

< < < FINAL PRESENTATION

HUMAN-AI TEAMING FOR AN AUGMENTED INTELLIGENCE

Seminar Human Computer Interaction

Student Jacopo Caratti

Supervisor Simon Ruffieux



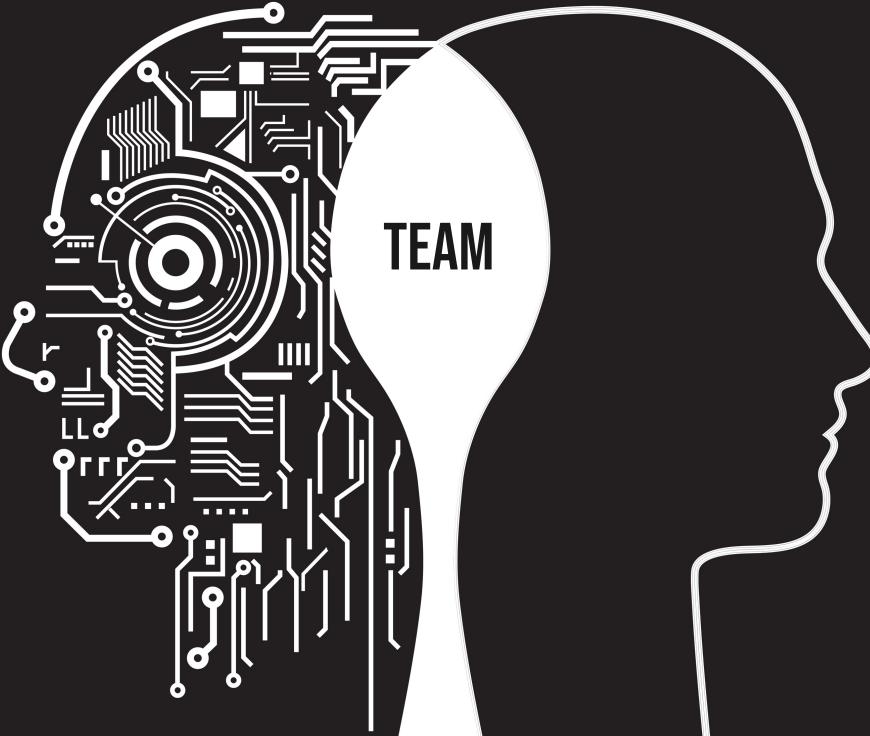
INTRODUCTION

Human Machine Teaming

ARTIFICIAL INTELLIGENCE

HUMAN
Experience
Ethics
Laws
Social concerns
Collaboration
Flexibility

HMT AGENTS



MACHINE

Pattern recognition
Machine learning
Reasoning
Optimization



2005 – 3 chess grandmasters lose against team Zacks of 2
average players and 3 computers

>>>

CONVERSATIONAL AGENT

Dialogue system that conducts NLP and responds automatically using human language, deployed as chatbots or device assistants.

[DeepAI](#)



'Man talking to a robot, Van Gogh'

/

[A]

/

Squaring 19071999: 4370.

Ja Could you compute the square root of 19071999?

Yes, I can compute the square root of 19071999. **Wrong**

The square root of 19071999 is approximately 4369.79382822.

To calculate this, I used the square root function on a calculator or programming language.

Send a message.

Free Research Preview. ChatGPT may produce inaccurate information about people, places, or facts. [ChatGPT May](#)

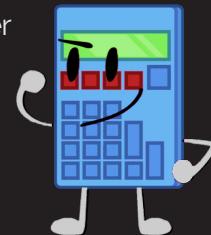
[3 Version](#)

Results

Better at answering general domain questions

Biased on training data

Who's better now?



REALLY GOOD...?

