Ranking and Measuring the Efficiency of AI Tools using Data Envelopment Analysis

Dr. Shivani Guru
Assistant Professor
Mr. Sanskar Sahu
Student
Amity Business School
Amity University, Raipur, Chhattisgarh

Abstract
The study evaluates the technical efficiency of ten leading Artificial Intelligence (AI) tools using Data Envelopment Analysis (DEA), a non-parametric operations research methodology. DEA assesses the relative performance of AI platforms using inputs like model complexity and computational demands against outputs such as response accuracy and user utility. An input oriented Constant Returns to Scale (CCR) model and a Super-Efficiency DEA model were employed to identify and rank efficient tools. The findings reveal that eight of the ten AI tools are technically efficient, with Claude AI emerging as the most super-efficient. The research provides performance benchmarks and managerial insights, facilitating strategic decision making in AI tool adoption and development.
Keywords: DEA, AI Tools, DEA Super efficiency, Ranking, Benchmarking

1.0 Introduction

The swift evolution of Artificial Intelligence (AI) has precipitated the emergence of a plethora of AI tools and platforms, each purporting to provide superior performance, precision, and user satisfaction. These tools are used for many purposes, encompassing healthcare, education, finance, and customer service. Nevertheless, as the marketplace becomes inundated with AI solutions, it is imperative for stakeholders—including developers, enterprises, and policymakers—to assess and juxtapose the efficacy of these tools in a systematic and objective manner. Conventional benchmarking methodologies frequently prove inadequate in encapsulating the multi-faceted nature of AI performance, which necessitates the reconciliation of intricate inputs such as algorithmic complexity and computational requirements with diverse outputs such as predictive accuracy, adaptability, and user experience.

In this framework, Data Envelopment Analysis (DEA) provides a rigorous and analytical paradigm for evaluating the relative efficiency of AI tools. DEA is a non-parametric approach used in operations research to measure the efficiency of decision-making units (DMUs). By implementing DEA in the context of AI tools, researchers can discern which platforms yield optimal outputs while minimizing resource utilization, thereby furnishing invaluable insights for efficiency-centric decision-making.
This research employs two DEA models—the input-oriented Constant Returns to Scale (CCR) model and the Super-Efficiency model—to assess and rank ten distinguished AI tools. The inputs analyzed consist of model complexity and computational requirements, whereas the outputs are quantified in terms of response accuracy and user utility. The investigation demonstrates that eight out of the ten tools exhibit technical efficiency, with Claude AI surpassing the others in terms of super-efficiency. These results not only underscore the strengths

and limitations of contemporary AI platforms but also establish a quantitative basis for forthcoming research and strategic planning within the domains of AI development and implementation.

Through this inquiry, we aspire to bridge the chasm between the perception of AI performance and quantifiable efficiency, thereby providing a meticulous and actionable evaluation framework for the AI ecosystem. This study applies Data Envelopment Analysis (DEA), a mathematical approach used to evaluate the efficiency of Decision-Making Units (DMUs), to benchmark ten top AI tools. DEA is particularly suited to this analysis due to its ability to handle multiple input-output variables without requiring predefined functional relationships.

2. Literature Review

Existing research proves the efficiency of DEA in various disciplines like healthcare, finance, education, and public policy. **Camanho et, al., 2024** shows Data Envelopment Analysis estimations of economic efficiency, from methodological advances to empirical implementations. **Fakhri, Bahri and Wulan, 2023** seeks to find and examine the efficient and effective value of zakat, ZIS fund management included in LAZNAS in Indonesia during 2017-2021**. Saini, Truong and Pan, 2023** They devised a two-phase, two-stage DEA model, and through linear programming, evaluated the relative operational efficiencies of 13 airlines over the period 2013 to 2015, taking into account the financial metrics of the airlines and the environmental performance measured by the operational carbon emissions of the airline's operations. **Hao et al., 2023** assesses the productivity of different clinical units attending to patients in a more advanced stage Chinese class A tertiary public hospital (Hospital M) to evaluate the distribution of hospital resources within these units and assist in management decision making.

Sun et al., 2022 attempted to identify the correlation between the clinical pharmacy workload and educational background of clinical pharmacists in China. There were 625 clinically trained pharmacists from 311 tertiary level Chinese hospitals who completed the survey. Expanded some elements of education or training for clinical pharmacists in tertiary hospitals in China could increase their capability to conduct clinical pharmacy services. **Park, Yoon & Lee, 2021** had as an objective the elaboration of a plan for effective management of a Healthy City Project aimed at local governments and tailored to their level. Zykova, 2019 aims for the system to establish an overall winner and rank all candidates. Gregorious, 2006 is investigating the trading efficiency of leading commodity trading advisors (CTAs) from January 1998 to June 2004. **Park et al. (2021)** used DEA to optimize local government health projects, while **Fakhri et al. (2023**) applied it to assess the efficiency of charitable organizations. In AI applications, DEA remains underutilized despite its potential to assess resource-intensive technologies. This study addresses this gap by applying DEA to AI tools, highlighting their performance efficiency in a data-driven framework.

