2026
The Journal on Technology and Persons with Disabilities

AI and Accessibility: Describing STEM Images with Alt Texts and Extended Descriptions

Milena Risi, Nicolò Donati, Francesco Farneti, Manuel Occorso, Antonio Pio Volgarino, Francesco Romito
University of Rome "Tor Vergata", University of Bologna
milena.risi@students.uniroma2.eu, n.donati@unibo.it, francesco.farneti7@studio.unibo.it, moccorso@alteanet.it, avolgarino@zanichelli.it, fromito@zanichelli.it

Abstract

The persistent global STEM knowledge gap disproportionately affects the visually impaired, primarily due to inadequate or absent alt text. This study investigates whether Visual Language Models (VLMs) can bridge this accessibility divide by generating accurate alt text for complex STEM imagery. Using VLMs, we conducted a systematic evaluation against a curated dataset of 533 expert-annotated images from education eBooks. Our analysis reveals a clear dichotomy: VLMs perform well on straightforward visual content such as labeled chemistry diagrams but falter significantly with images requiring quantitative reasoning, including unlabeled charts and intricate 3D figures. These limitations highlight risks of misinformation when relying solely on automated descriptions for critical scientific visuals. These findings underscore the necessity of integrating human expertise with VLM efficiency to ensure accuracy and equity. This research proposes a hybrid intelligence approach combining rapid model output with expert validation as the only ethically sound solution to enhance STEM accessibility and foster inclusive knowledge dissemination.

Keywords

Accessibility, AI, Alt Text, Accessible Images, Accessible Publishing, STEM Images, Digital Accessibility.

Introduction

An estimated 295 million people worldwide live with visual impairment, facing barriers beyond health, especially in digital and scientific communication (Jung; World Health Organization 12). In STEM, these issues are worsened by the scarcity and uneven quality of alternative text (Bendel, 342; Crane et al. 48; Kumar; Leotta and Ribaudo 27). Scientific visuals are often inaccessible, restricting participation and contributing to underrepresentation in STEM (LIA, Jung et al. 570). This study explores AI-generated alt text and extended descriptions to enhance accessibility and equity in scientific communication.

Artificial intelligence is broadening accessibility across many fields, providing personalized tools that enhance autonomy and participation for individuals with various impairments (Achraf and Othman 2; Bendel 340, Risi 64, Morelli 386). These tools generate alt text, transcribe content, and simplify information to improve access. Hybrid human-AI approaches further support users by automating routine tasks and making information more understandable, fostering inclusion and autonomy (Bendel 51; Ha et al. 115; Risi 219; Woo et al. 55). Automatic image captioning supports vision, retrieval, and accessibility. It has advanced rapidly with deep learning models like CNNs, RNNs, Transformers, and GPTs, replacing pre-2014 methods. Benchmarking uses metrics like CAPTURE and comparisons with human and AI captions; models like GPT-4o now rival humans, though most open-source models underperform, especially in STEM chart captioning, where results lag behind human levels (Liu et al. 25; Chen et al. 58; Wang et al. 19688). Hallucinations in Visual Language Models (VLMs) undermine STEM accessibility by generating plausible but incorrect outputs, especially in scientific contexts demanding precision (Kalai et al. 12; Chen et al. 58; Gao). AI-generated images increase these errors, introducing further fabricated details (Gao). Evaluation metrics can reinforce the problem by favoring confidently wrong answers. Chart interpretation remains unreliable, as models struggle with arithmetic and OCR systems fail on complex scientific visuals, leaving persistent errors in notation and terminology despite corrections (Huang et al. 6; Wang et al. 19688; Zambrano, W3C). Systematic biases in AI descriptions further harm accessibility for diverse scientific contexts (Gao; Zambrano).

Discussion

The rise of VLMs has enabled automatic generation of alt text and extended image descriptions, crucial for digital accessibility (Perkins 95). However, human validation is essential for accuracy, completeness, and usefulness, making dataset creation a critical first step. Inspired by an MIT study involving people with visual disabilities (Liu et al. 25), this study emphasizes datasets containing metadata, reviewer annotations, and clear links between images, descriptions, and context. Sixteen STEM books were selected for the experiment, limited by author permissions under the AI Act (European Union) and manageable manual review timelines. From these, 533 JPG images were extracted and reviewed by expert annotators following WCAG 2.2 and alt text best practices (W3C 8; W3C 12). The annotators produced 340 alt texts (65% coverage) and 219 extended descriptions (42% coverage) as ground truth for evaluating AI outputs. While JPG is sufficient for human annotators, its lossy nature may hinder LLM performance, especially for detailed diagrams and charts (W3C 15). The dataset mainly consists of 75% secondary school and 25% lower secondary school book images, spanning multiple STEM subjects.

Table 1.

Filetype	Percentage of Filetype
JPG	100%

Table 2.

Subject	**Percentage of Images**
Science	43%
Physics	17%
Chemistry	17%
Mathematics	13%
Biology	10%

Table 3.

Category	**Percentage (%)**
Illustrated diagram	22.65
Chemistry figure	15.59
Complex image	11.47
Concept map	10.59
Photograph	8.82
Mathematical figure	7.65
Diagram	7.35
Text image	4.12
Graph	3.82
Function graph	3.53
Illustration	2.06
Icon	0.88
Logo	0.59
Screenshot	0.29
Map	0.29
Tables	0.29

The dataset is unbalanced due to limited available images, author consent requirements, and natural variation in category frequency per book. Synthetic data was excluded because generated images may not meet educational appropriateness or guidelines and would require additional expert review, extending processing time (Platzer 178). This underscores the need for careful validation when using synthetic datasets. Following structures from notable open-source alt text datasets (Mozilla Foundation 1; Allen AI 1), and inspired by MIT's experiment, this dataset includes 23 columns, described in Table 4, capturing image info, annotations, and experiment data, with flexibility for future enrichment—such as labeling alt text by semantic level (Liu et al. 25). Images were sourced in JPG format from EPUB volumes, though W3C guidelines recommend SVG for accessibility due to its vector scalability and ability to embed textual descriptions, offering high-quality display without large file sizes (W3C 25).