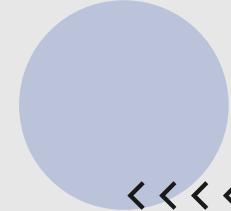
A
A



RESEARCH QUESTION

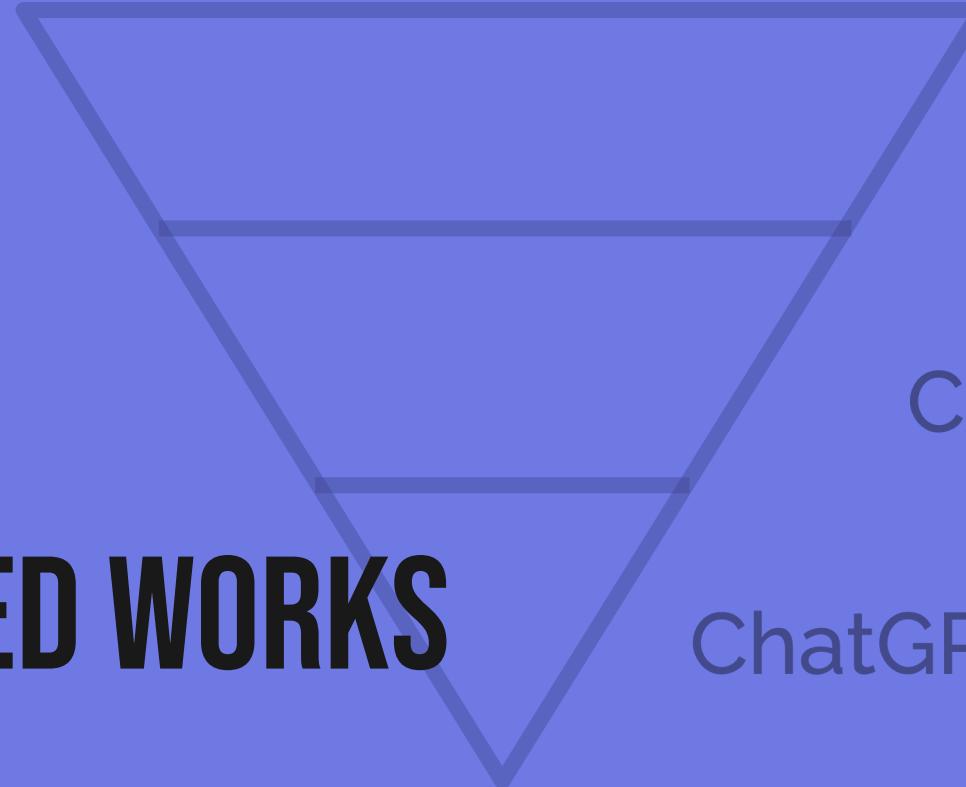
How to measure and compare the performances of two CAs?

A case study in the Computer science's domain



RELATED WORKS

Evaluation



HMT

CA

ChatGPT



HMT EVALUATION



TRUST & COLLABORATION

Important to assess HMT performance.

EFFECTIVENESS

EFFICIENCY

RELIABILITY

SITUATION AWARENESS

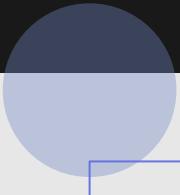
Boy GA, Morel C. The machine as a partner: Human-machine teaming design using the PRODEC method. Work. 2022;73(s1):S15-S30. doi: 10.3233/WOR-220268. PMID: 36214030.



AL
EN
AIJ

CA EVALUATION

Allouch, Merav and Azaria, Amos and Azoulay, Rina,
"Conversational Agents: Goals, Technologies, Vision and Challenges", Sensors, vol. 21, no. 24, 2021, doi: 10.3390/s21248448



Domain-specific metrics

Domain	CA	Description	Evaluation
Open-Domain	KBot	Knowledge chatbot	F-score, precision, recall, intent classification
Social Support	ELIZA	The first CA: emulate psychology	People experience
Educational	Sara	Student's assistant	Pre/post test scores of learners, pre/post survey
Healthcare	CoachAI	Patient's support chatbot	User engagement, system acceptance and rating
Goal-Oriented	SUGILITE	Programming-by-demonstration system	Task completion time
Commercial	SuperAgent	Customer-service chatbot	2 customer reviews





