında transformer tabanlı bir **büyük dil modeli** eğitebiliriz.



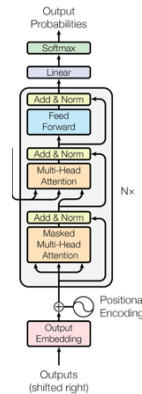
Aspirin

CC(=O)OC1=CC=CC=C1C(=O)O

Aspirin için SMILES dizisi

Bağlanma enerjisi

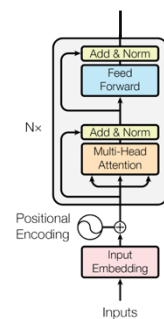
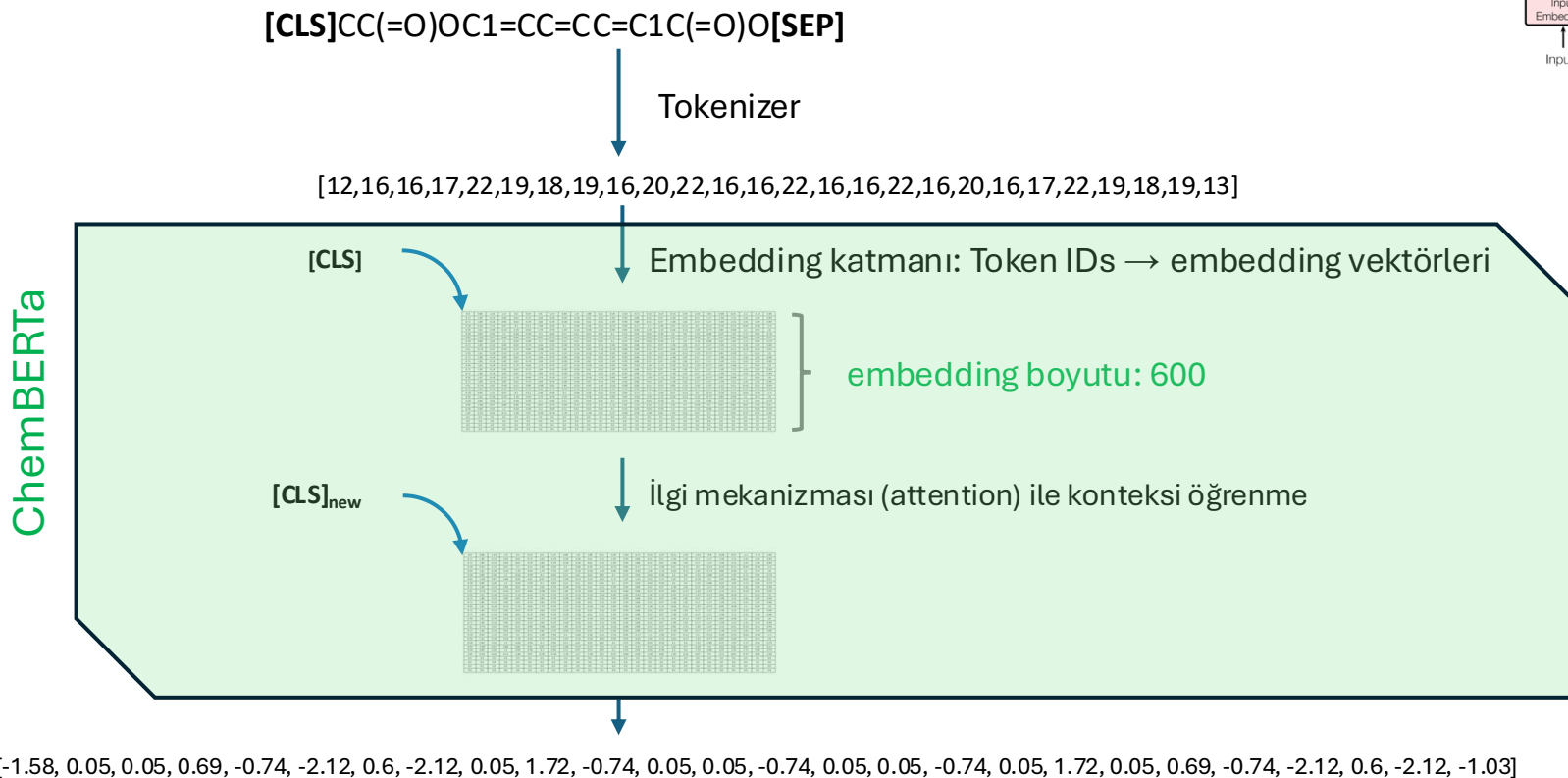
= -11.2 kcal/mol



CC(=O)OC1=CC=CC=C1C(=O)O

ChemBERTa

- ChemBERTa ön eğitilmiş bir encoder-only transformer modelidir.



RapidDrug

