

# YAPAY ZEKA DESTEKLİ İLAÇ

Büyük Dil Modelleri ve HPC ile Trilyon Ölçeğinde

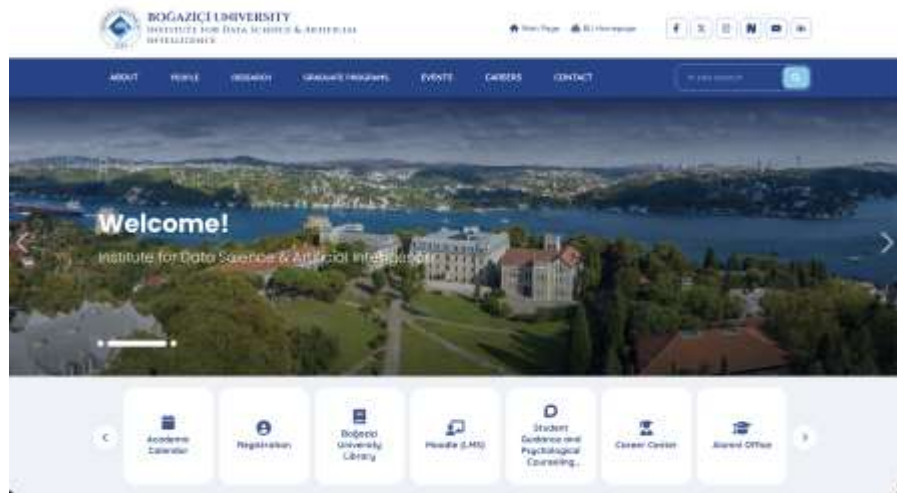
**KEŞFİ:**  
Şenol Özener

**Sanal Molekül Tarama**  
Kurucu

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



# DSAI, Boğaziçi Üniversitesi



[dsai.bogazici.edu.tr](https://dsai.bogazici.edu.tr)

**BOĞAZIÇI UNIVERSITY**  
Institute for Data Science & Artificial Intelligence  
[dsai.bogazici.edu.tr](https://dsai.bogazici.edu.tr)

## AI for Drug Discovery and Protein Design

DSAI 583, Spring 2026

**Instructor:** Assoc. Prof. Şenar Özünder ([senar.ozunder@bogazici.edu.tr](mailto:senar.ozunder@bogazici.edu.tr))

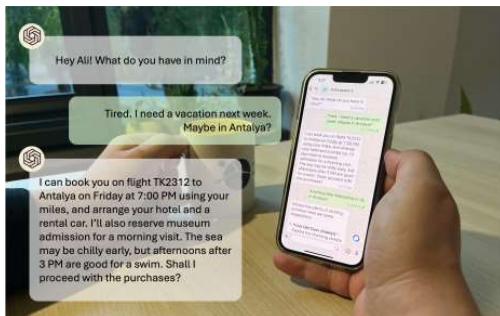
### Course Plan:

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 <i>Midterm</i>
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	<i>De novo</i> protein design with protein language models
11	<i>De novo</i> drug design with generative models and Reinforcement Learning
12	Lead optimization with large language models
13	Current trends in AI for drug discovery
	<i>Final exam</i>