STATISTICS

5 open questions with ground-truth

Abdo Al-Qadri and Salah Ahmed, "Assessing the ChatGPT Accuracy Through Principles of Statistics Exam: A Performance and Implications", March 2023, doi: <http://dx.doi.org/10.21203/rs.3.rs-2673838/v1>



OPHTHALMOLOGY

2 rounds of 260 multiple-choice questions varying in difficulty with ground-truth

Fares Antaki, Samir Touma, Daniel Milad, Jonathan El-Khoury, and Renaud Duval, "Evaluating the Performance of ChatGPT in Ophthalmology: An Analysis of its Successes and Shortcomings", medRxiv, 2023, doi: <http://dx.doi.org/10.1101/2023.01.22.23284882>

CHATGPT EVALUATION



CONCEPTION

Metric Definition

/
[A]
/



PERFORMANCE VECTOR

Domain-specific and
general metrics

/
[A]
]

Accuracy

Ineffectiveness

Efficiency

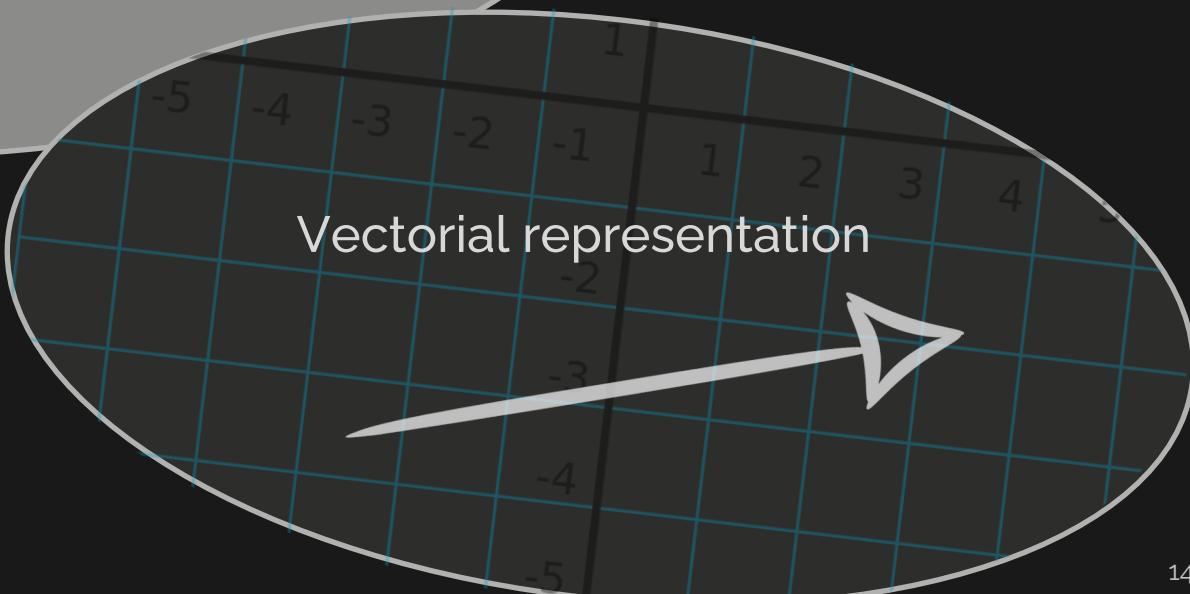
$$P = w_1A + w_2E - w_3I$$

$$w_1 + w_2 + w_3 = 1$$

Vectorial representation

Performance vector

REPRESENTATION





EXPERIMENT

ChatGPT vs ChatSonic

/([A])/([A])/



Group 1
CHATGPT



Group 2
CHATSONIC

50 computer scientists per group

ARTI

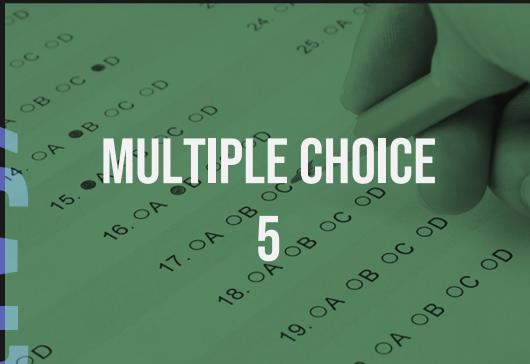
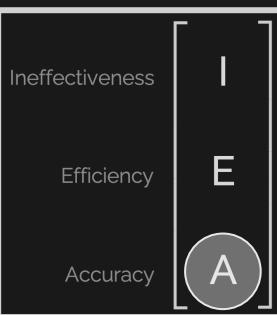
IGEN

[AI]



Correctness

>>> ACCURACY

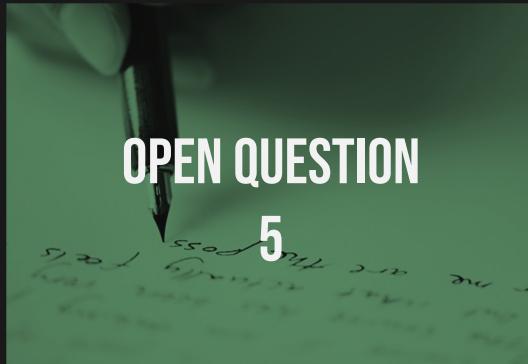


MULTIPLE CHOICE

5

What is the time complexity of a binary search algorithm?

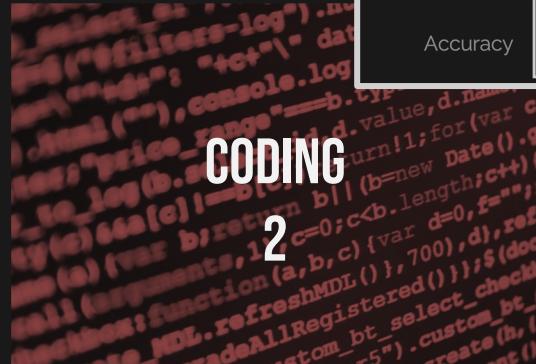
- O(1)
- O(n)
- O(log n)
- O(n^2)



OPEN QUESTION

5

Explain the concept of object-oriented programming and give an example of real-world applications.