3.0 Research Objectives

Based on Literature we come up with five objectives

I. To measure the efficiency and rank the AI tools using Data Envelopment Analysis.
II. To benchmark the AI tools.
III. To rank the most Efficient AI tools using DEA Super efficiency model.

4.0 Research Methodology

This research utilizes a quantitative methodology, specifically implementing Data Envelopment Analysis (DEA), to assess and juxtapose the performance efficiency of artificial intelligence instruments. The input-

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 06 | June – 2025 DOI: 10.55041/ISJEM03920
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

oriented Charnes-Cooper-Rhodes (CCR) model, which operates under the assumption of constant returns to scale, is employed in conjunction with the Super-Efficiency model to systematically rank Decision-Making Units (DMUs) that attain technical efficiency scores of unities. A cohort of ten AI instruments—ChatGPT, Claude AI, Google Gemini, Microsoft Copilot, DeepSeek AI, Meta AI, Copy AI, Perplexity AI, Writesonic AI, and Rytr AI—was curated based on their prevalence, functionality, and the accessibility of pertinent data. Each AI instrument is conceptualized as a distinct DMU. Secondary data were amassed from technical whitepapers, performance assessments, scholarly articles, and industry benchmarking reports. The input variables incorporated into the DEA model encompass the quantity of monthly active users, monthly subscription fees, word limit per output response, and average response duration measured in seconds. The output variables consist of user satisfaction, gauged from a sample of 50 respondents, and the precision of content. The DEA Frontier software was employed to scrutinize the data and derive the efficiency scores and hierarchical rankings of the AI instruments.

5.0 Analysis And Interpretation

5.1 Application Of DEA

To order the AI tools, we apply input-oriented DEA analysis for CCR model. In the study, since input-oriented approach is adopted, emphasis is placed on minimizing the input to the largest extent possible either by keeping the output unchanged or by its increase. The AI tools having efficiency score 1.000 is taken as efficient, the AI tools having efficiency score below 1 indicate that they are having scope for improvement and if output is kept constant inputs could be reduced further in turn better efficiency could be increased. Input-oriented efficiency score, ranking results and benchmarking results of various AI tools are demonstrated in Figure (1) and Table (1).

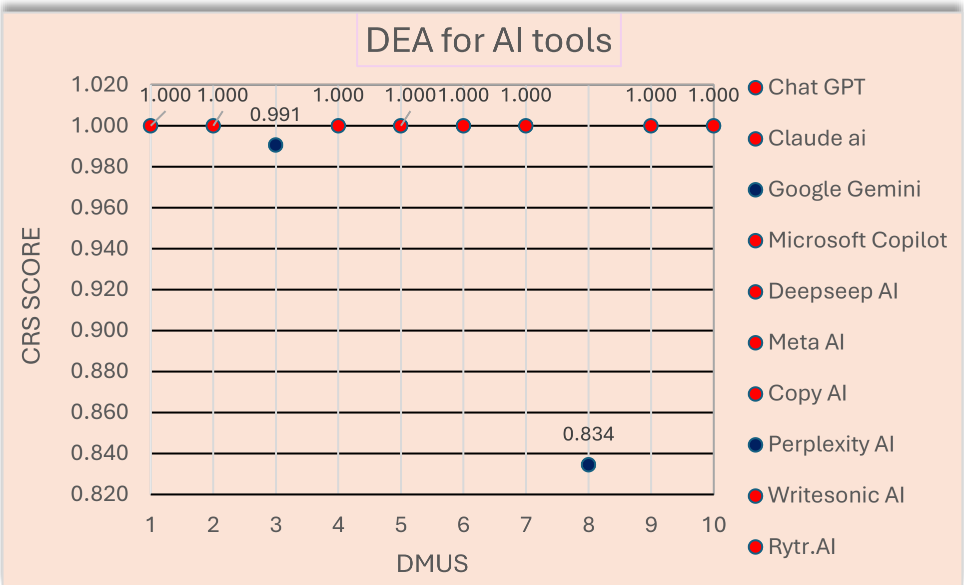

Fig 1: CRS Benchmarking

Table 1: Performance Measurement and Benchmarking Result of DEA- CRS technique

DMUs	DMUs Name	CRS Efficiency	Rank	Bench Mark	Target Value (λ)
1.	Chat GPT	1.00000	1	1	
2.	Claude ai	1.00000	1	1	
3.	Google Gemini	0.99051	2	1,6	0.93002, 0.07410
4.	Microsoft Copilot	1.00000	1	1	
5.	Deepseep AI	1.00000	1	1	
6.	Meta AI	1.00000	1	1	
7.	Copy AI	1.00000	1	1	
8.	Perplexity AI	0.83445	3	2,4,5,9	0.16212,0.18368,0.00159,0.77017
9.	Writesonic AI	1.00000	1	1	
10.	Rytr.AI	1.00000	1	1	

Source: Own Calculation using DEA Frontier.