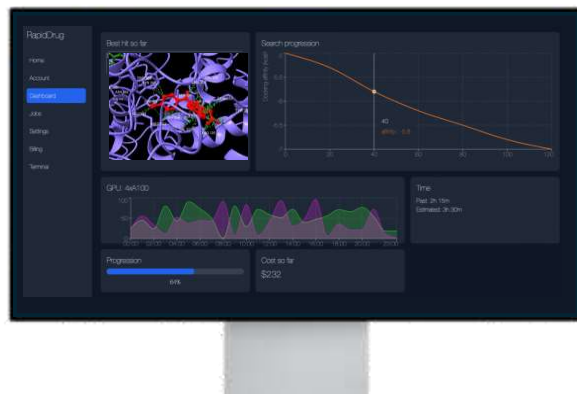
# ArtificaX

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

## AXAgents



## RapidDrug

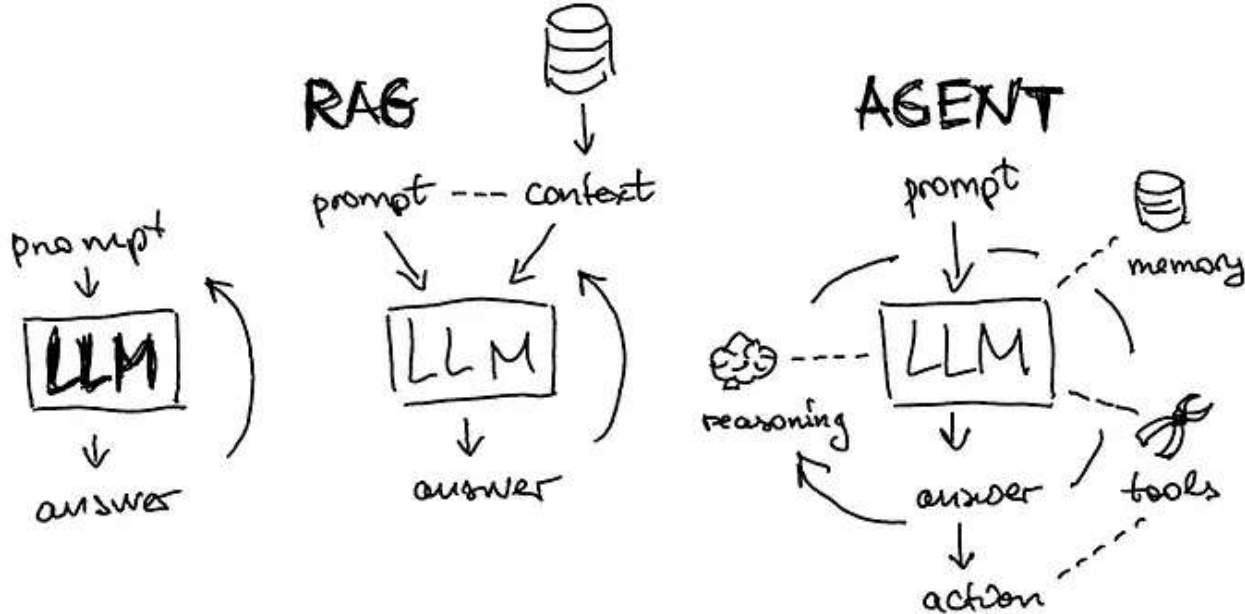


## Draw2Cost



# 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



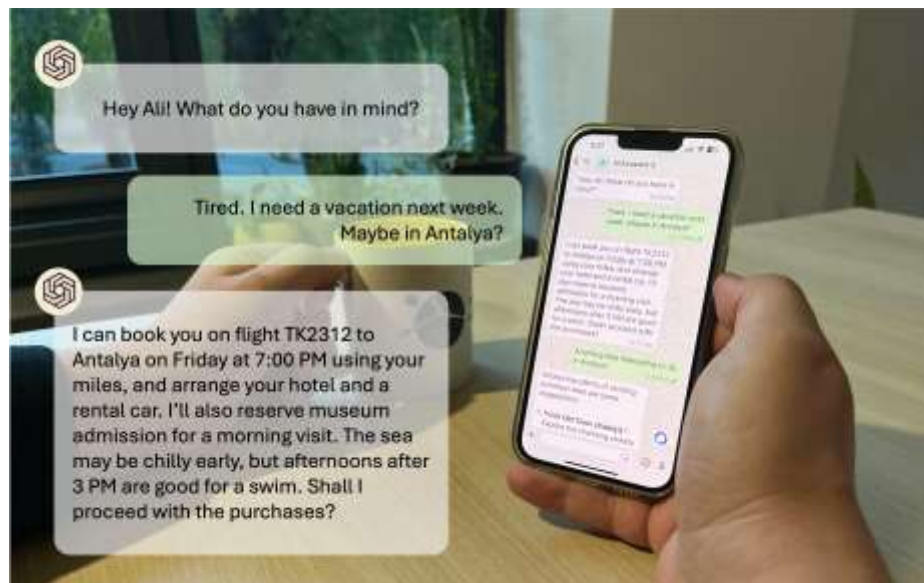
**EuroHPC**  
Joint Undertaking



**FORTISSIMO**  
PLUS

ARTIFICAX  
TECHNOLOGIES

**TURKCELL**  
The picture can't be displayed.



# FFPlus & EuroHPC (2025)

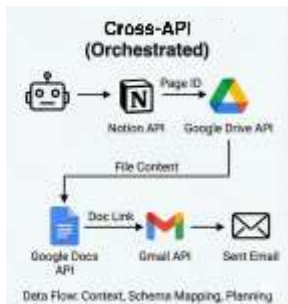
- Yapay zeka ajanlarını eğitmek için,
  - açık kaynaklı LLM'ler ile araç ve API kullanımını simüle eden çok sentetik veri üretilmelidir,
  - LLM bu veri kümesi ile uyarlanmalıdır (fine-tune),
  - bu LLM'leri mobil cihazlarda çalıştırabilmek için damıtma (distillation) yapılmalıdır.
- **LLAMA 405b** ve **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



