

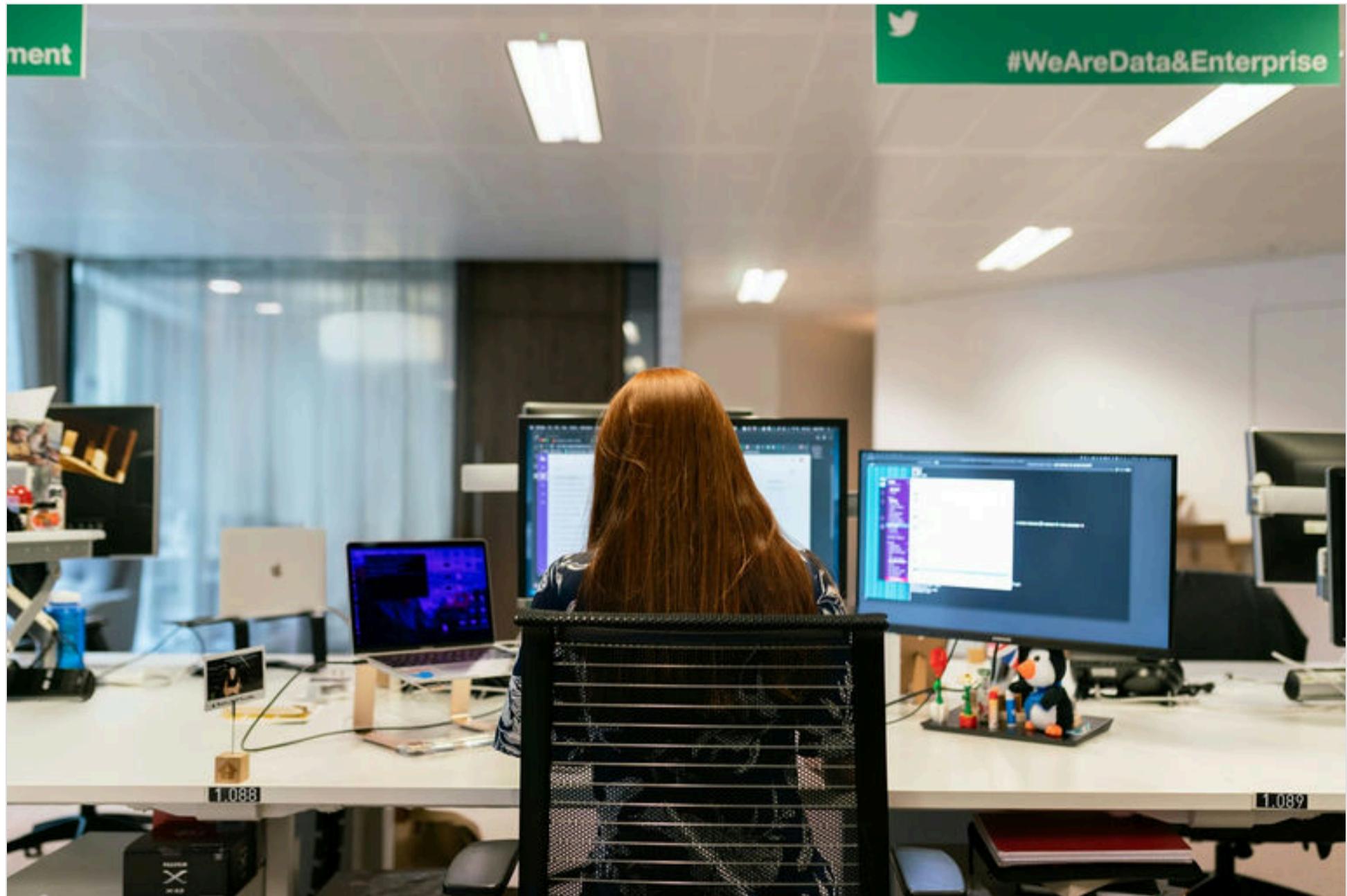
 PaaS for AI Development — And Why AI Development Differs from Traditional Software Development [Read more](#)



■ [Projects / Local Chapter Project](#)

AI Agents Inference Benchmarking Challenge

■ Start Date: December 13, 2024 | 6 months ago



[X Close](#)

Challenge Background

With the exponential growth of AI agents across different frameworks and platforms, the challenge is to identify the key performance factors in real-time applications. AI processing represents less than 10% of overall transaction time in many cases, with most time spent on data preparation and reference data retrieval. The challenge of optimizing inference speed while maintaining accuracy and reliability is a significant one.