Interpretation 1:

DEA result also provides the benchmark for other DMU. Table 1 shows efficiency score, ranking and benchmarking results. Figure 1 shows under CRS model eight AI tools are with efficiency score 1. Efficient AI tools do not have to refer any other but for inefficient DMUS their benchmark are efficient DMUs (may be one or more).

Benchmarks for DMU 3 are DMUs 1, and 6. To turn into efficient, DMU 3, use a combination of two DMUs 1, and 6. DMU 3 to become efficient like DMUs 1, and 6 will have to reduce one of its input either to 0.93002 or 0.07410 or make a combination, as it observed from λ weight DMU 3 will prefer to become more like DMU 6 than DMUs 1 ($\lambda 1$=0.93002, $\lambda 6$=0.07410)

Benchmarks for DMUs 8 are DMUs 2,4,5 and 9. To turn into efficient, DMU 8 use a combination of the four DMUs 2,4,5 and 9. DMU 8 to become efficient like DMUs 2,4,5 and 9 will have to reduce one of its input either to 0.16212, or 0.18368, or 0.00159, or 0.77017 or make a combination, as it observed from λ weight DMU 8 will prefer to become more like DMU 5 than DMU 2, 4 and 9 ($\lambda 2$=0.16212, $\lambda 4$ =0.18368 , $\lambda 5$=0.00159, $\lambda 9$ = 0.77017).

Table 2: Result of DEA Super efficiency

DMU#	**DMU Name**	CRS Super efficiency	Rank
1	Chat GPT	1.765	4
2	Claude ai	3.283	1
3	Google Gemini	0.991	9
4	Microsoft Copilot	1.598	5
5	Deepseep AI	1.070	7
6	Meta AI	1.110	6
7	Copy AI	1.063	8
8	Perplexity AI	0.834	10
9	Writesonic AI	1.878	2
10	Rytr.AI	1.814	3

Source: Own Calculation using DEA Frontier

Interpretation 2:
From figure 1 it clear that the larger the number of AI tools is with efficiency score 1, out of these numerous efficient AI tools study attempt to get most efficient AI tools using super-efficiency DEA model. Analysis of DEA super efficiency model shows that there is one super-efficient AI tool, super efficiency score is shown in Table 2 for CRS model. The analysis interprets that under inputs-oriented CRS super efficiency model Claude AI arises as super-efficient.

6.0 Conclusion

This investigation highlights the increasing imperative for the adoption of robust, data-centric methodologies to assess the efficacy of artificial intelligence tools within a progressively congested technological milieu. As artificial intelligence persistently transforms various sectors, including healthcare and finance, the necessity to evaluate these instruments beyond mere superficial performance assertions becomes increasingly paramount. Through the application of Data Envelopment Analysis (DEA), particularly the input-oriented CCR and Super-Efficiency models, this inquiry provides an objective and replicable framework to juxtapose the relative efficiency of ten prevalent AI platforms.
The results indicate that eight out of the ten AI tools attain technical efficiency, with Claude AI distinguished as the most super-efficient. These findings not only substantiate the applicability of DEA in benchmarking AI systems but also furnish developers, users, and policymakers with a lucid understanding of which tools yield optimal performance while minimizing resource expenditure. Moreover, this research addresses a notable deficiency in the literature, wherein DEA remains insufficiently leveraged in the assessment of artificial intelligence, notwithstanding its demonstrated efficacy in other domains.
In summation, this study presents a groundbreaking model for evaluating AI performance that aligns complexity, computational expenditure, and user-centered outcomes. By incorporating DEA into the strategic formulation and development of AI tools, stakeholders are enabled to make more informed decisions that enhance innovation, user satisfaction, and operational efficacy throughout the digital ecosystem. Subsequent research may expand this framework to encompass dynamic DEA models, cross-sectoral comparisons, or integrate qualitative aspects such as ethical considerations and transparency to further enrich the assessment of artificial intelligence technologies.