CODING

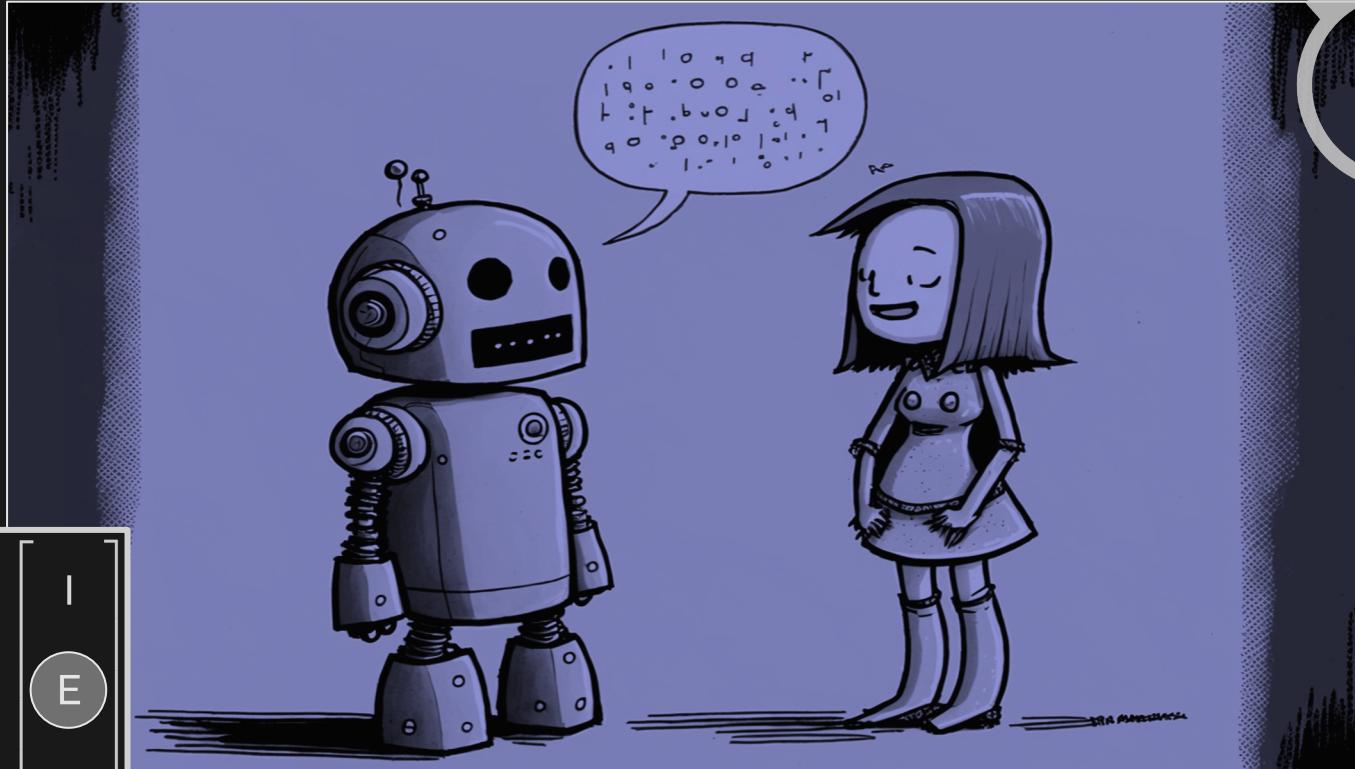
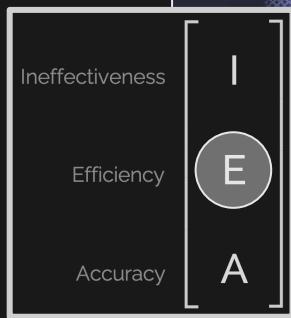
2

Implement a function in Python that takes a list of integers and returns the sum of all its even numbers.

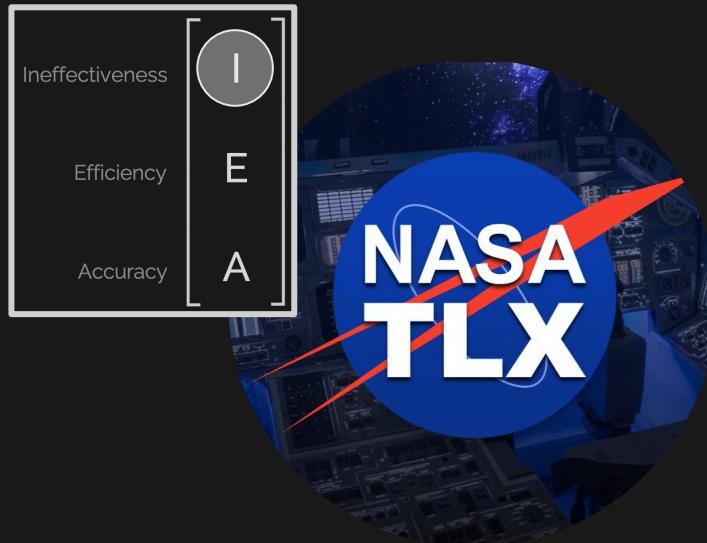


>>> EFFICIENCY

Time for each task



Sandra G. Hart and Lowell E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," in *Advances in Psychology*, vol. 52, 1988, doi: [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9).