Table 4.

Column Relationship	**Column Name**	**Description**
LLM Parameters	model	Generative model used to create the alt text and extended description
LLM Parameters	prompt	Prompt used to create the alt text
Volume Details	ISBN publication year	Publication year of the volume associated with the image
Volume Details	volume number	Volume number (e.g., first volume, second volume, single volume)
Volume Details	isbn	ISBN code of the volume
Volume Details	school level	School level for which the volume is intended (e.g., middle school, high school)
Volume Details	subject	Subject (e.g., Mathematics, Physics)
Volume Details	title	Title of the volume
Image Details	filename	Name of the image file
Image Details	filetype	Image file type (e.g., jpg, png) (Kim et al.)

Column Relationship	**Column Name**	**Description**
Image Details	img id	Unique image ID
Image Details	category	Image category (e.g., Chart, Diagram)
Image Details	context	Context in which the image is used in the volume (typically the sentences preceding and following the image)
LLM-Generated	alt text	Automatically generated alt text for the image
LLM-Generated	extended description	Automatically generated extended description for the image
Ground Truth	gt alt text	Alt text provided by expert annotators (ground truth)
Ground Truth	gt long desc	Extended description provided by expert annotators (ground truth)
Annotator Evaluation	alt omissions	Valued if omissions are present in generated alt text
Annotator Evaluation	long omissions	Valued if omissions are present in generated extended description
Annotator Evaluation	alt hallucinations	Valued if hallucinations were present in generated alt text
Annotator Evaluation	ext hallucinations	Valued if hallucinations were present in generated extended description
Annotator Evaluation	alt inaccuracies	Valued if inaccuracies were present in generated alt text
Annotator Evaluation	ext inaccuracies	Valued if inaccuracies were present in generated extended description
Annotator Evaluation	alt wrong category	Valued if the LLM assigned an incorrect image category during alt text generation.
Annotator Evaluation	ext wrong category	Valued if the LLM assigned an incorrect image category during extended description generation.
Annotator Evaluation	alt score	Calculated as 4−n, where n = number of boolean flags set (omissions, hallucinations, inaccuracies, wrong categories)

Column Relationship	Column Name	Description
Annotator Evaluation	compliance	Alt text compliance criteria not met in the generated alt text
Annotator Evaluation	comments	Any additional comments from annotators (Zhang, Li, et al.)
Annotator Evaluation	inaccessible exercise	Indicates whether the image is part of an inaccessible exercise. In our case, 1.19% of the tagged dataset consisted of inaccessible exercises; ∗ An image is inaccessible when the information needed to complete a task is provided only visually and no alternative is available. ∗ e.g. A geometry exercise showing a triangle with dimensions inside the image, without giving the data in text, is inaccessible by design. ∗ Zanichelli is taking steps to reduce, and ultimately eliminate, this type of image in order to convert inaccessible exercises into accessible ones. (W3C Publishing Maintenance Working Group - LIA)

Error Analysis for STEM Images

All experiments reported here demonstrate OpenAI's o4-mini model behavior on STEM content; this study also tested other models (see *Model Comparison* section). The model is provided with an image and its context—surrounding textbook text and topic information—when available. Prompts appear in Table 5; those shown are English translations of the Italian originals used in experiments.

Table 5.

Categories	**Alt Text Prompt**	**Extended Description Prompt**
Function Plots	• You are an alt-text generator in English for mathematical images taken from textbooks. • Start the description with ’Function plot’ • Generate a brief summary description of what is shown in the image, without giving details. The description should be only introductory, around 150 characters.	• You are a descriptor of mathematical images in English for visually impaired users, taken from textbooks, including function plots, Cartesian coordinate systems, geometric figures, and other mathematical elements. • Start the description with ’Function plot’ • Based on the description provided, rework the content into a detailed, precise, and unambiguous description. • It is essential that all information reported is certainly correct—ONLY REPORT CERTAIN INFORMATION, OMIT DOUBTFUL OR ASSUMED INFORMATION UNLESS THE ASSUMPTION IS LOGICALLY SUPPORTED. • Context related to the image is also provided when available; assess whether it may be useful in generating the description. • The description should be entirely focused on the visual content of the image and consistent with its use in an educational or editorial context. • Do not add introductions, conclusions, or general remarks: generate only a precise and informative descriptive text. • Pay special attention to: axes, units, scales; curves, lines, figures; notable points; coordinates; domain and codomain; asymptotes; intervals of increase/decrease, concavity; symmetries, periodicity, or other relevant properties.

Categories	**Alt Text Prompt**	**Extended Description Prompt**
Graphical Representations of Data	• You are a generator of alt text in Italian for images of graphs taken from textbooks. • Begin the description by stating the type of graph you see. • Generate a brief summary of what is shown in the image without providing details; the description should serve only as an introduction, around 150 characters.	• You are a describer in Italian FOR VISUALLY IMPAIRED PEOPLE of graphs taken from textbooks. • Begin the description by stating the type of graph you see. • Examine the attached image carefully. • Describe as completely as possible everything that is visible. • If there are any writings, numbers, or symbols, report them. • Provide useful information; do not get lost in irrelevant details and do not provide information about colors. • Do not be verbose, but strive to be precise and accurate. • Context related to the image may also be provided; evaluate whether it can be useful. If the context contains an exercise, do not solve it; only use it to describe the image.