3. Findings

High Efficiency Among Most Tools: Among the ten AI instruments scrutinized, eight (comprising ChatGPT, Claude AI, Microsoft Copilot, DeepSeek AI, Meta AI, Copy AI, Writesonic AI, and Rytr AI) attained a technical efficiency score of 1.000 according to the DEA-CCR model, signifying optimal utilization of resources to generate desired outputs.

Super-Efficiency of Claude AI: When assessed through the Super-Efficiency DEA model, Claude AI emerged as the foremost performer, setting itself apart from its technically efficient counterparts by delivering enhanced outputs with relatively diminished input requirements.

Performance Gaps in Select Tools: Google Gemini and Perplexity AI exhibited inferior efficiency scores (0.99051 and 0.83445 respectively), which underscores substantial potential for improvement in aligning their input expenditures with resultant performance outputs. These instruments necessitate optimization strategies to attain the efficiency frontier.

Benchmarking Insights: The outcomes of the DEA analysis elucidate efficient tools that can serve as performance benchmarks for their less efficient counterparts. For instance, Google Gemini could adopt the input-output methodologies exemplified by ChatGPT and Meta AI to bolster its efficiency.

Diverse Input and Output Metrics: The investigation adeptly amalgamated multifaceted input metrics (such as user base, subscription cost, response time, etc.) with output metrics (including user satisfaction and precision), thereby illustrating DEA's efficacy in managing intricate performance evaluations within the realm of AI.

4. Suggestions

Optimize Resource Utilization for Inefficient Tools: AI tools such as Google Gemini and Perplexity AI ought to undertake internal assessments aimed at diminishing computational demands or augmenting algorithmic precision, leveraging strategies employed by their efficient equivalents, including Claude AI and ChatGPT.

Adopt DEA for Continuous Benchmarking: Organizations engaged in the development or deployment of AI should incorporate DEA into their performance management frameworks to facilitate the continuous monitoring of efficiency and the corresponding adjustment of operational strategies.

Expand Performance Metrics: Future evaluations should integrate additional qualitative factors such as explainability, ethical congruence, and user privacy to furnish a more holistic assessment of efficiency.

Policy and Regulatory Support: Policymakers ought to contemplate the incorporation of DEA based assessments within AI certification or compliance frameworks to ensure the deployment of resource-efficient and user-centric AI across various sectors.

Encourage Transparent Reporting: Developers should publicly disclose critical performance metrics and resource utilization data, thereby enabling independent benchmarking and informed decision-making by users and enterprises.

Periodic Reassessment with Updated Data: Considering the swift advancements in AI technology, DEA-based evaluations should be periodically revisited to accurately reflect shifts in technological capabilities, usage patterns, and market dynamics.

References

- Camanho, A. S., Silva, M. C., Piran, F. S., & Lacerda, D. P. (2024). A literature review of economic efficiency assessments using Data Envelopment Analysis. In European Journal of Operational Research (Vol. 315, Issue 1, pp. 1–18). Elsevier B.V. https://doi.org/10.1016/j.ejor.2023.07.027
- Fakhri, M. Y., Bahri, E. S., & Wulan, M. (2023). Efficiency Of Zakat Institutions In Indonesia During 2017-2021: An Application Of Data Envelopment Analysis. Jurnal Akuntansi Dan Keuangan Islam, 11(1), 101–135. https://doi.org/10.35836/jakis.v11i1.456
- Gregoriou, G. N. (2006). Trading efficiency of commodity trading advisors using Data Envelopment Analysis. Derivatives Use, Trading & Regulation, 12(1). https://doi.org/10.1057/palgrave.dutr.1840044
- Hao, X., Han, L., Zheng, D., Jin, X., Li, C., Huang, L., & Huang, Z. (2023). Assessing resource allocation based on workload: a data envelopment analysis study on clinical departments in a class a tertiary public hospital in China. BMC Health Services Research, 23(1). https://doi.org/10.1186/s12913-023-09803-y
- Risk Park, S., Yoon, K., & Lee, M. (2021). Healthy City Project: An application of data envelopment analysis. Management and https://doi.org/10.2147/RMHP.S325825 Healthcare Policy, 14, 4991–5003.
- Saini, A., Truong, D., & Pan, J. Y. (2023). Airline efficiency and environmental impacts – Data envelopment analysis. International Journal of Transportation Science and Technology, 12(2), 335–353. https://doi.org/10.1016/j.ijtst.2022.02.005
- Sun, Q., Chen, L., Huang, Y., & Xi, X. (2022). Association of education background with clinical pharmacists' clinical pharmacy workload in tertiary hospitals of China. BMC Medical Education, 22(1). https://doi.org/10.1186/s12909-022-03859-w
- Zýková, P. (2019). Comparison between data envelopment analysis models with penalties. Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 67(6), 1665–1669. https://doi.org/10.11118/actaun201967061665