# 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, 2024)



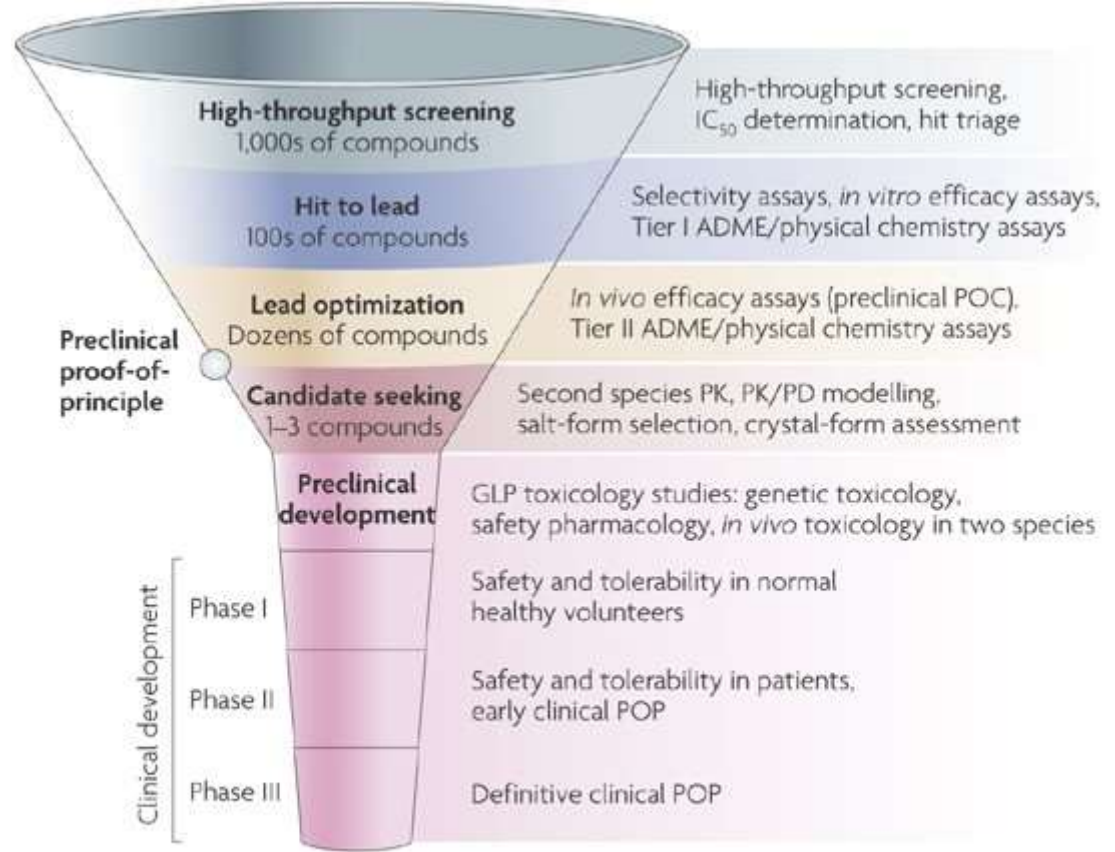
- *EUPilot*

(EU Project, 2024)