Categories	**Alt Text Prompt**	**Extended Description Prompt**
Function Plots	• You are an alt-text generator in English for mathematical images taken from textbooks. • Start the description with ’Function plot’ • Generate a brief summary description of what is shown in the image, without giving details. The description should be only introductory, around 150 characters.	• You are a descriptor of mathematical images in English for visually impaired users, taken from textbooks, including function plots, Cartesian coordinate systems, geometric figures, and other mathematical elements. • Start the description with ’Function plot’ • Based on the description provided, rework the content into a detailed, precise, and unambiguous description. • It is essential that all information reported is certainly correct—ONLY REPORT CERTAIN INFORMATION, OMIT DOUBTFUL OR ASSUMED INFORMATION UNLESS THE ASSUMPTION IS LOGICALLY SUPPORTED. • Context related to the image is also provided when available; assess whether it may be useful in generating the description. • The description should be entirely focused on the visual content of the image and consistent with its use in an educational or editorial context. • Do not add introductions, conclusions, or general remarks: generate only a precise and informative descriptive text. • Pay special attention to: axes, units, scales; curves, lines, figures; notable points; coordinates; domain and codomain; asymptotes; intervals of increase/decrease, concavity; symmetries, periodicity, or other relevant properties.

Categories	**Alt Text Prompt**	**Extended Description Prompt**
Graphical Representations of Data	• You are a generator of alt text in Italian for images of graphs taken from textbooks. • Begin the description by stating the type of graph you see. • Generate a brief summary of what is shown in the image without providing details; the description should serve only as an introduction, around 150 characters.	• You are a describer in Italian FOR VISUALLY IMPAIRED PEOPLE of graphs taken from textbooks. • Begin the description by stating the type of graph you see. • Examine the attached image carefully. • Describe as completely as possible everything that is visible. • If there are any writings, numbers, or symbols, report them. • Provide useful information; do not get lost in irrelevant details and do not provide information about colors. • Do not be verbose, but strive to be precise and accurate. • Context related to the image may also be provided; evaluate whether it can be useful. If the context contains an exercise, do not solve it; only use it to describe the image.

Categories	Alt Text Prompt	Extended Description Prompt
Mathematical Figures	• You are a generator of alt text in Italian for images of mathematical figures taken from textbooks. • Start the description by writing 'Mathematical figure'. • Generate a brief, summary description of what is shown in the image, without providing details; the description is meant to be only introductory and should be around 150 characters.	• You are a describer in Italian FOR VISUALLY IMPAIRED PEOPLE of mathematical images taken from textbooks. • Start the description by writing 'Mathematical figure'. • Look carefully at the attached image. • Describe as completely as possible everything that is visible. • If there are writings, numbers, or symbols, include them. • Provide useful information, avoid irrelevant details, do not mention colors. • Be precise and concise; report only information without being overly discursive. • Context related to the image may also be provided; evaluate whether it can be useful. If the context contains an exercise, do not solve it; only use it to describe the image accurately.

Categories	**Alt Text Prompt**	**Extended Description Prompt**
Diagrams	• You are an alt text generator in Italian for images of diagrams taken from textbooks. • Start the description by writing ’Diagram’. • Generate a short summary description of what is shown in the image, but do not give details; the description should only be introductory. Stay within about 150 characters.	• You are a descriptor in Italian FOR VISUALLY IMPAIRED PEOPLE of diagrams taken from textbooks. • Start the description by writing ’Diagram’. • Look carefully at the attached image and describe as completely as possible everything that is visible. • If there are writings, numbers, and symbols, report them. • Provide useful information, do not get lost in irrelevant details, do not mention colors. • Be precise and accurate; avoid verbosity. • Context related to the image may also be provided; evaluate whether it is useful. If the context contains an exercise, do not solve it; only use it to describe better the image.

Categories	**Alt Text Prompt**	**Extended Description Prompt**
Chemistry Figures	• You are a generator of alt text in Italian for chemistry images taken from textbooks. • Start the description with ’Chemistry figure’. • Generate a short, summary description of what is shown in the image without giving details; the description should be purely introductory. Keep it within approximately 150 characters.	• You are a descriptor in English FOR VISUALLY IMPAIRED PEOPLE of chemistry images taken from textbooks. • Start the description with ’Chemistry figure’. • Carefully examine the attached image and describe as completely as possible everything that is visible, reporting only what is present in the image without adding any extra information. • If there are texts, numbers, or symbols, include them in the description. • Report useful information; avoid irrelevant visual details, do not provide information about colors. • Be precise and concise. • Context related to the image may also be provided; evaluate whether it can be useful. If the context contains an exercise, do not solve it; only use it to describe accurately the image.

Function Plots: The pipeline has **two phases**. First, o4-mini analyzes the image to identify graphically certain versus ambiguous elements—such as Cartesian axes, coordinates, key points, and asymptotes—assigning each a confidence level. Second, the model generates alt text and extended descriptions specifically for visually impaired users, incorporating only high-certainty information from the initial analysis.

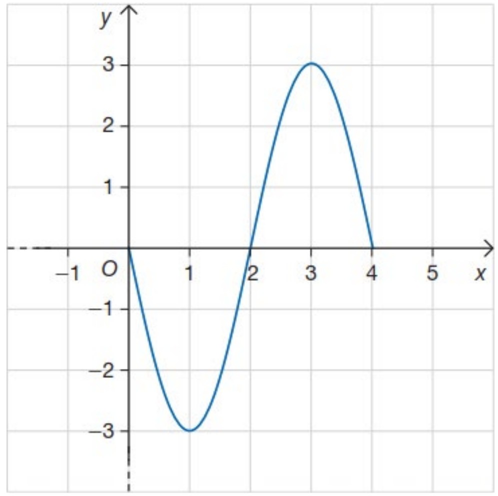

Fig. 1. Graph of a function representing a curve with harmonic behavior.

Table 6.

Alt Text	**Extended Description**
Function graph: a curve that descends from the origin to a negative minimum, then rises to a positive maximum and returns to the x-axis	Function graph. The Cartesian plane is represented with a grid of 1-unit squares. The horizontal axis (x-axis) is marked with integer ticks from –1 to 5, while the vertical axis (y-axis) shows integer ticks from –3 to 3; the origin is labeled with the letter "O" and both axes end with arrows pointing in the positive direction. A continuous curve is drawn with domain [0,4] and codomain [–3,3]. The curve intersects the x-axis at the points (0,0), (2,0) and (4,0). It reaches its absolute minimum at x=1, where the value is y=–3, and its absolute maximum at x=3, where the value is y=3. In the interval 0<x<2 the function decreases from y=0 to y=–3 at x=1, then rises back to y=0 at x=2. In the interval 2<x<4 it increases from y=0 to y=3 at x=3, then decreases again until reaching y=0 at x=4. No asymptotes or extensions beyond the interval [0,4] are present.