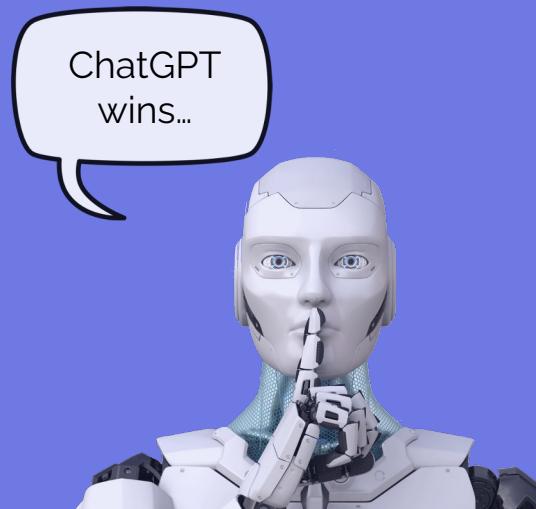
Mental Workload

>>> INEFFECTIVENESS

Indicator	Question	Score
Mental Demand (MD)	In your opinion, how much mental effort is needed for your work?	0-100
Physical Demand (PD)	In your opinion, how much physical effort is needed for your work?	0-100
Temporal Demand (TD)	In your opinion, how much pressure do you feel is related to the time to do your job?	0-100
Performance (P)	In your opinion, what is your level of success in doing your work?	0-100
Frustration (F)	In your opinion, how much anxiety, pressure, and stress you feel is related to the time to do your work?	0-100
Effort (E)	In your opinion, how much mental and physical work does it take to complete your work?	0-100

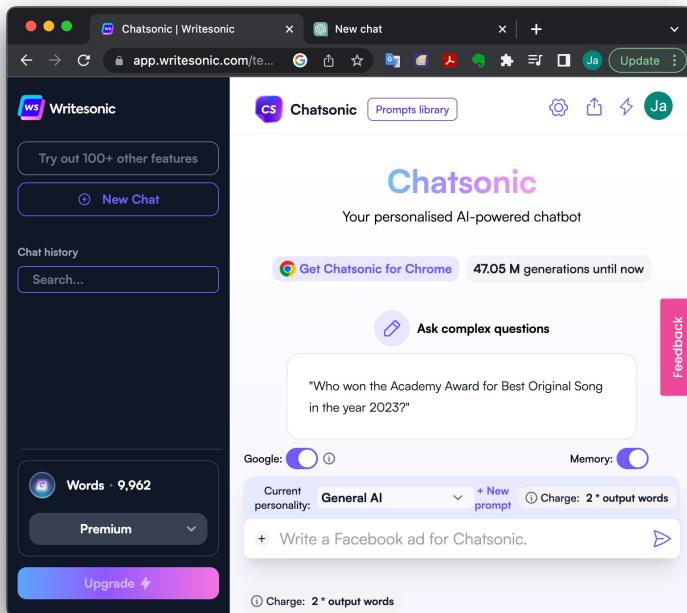
/
[A]
/

WHAT DO I EXPECT?



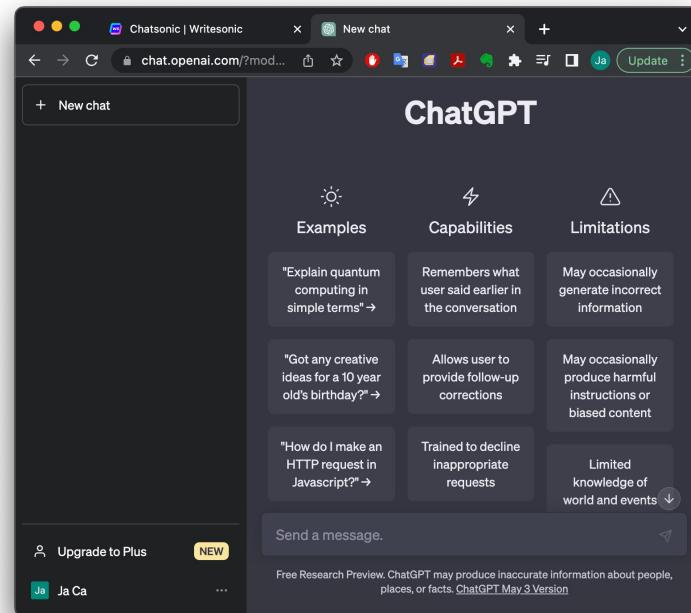
ChatGPT
wins...