¿Listo para mejorar tus operaciones con IA?

maintaining accuracy has become increasingly important, especially since up to 90% of an AI model's life is spent in inference mode.



The Problem

While numerous AI agent frameworks exist, there's no standardized way to compare their inference performance across different scenarios. The research community needs a comprehensive benchmark system that can:

- Evaluate inference speed across different AI agent architectures/Frameworks
- Consider both end-to-end latency and throughput metrics
- Account for various optimization techniques and their impact
- Assess real-world performance under different computational constraints

Goal of the Project

- **Develop Comparison Metrics:** Establish metrics for effectively comparing inference times across different AI agent implementations.
- **Define Scenarios:** Include two distinct scenarios—Simple AI Agent Tasks and Complex AI Agent Tasks—to evaluate performance comprehensively.
- **Framework Comparison:** Conduct comparative analyses between frameworks such as CrewAI, Langchain, LangGraph, Swarn and custom AI agents within the defined scenarios.
- **Parameter Tuning:** Optimize parameters within the frameworks such as CrewAI to enhance performance metrics.
- **Public Leaderboard:** Create and maintain a public leaderboard to facilitate transparent comparisons and track performance across different frameworks.

Project Timeline

- 1
 - **Design Standardized Testing Methodology:** Define protocols for evaluating different AI agent frameworks under consistent conditions.
- 2
 - **Establish Baseline Metrics:** Creating two distinct scenarios—Simple AI Agent Tasks and Complex AI Agent Tasks.
- 3
 - **Build Pipeline for Frameworks:** Create scripts, tasks , agents and CrewAI, Autogen, Langchain, LangGraph, Semantic Kernel, TxTAI |
 - **Run & Test initial task:** Execute tests on selected AI agent frame

¿Listo para mejorar tus operaciones con IA?

4

- **Parameter Tuning:** For all AI Agents frameworks for optimum inference performance.



5

- **Create Visualization Tools (Optional):** Develop dashboards and visualisation interfaces to display benchmarking results clearly and intuitively.
- **Validate Results:** Ensure the accuracy and reliability of the benchmarking tasks through repeated tests and cross-verification.

6

- **Deploy Public Leaderboard or Research Article**
- **Create Comprehensive Documentation:** Develop detailed guides and documentation to help users understand and utilize the benchmarking tasks effectively.

What you'll learn

- Understanding AI inference optimization techniques
- Mastering performance measurement and benchmarking
- Analyzing trade-offs between different AI agent architectures
- Implementing various optimization strategies
- Research Article Publication

First Omdena Local Chapter Project?

- Beginner-friendly, but also welcomes experts
- Education-focused
- Duration: 4 to 8 weeks
- Open-source

Your Benefits

- ✓ Address a significant real-world problem with your skills
- ✓ Build your project portfolio
- ✓ Access paid projects (as an Omdena Top Talent)
- ✓ Get hired at top organizations

¿Listo para mejorar tus operaciones con IA?

Requirements

Good English Suitable for AI/ Data Science beginners but also more senior collaborators

Application Form

Application Closed.

This Challenge is hosted by:



Omdena Knowledge Chapter

[Visit Local Chapter →](#)

Other Projects



[DeepTruth: AI for Authenticity](#)



[LibreTools AI Assistant Plugin](#)



[Building EduFundAI – \(Education + Funding + AI\)](#)

[Back to all Projects →](#)

Become an Omdena Collaborator



[Visit the Omdena Collaborator Dashboard](#)

■ [Learn More →](#)

¿Listo para mejorar tus
operaciones con IA?



Join Omdena today

Let's co-create the AI future,
by the people, for the people.

■ For Organizations > ■ For Developers



■ NAVIGATION	■ COMPANY	■ SOLUTIONS	■ OFFICES & LOCATIONS
Company	About us	Geospatial AI	● New York
AI Solutions	AI Centres of Excellence	NLP	● Warsaw
For Businesses	Innovation Challenges	Computer Vision	● Dublin
For Developers	Top Talents	Data Engineering	● Cologne
Projects	Collaborator Dashboard	Generative AI	■ SOCIAL LINKS
Resources	FAQs		in LinkedIn
AI Program	Omdena Policies		f Facebook
Career Paths	Testimonials		t Twitter
	Privacy Policy		ig Instagram
	Terms and Conditions		

Omdena Inc. | © 2025 Copyright

¿Listo para mejorar tus operaciones con IA?