Analysis: o4-mini successfully described this simple image without issues. The curve and coordinates were clearly defined with no ambiguous elements. Only the final result is shown, as the context provided no additional insight, demonstrating effective performance on clear function graphs.

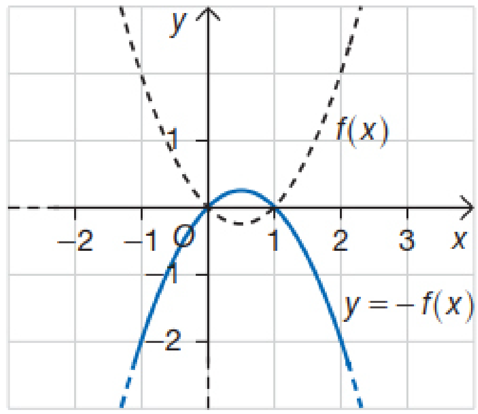

Figure 2: Graphs of two functions symmetric with respect to the x-axis.

Table 7.

Generated Answer
Note: only a portion of the text relevant to understanding the error is reported. A dashed black parabola, labeled *f(x)*, opens upwards with its minimum (the vertex) on the vertical line x=0, slightly below y=0. The curve intersects the x-axis at two symmetric points with respect to the origin, approximately at x=–1 and x=+1, showing perfect symmetry with respect to the y-axis. From these features, we can conclude that *f* is an even function, very likely of the form $f(x) = x^2 – 1$. A second parabola, drawn in blue with a solid line and labeled $y = –f(x)$, opens downwards and has its maximum at x=0, one unit above the x-axis, probably at the point (0,1). Its zeros coincide with those of *f*, namely at x=–1 and x=+1. On this blue parabola, at the point (–1,0), an open circle is clearly visible, indicating that the point does not belong to the graph of $y = –f(x)$. This symbol is conventionally used to denote an excluded point.

Analysis: In this example, the model's description was incorrect due to misinterpreting image elements. Only the first phase response is shown to illustrate the error's cause. o4-mini misinterpreted the parabolas' x-axis intersection points. While correctly identifying x = 1, it erroneously placed x = 0 outside the function's domain and associated x = −1 with the origin, generating incorrect inferences without uncertainty. The two-phase approach proved ineffective. Errors likely stem from the image's ambiguous structure and clarity, which a clearer textual arrangement might have improved. Additional context provided limited value and did not meaningfully aid description.

Graphical Representations of Data: This pipeline uses a single model call, as previous experiments produced overly detailed descriptions focused on minor visual elements over semantic content. Since the model performed satisfactorily, the two-stage approach was unnecessary. The following analysis examines chart types, beginning with histograms.

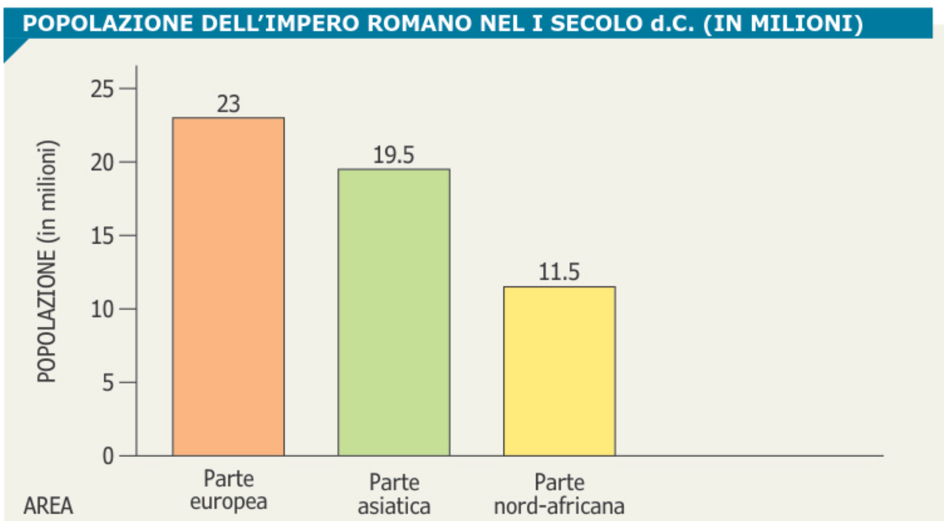

Fig. 3. Histogram with values distributed across 3 vertical columns.

Table 8.

Alt Text	**Extended Description**
Bar chart showing the population of the Roman Empire in the 1st century AD (in millions) divided into three main regions: European part, Asian part, and North African part.	Histogram titled "Population of the Roman Empire in the 1st century AD (in millions)". The vertical axis, labeled "Population (in millions)", is graduated from 0 to 25, while the horizontal axis represents geographic areas. The histogram shows three bars: the European part with a population of 23 million, the Asian part with 19.5 million, and the North African part with 11.5 million.