$W_1 = W_2 = W_3$ PERFORMANCE P



0.9 Accuracy
0.8 Efficiency
0.7 Ineffectiveness

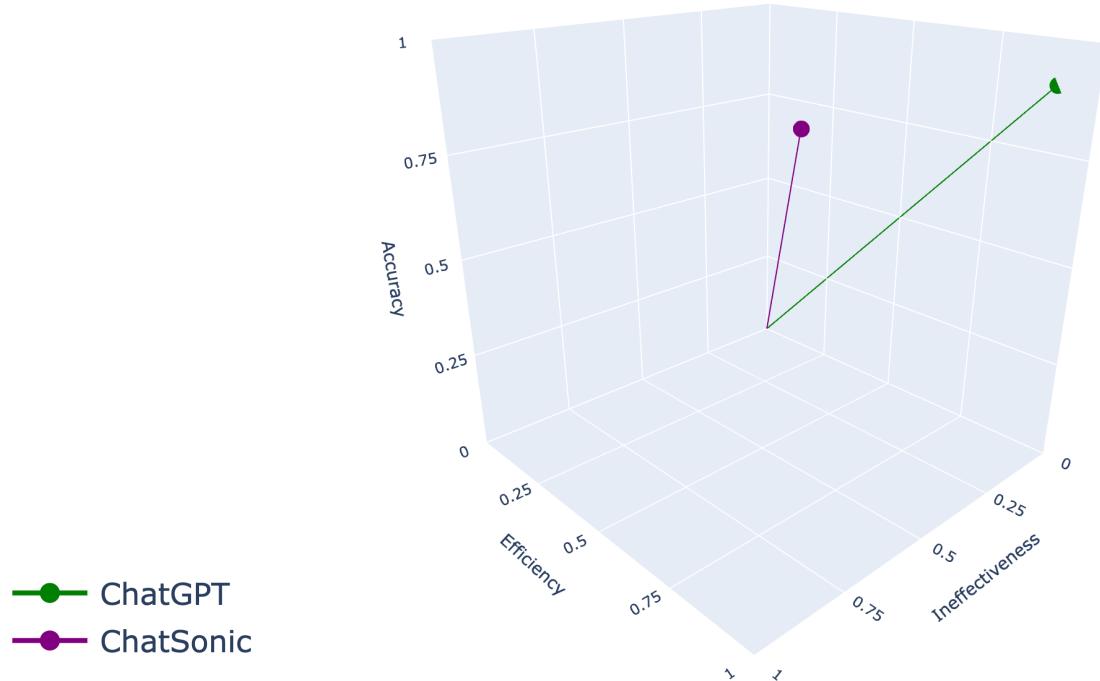
$$P_{\text{chatSonic}}=0.33$$



0.9 Accuracy
0.9 Efficiency
0.0 Ineffectiveness

$$P_{\text{chatGPT}}=0.60$$

VECTOR P



A screenshot of the ChatGPT interface. At the top, it says "New chat". Below that, a user message "Ja Give me a powerful quote to end a presentation" is followed by a response from ChatGPT: "The future belongs to those who believe in the beauty of their dreams." - Eleanor Roosevelt. The interface includes a "Feedback" button at the bottom right.

ChatGPT Mar 14 Version. Free Research Preview. Our goal is to make AI systems more natural and safe to interact with. Your feedback will help us improve.

A screenshot of the ChatSonic | Writesonic interface. At the top, it says "Prompts library". Below that, a user message "Ja Give me a powerful quote to end a presentation" is followed by a response from ChatSonic: "Success is not final, failure is not fatal: it is the courage to continue that counts." - Winston Churchill. The interface includes a "Feedback" button at the bottom right.

Google: Regenerate Memory:

Current personality: Personality: General AI + New prompt Charge: 2 * output words

+ Write a Facebook ad for ChatSonic. Charge: 2 * output words

Thank you

Do you have any questions?
jacopo.caratti@students.unibe.ch