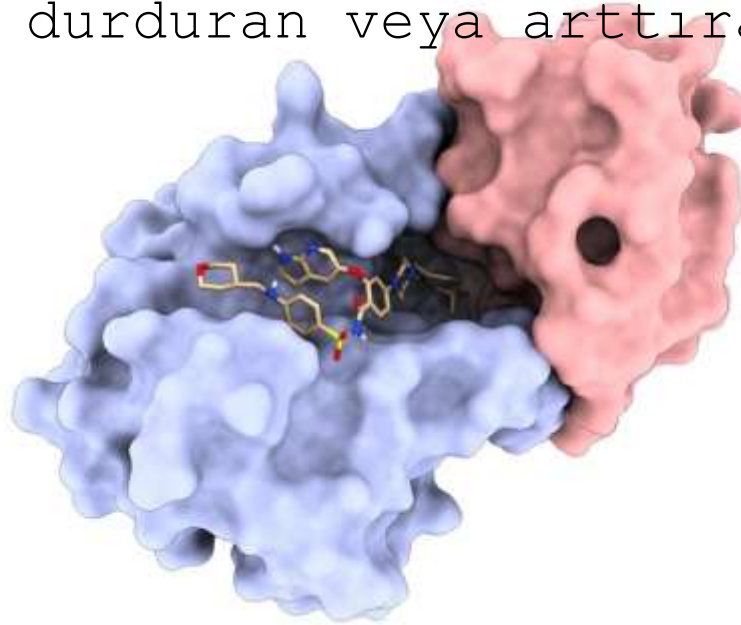
# İLAÇ KEŞFİ

- 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

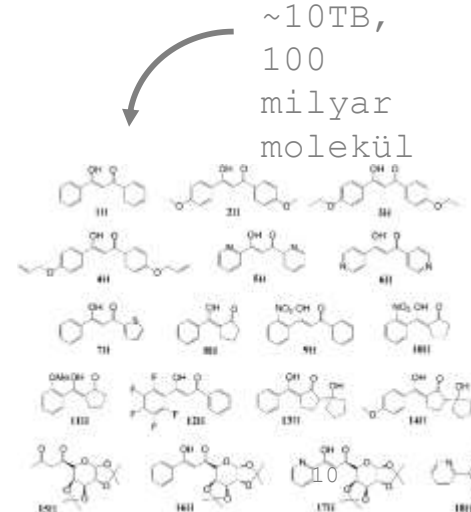


# İlaç Keşfi

- 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: 387.46

Purity: 95%

10 mg - \$167.5 - In stock

1

Add to cart

The image shows a chemical structure on the left and a product information panel on the right. The panel includes the product ID, product class, molecular weight, purity, and a quantity selector.



# High throughput sanal tarama



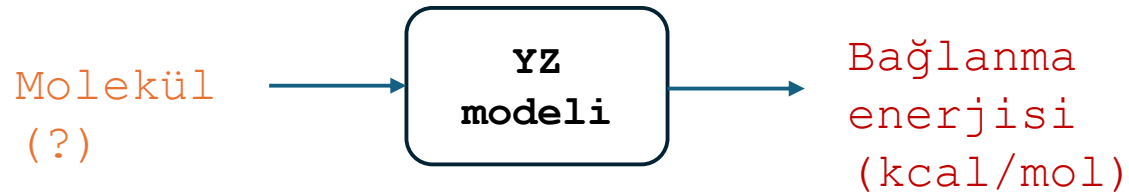
Kenetleme simülasyonu 2 dk/kenetleme 0.1 sn/kenetleme

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

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

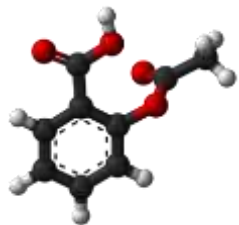
- Bu pratik değil! Yapay zekanın yardımı gerekiyor.

# İlaç Keşfi: Yapay Zeka



# SMILES büyük dil modelleri

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

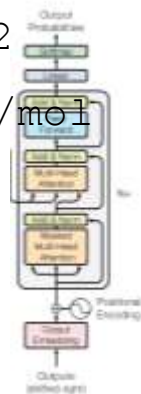


*Aspirin*

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.

**[CLS]CC(=O)OC1=CC=CC=C1C(=O)O[SEP]**

Tokenizer

[12, 16, 16, 17, 22, 19, 18, 19, 16, 20, 22, 16, 16, 22, 16, 16, 22, 16, 20, 16, 17, 22, 19, 18, 19, 13]

ChemBERTa

[CLS]

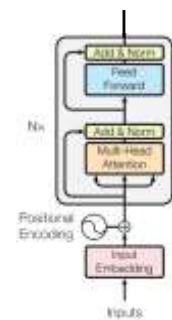
Embedding katmanı: Token IDs → embedding vektörleri

embedding boyutu: 600

[CLS]<sub>new</sub>

İlgi mekanizması (attention) ile konteksi öğrenme

[CLS]<sub>new</sub>=[-1.58, 0.05, 0.05, 0.69, -0.74, -2.12, 0.6, -2.12, 0.05, 1.72, -0.74, 0.05, 0.05, -0.74, 0.05, 0.05, -0.74, 0.05, 1.72, 0.05, 0.69, -0.74, -2.12, 0.6, -2.12, -1.03]



# RapidDrug

