Piano Nazionale di Ripresa e Resilienza Missione 4 – Componente 2, Investimento 1.1 CALL PRIN 2022 D.D. 104 02-02-2022 – (STEMMA) Codice CUP B53D23019500006

Is Generative AI Mature for Alternative Image Descriptions of STEM Content?

Marina Buzzi[1], Giulio Galesi[2], Barbara Leporini[2,3] and Annalisa Nicotera[3]

[1] IIT-CNR, Institute of Informatics and Telematics, National Research Council, Pisa, Italy
[2] ISTI-CNR, Institute of Information Science and Technologies, National Research Council, Pisa, Italy
[3] University of Pisa, Largo B. Montecorvo, Pisa, Italy
marina.buzzi@iit.cnr.it, giulio.galesi@cnr.it, a.nicotera2@studenti.unipi.it, barbara.leporini@unipi.it

INTRODUCTION

Artificial intelligence (AI) supports visually impaired people in getting information about the world around them.
This study investigates whether existing AI-based tools on the market are mature for describing images related to scientific content.

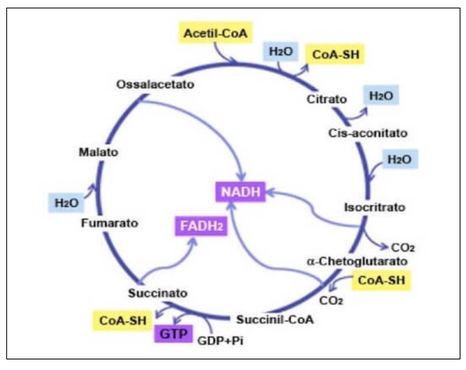
THE STUDY

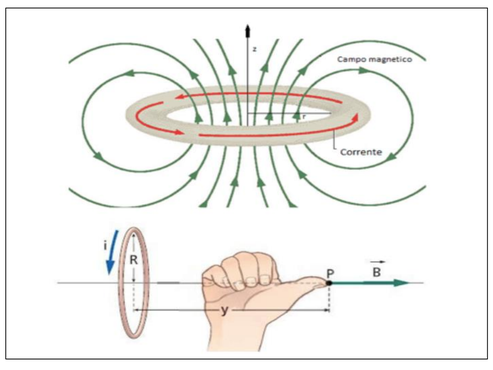
Comparison of 5 popular AI-based tools
- Seeing AI
- Be My Eyes (mobile version)
- Microsoft Copilot
- Google Gemini
- Google Gemini Advanced

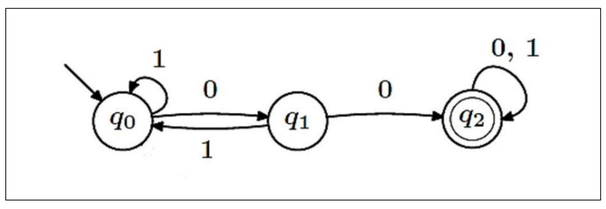
in describing 4 STEM images of different science domains
- Equation (Mathematics, easy, no difficulty)
- Krebs cycle (Chemistry, basic difficulty).
- Magnetic field (Physics, medium difficulty).
- Diagram of a Finite State Machine (computer science, moderate difficulty).

3 types of prompts:
- Prompt 1: ‘What is in this picture?’.
- Prompt 2: ‘Can you describe this picture?’.
- Prompt 3: I’m a blind person. Can you describe this picture?’.

1+1=2





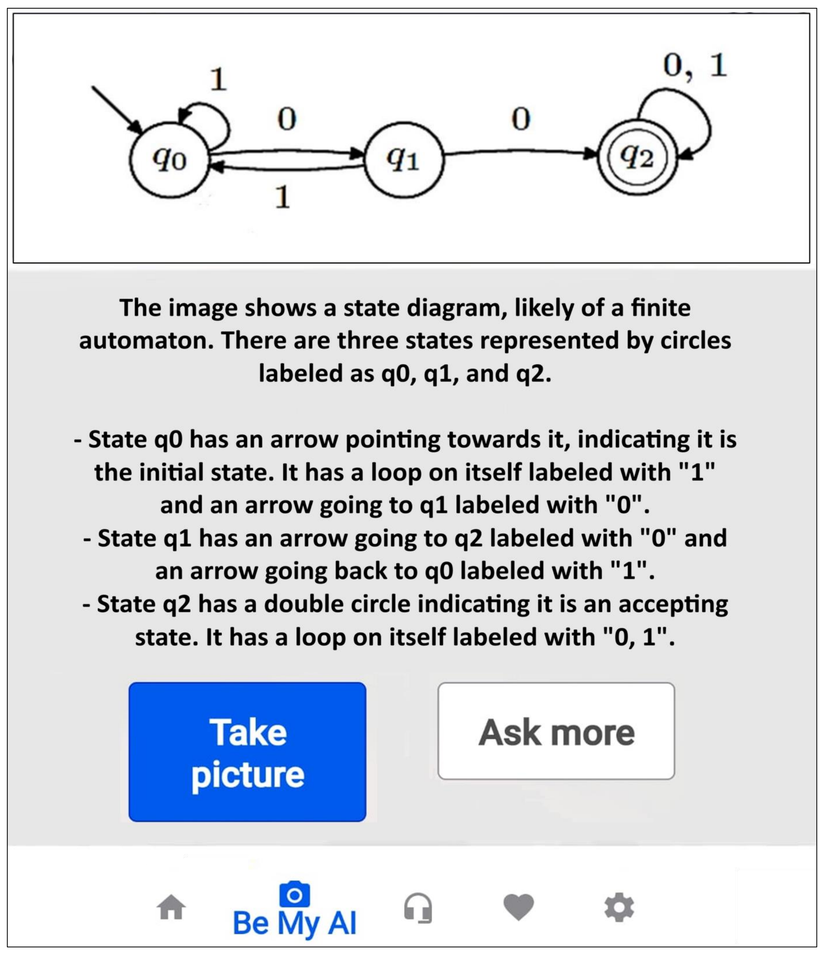


RESULTS

AI tools can generate image descriptions in STEM education, but more effort is needed to get effective content for satisfying and accurate image exploration by visually impaired or blind students.

DISCUSSION

- The behaviour of the tools in interpreting the target images may change over time.
- It seems that the Gemini tool is trained with images with backgrounds, which may be more suitable for photos and pictures.
- Results may be affected by image quality. ‘Be My Eyes’ requires taking a photo, so the image resolution and quality can vary (taking a photo is complex for blind users).
- The more accurate the answer the more time is needed for the AI assistant to describe the picture, as it happens with Bing Copilot.



CONCLUSIONS

Generative AI can be used for STEM image descriptions. Current AI tools provide good descriptions for STEM images, but accuracy and educational context need further investigation and enhancement to generate more suitable STEM image content.
Thus, for education purposes, image descriptions should be generated more effectively according to:
(1) the learning level in the specific subject;
(2) The education purpose (learning, review/practice, exam).

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
Università di Pisa