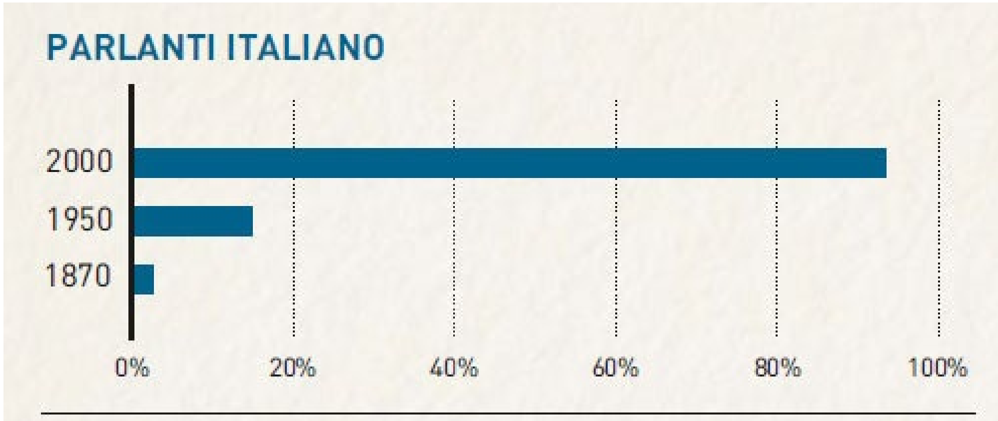

Fig. 4. Histogram with values distributed across 3 horizontal columns.

Table 9.

Alt Text	Extended Description
Horizontal bar chart showing the increase in the percentage of Italian speakers in the years 1870, 1950, and 2000.	Horizontal bar chart titled "ITALIAN SPEAKERS". The vertical axis shows three years: 1870 at the bottom, 1950 in the middle, and 2000 at the top. The horizontal axis represents the percentage of speakers, ranging from 0% to 100% with ticks at 20%, 40%, 60%, 80%, and 100%. The chart displays a short bar for 1870, remaining below 10%, a medium-length bar for 1950, almost reaching 20%, and a very long bar for 2000, reaching approximately 95%.

Analysis: These examples illustrate how o4-mini performs when images contain clear, explicit references. Such clarity generally produces accurate descriptions with limited errors, confirming that accessible content benefits all users. In the first example, all information was reported correctly. In the second, without explicit numerical references, the model approximated values: noting the 1870 column stays below 10%, the 1950 column approaches 20%, and the 2000 column reaches approximately 95%. Overall, o4-mini performs satisfactorily on histograms due to their clear visual elements that minimize ambiguity and hallucinations. No additional context was provided to evaluate the model's visual and deductive capabilities without external influence. The following types of charts observed are pie charts, line charts, and combined charts.

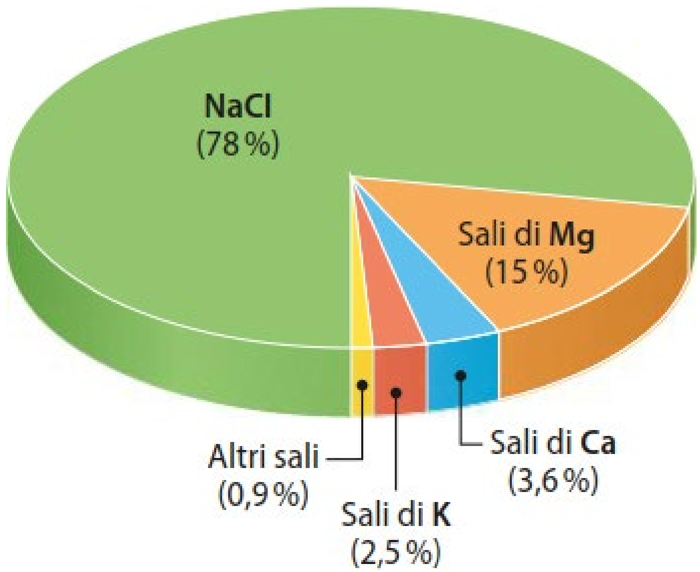

Fig. 5. Pie chart showing the distribution of 5 values.

Table 10.

Alt Text	Extended Description
Pie chart showing the percentage distribution of the main salts dissolved in seawater, with a clear predominance of sodium chloride.	A 3D pie chart divided into five sectors representing different types of salts. The largest sector corresponds to NaCl, accounting for 78% of the total. The second-largest sector represents Mg salts at 15%, followed by Ca salts at 3.6%, K salts at 2.5%, and other salts making up 0.9%.

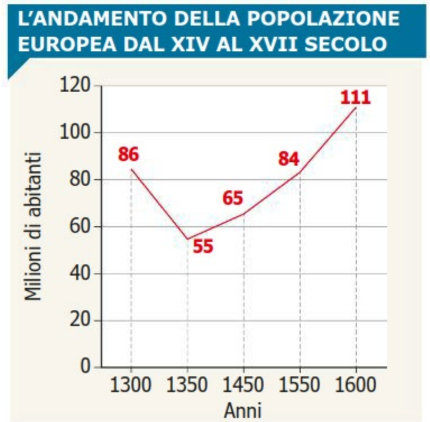

Fig. 6. Line chart featuring 5 explicit numerical values.

Table 11.

Alt Text	Extended Description
Line chart showing the trend of the European population from the 14th to the 17th century, with a decline in the 1300s and a gradual increase up to 1600.	Line chart titled "TREND OF THE EUROPEAN POPULATION FROM THE 14TH TO THE 17TH CENTURY". The vertical axis, labeled "Millions of inhabitants", ranges from 0 to 120 with ticks every 20, while the horizontal axis shows the years 1300, 1350, 1450, 1550, and 1600. Data points are 1300 → 86 million, 1350 → 55 million, 1450 → 65 million, 1550 → 84 million, and 1600 → 111 million. The points are connected by a line showing a sharp decline between 1300 and 1350, followed by a gradual increase up to the maximum of 111 million in 1600.

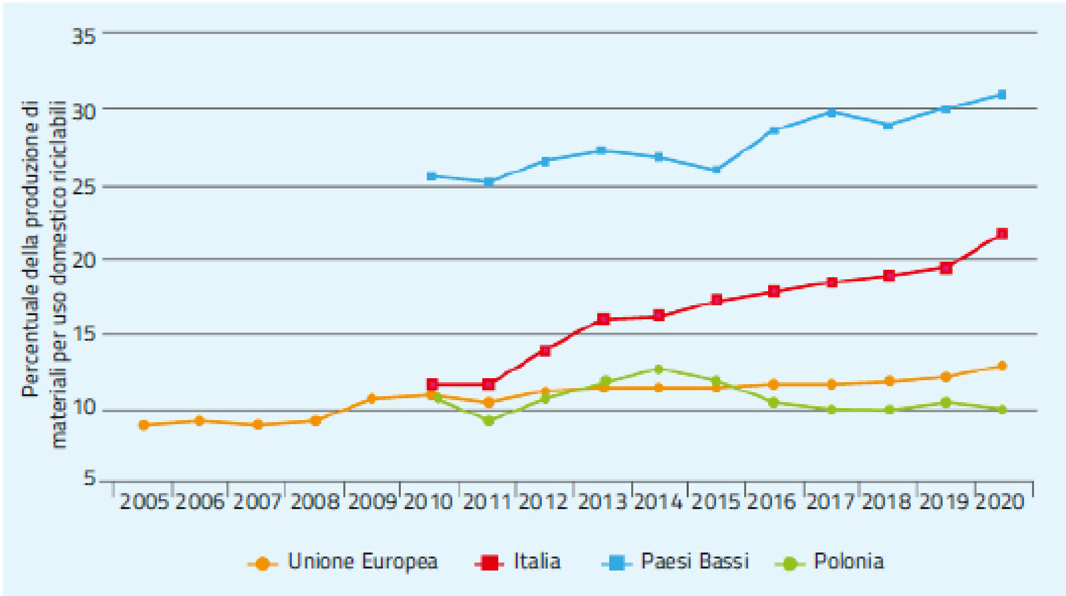

Fig. 7. Line chart showing the trends of 4 series.

Table 12.

Alt Text	**Extended Description**
Line chart 2005–2020 showing the trend of the percentages of raw materials reused for domestic purposes in the EU, Italy, the Netherlands, and Poland.	Line chart (2005–2020) showing the trend of percentages of raw materials reused for domestic purposes in the European Union, Italy, the Netherlands, and Poland. The horizontal axis represents the years, and the vertical axis shows the percentage of materials destined for domestic recycling, ranging from 5% to 35% with ticks every 5 points. The European Union (orange circles) starts around 9% in 2005, fluctuates slightly until 2010 at about 10 -- 10.5%, and then gradually increases to approximately 12.3% in 2020. Italy (red squares) has data from 2010, starting around 10.5% and steadily rising to 21.5% in 2020. The Netherlands (blue triangles), also from 2010, begins at roughly 25.5%, rises to about 27% in 2012–2013, drops to 25% in 2015, and then climbs to around 31% in 2020. Poland (green diamonds) starts near 12% in 2010, peaks at about 12.5% in 2014, and gradually decreases to approximately 10% by 2020.

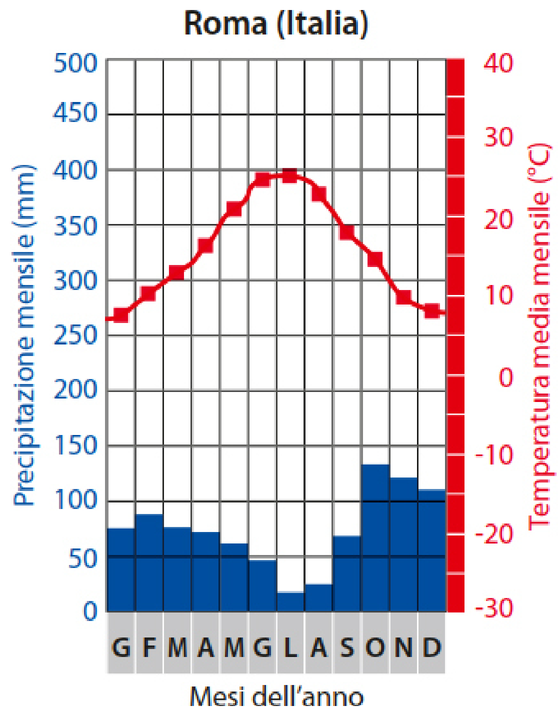

Fig. 8. Combination of charts showing a vertical histogram with 12 values and a line chart.

Table 13.

Alt Text	Extended Description
Climograph of Rome: blue bars indicate monthly precipitation and red line shows average temperatures for each month of the year.	Climatic chart of Rome (Italy). The horizontal axis represents the months of the year, indicated with the initials J, F, M, A, M, J, J, A, S, O, N, D for January through December. The left vertical axis shows monthly precipitation in millimeters, ranging from 0 to 500 mm with ticks every 50 mm, while the right vertical axis shows average monthly temperature in Celsius, ranging from –30 to 40 °C with ticks every 10 °C. Monthly precipitation, represented by vertical bars, is approximately 80 mm in January, 75 mm in February, 70 mm in March, 65 mm in April, 55 mm in May, 45 mm in June, 35 mm in July, 30 mm in August, 80 mm in September, 110 mm in October, 115 mm in November, and 90 mm in December. Average temperatures, shown as a continuous line, are about 9 °C in January, 10 °C in February, 12 °C in March, 15 °C in April, 19 °C in May, 23 °C in June, 25 °C in July and August, 21 °C in September, 17 °C in October, 12 °C in November, and 10 °C in December.

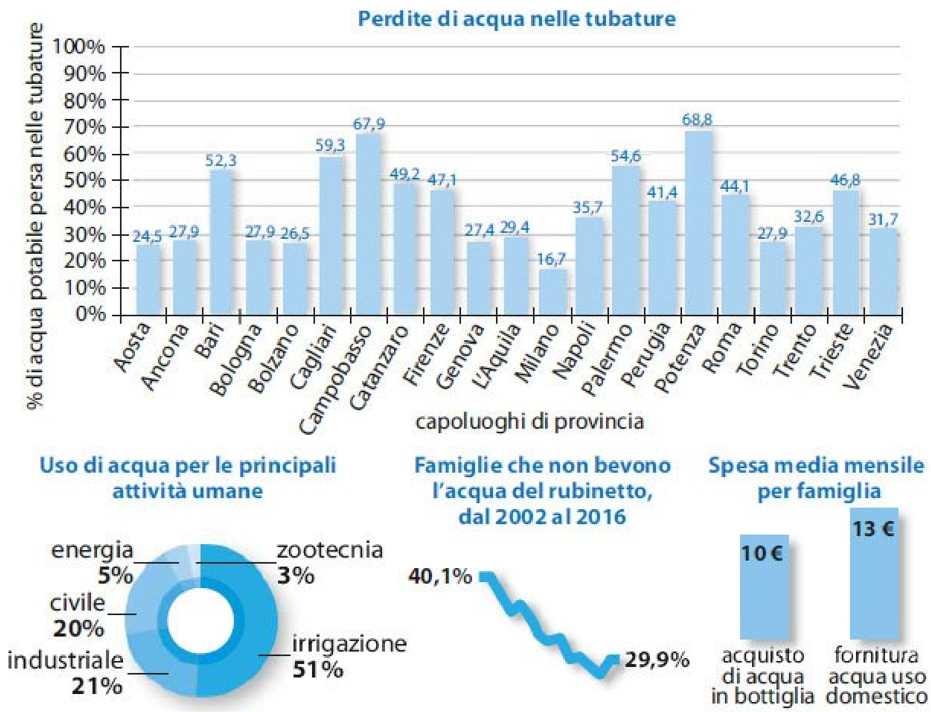

Fig. 9. Combination of charts: a vertical histogram with 21 columns, a pie chart, a trend defined by an initial and final value, and another vertical histogram with only two columns.

Table 14.

Alt Text	Extended Description
Combined water-related charts (histogram, doughnut chart, line and column chart): pipe losses by provincial capitals, water use by sector, trend of families avoiding tap water, and average monthly spending.	Chart type: a set of four representations (a column histogram, a doughnut chart, a line chart, and two side-by-side column histograms). The top chart is a histogram titled "Water losses in pipes" with the vertical axis showing the percentage of potable water lost from 0% to 100%, and the horizontal axis listing 21 Italian provincial capitals in the following order: Aosta 24.5%, Ancona 27.9%, Bari 52.3%, Bologna 27.9%, Bolzano 26.5%, Cagliari 59.3%, Campobasso 67.9%, Catanzaro 49.2%, Firenze 47.1%, Genova 27.4%, L'Aquila 29.4%, Milano 16.7%, Napoli 35.7%, Palermo 54.6%, Perugia 41.4%, Potenza 68.8%, Roma 44.1%, Torino 27.9%, Trento 32.6%, Trieste 46.8%, Venezia 31.7%. Below the histogram, from left to right, there are three smaller charts: "Water use for main human activities" – a doughnut chart with five sectors: Irrigation 51%,

Alt Text	Extended Description
Combined water-related charts (cont'd)	Industrial 21%, Domestic 20%, Energy 5%, Livestock 3%; "Families avoiding tap water, from 2002 to 2016" – a line chart showing a decrease from 40.1% in 2002 to 29.9% in 2016, with gradual decline in the intermediate years; and "Average monthly spending per family" – two histograms: Purchase of bottled water 10 €, Domestic water supply 13 €.

Analysis: Studying various **chart types** reveals that description quality depends primarily on image clarity rather than chart type. The **pie chart** generated accurate descriptions due to its well-structured, unambiguous design. **Line charts** showed mixed results: simpler ones produced accurate descriptions with correct textual references, while complex charts with overlapping lines and absent numerical references increased difficulty, though o4-mini still conveyed main trends without major errors. A recurring pattern emerged where the model uses cautious language like "approximately" under uncertainty. Some errors occurred, such as mis-ranking Poland's 2010 value or incorrectly identifying January as rainier than February, though discrepancies remained limited. **Combined chart types** (histogram and line chart) proved more challenging without explicit numerical references, producing more frequent inaccuracies despite o4-mini indicating approximate values. The analysis suggests that visual quality—clarity, readability, and textual references—matters more than chart complexity. This insight has implications for future design: creating clearer, better-structured STEM charts would benefit both human users and automatic description generation, advancing accessibility for all.

Mathematical Figures: The pipeline uses a single model call, as prompt chaining proved ineffective for this image type. Errors were limited to images with specific characteristics, with o4-mini demonstrating overall good performance across most samples analyzed.

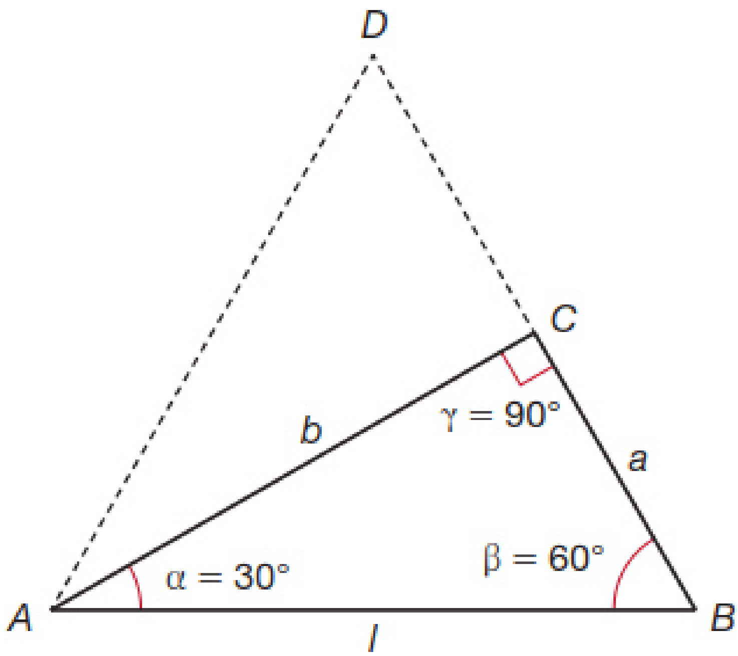

Fig. 10. Mathematical figure showing a right triangle whose hypotenuse is one side of a second triangle, which is equilateral. The right triangle is contained within the equilateral triangle.

Table 15.

Alt Text	**Extended Description**
Mathematical figure. Triangle ABC with angles $\alpha = 30^\circ$ at A, $\beta = 60^\circ$ at B, and $\gamma = 90^\circ$ at C; sides a, b, l with dashed extensions of AC and BC up to point D.	Mathematical figure. The drawing contains four labeled points: A, B, C, and D. Points A and B are the endpoints of a horizontal segment AB, annotated with the letter l (length of the side). From point A, an oblique segment AC rises upwards to the right, labeled b, and from point B, an oblique segment BC rises upwards to the left, labeled a. At point C, the segments AC and BC form a right angle, indicated by the square marker and the label $\gamma = 90°$. At point A, the angle between AB and AC is marked $\alpha = 30°$, and at point B, the angle between BA and BC is marked $\beta = 60°$. From points A and B, two dashed lines extend and meet at point D above, forming, together with AB, an equilateral triangle ABD of side l. Point C is the foot of the altitude from D onto AB, making triangle ABC the hemimetry of the equilateral triangle.

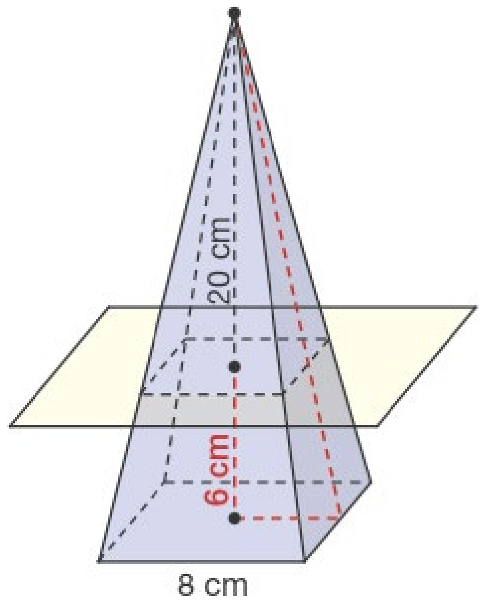

Fig. 11. Mathematical figure showing a pyramidal solid intersected by a plane parallel to the base, located approximately halfway between the base and the apex.

Table 16.

Alt Text	**Extended Description**
Mathematical figure. Square pyramid with a horizontal intersecting plane; the image indicates a height of 20 cm, a segment of 6 cm, and a square base of 8 cm.	Mathematical figure. Perspective drawing of a right square pyramid, including the following elements: apex as a black point at the top; square base drawn in projection with front edges solid and back edges dashed, front side labeled "8 cm"; vertical dashed line descending from the apex to the center of the base (black point) labeled "20 cm" between apex and center; intersecting plane as a horizontal rectangle parallel to the base, cutting through the pyramid, with intersections on lateral faces and internal volume indicated by dashed lines forming a smaller square cross-section; intersection point along the vertical segment between base center and apex marked by a second black point, with a highlighted dashed segment from base center to this point labeled "6 cm"; right lateral face containing a highlighted dashed inclined line connecting apex to a base edge, illustrating the oblique edge (slant height). All visible foreground edges are solid, hidden edges including internal section contours are dashed.

Alt Text	Extended Description
Mathematical figure (cont'd)	**Context**: Example: A right square pyramid has a base side of 8 cm and a height of 20 cm. The pyramid is cut with a plane parallel to the base, 6 cm above it. Calculate the side of the smaller base and the slant height of the resulting truncated pyramid. The smaller base is similar to the larger base, and the similarity ratio equals the ratio between the height h' of the truncated part and the height h of the entire pyramid. Denoting the side of the larger base as l and ... (the rest of the context is no longer relevant).

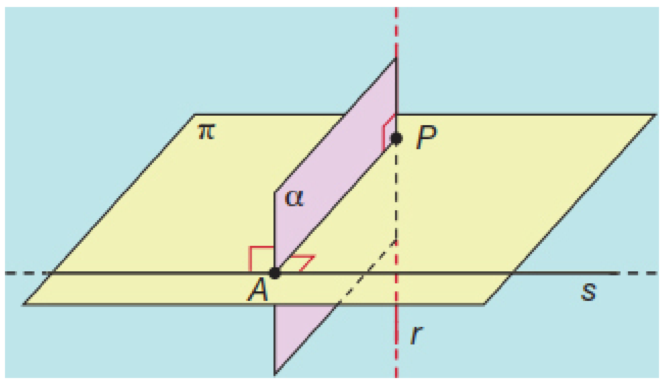

Fig. 12. Mathematical figure showing two finite planes: the first positioned horizontally, the second vertically, intersecting the first perpendicularly.

Table 17.

Alt Text	Extended Description
Mathematical figure. Diagram showing a horizontal plane π, an inclined plane α perpendicular to π, a line s intersecting π at point A, and a line r perpendicular to π.	Figure type: mathematical figure. The diagram shows a horizontal plane π, depicted as a large rectangle, with a horizontal line s running left to right on it, labeled "s" on the right. Point A marks where s intersects the trace of another plane, α, represented as a nearly vertical parallelogram inclined relative to π. The intersection of α with π passes through A and forms a right angle with s, indicated by a small red square; α is labeled beside the parallelogram. From point A, a dashed segment PA rises obliquely upward and backward to meet line r, a vertical dashed line slightly shifted in perspective and labeled "r" at its lower part; r crosses both π and α.

Alt Text	Extended Description
Mathematical figure (cont'd)	The intersection of PA with r, lying in α above π, is marked by a solid point labeled "P", with a small red square indicating that PA is perpendicular to r within plane α. In summary, the figure includes horizontal plane π, line s on π, point A (intersection of s with α's trace on π), vertical plane α, vertical dashed line r, point P (intersection of r with PA in α), segment PA from A to P perpendicular to s in π, and right angles marked at A and P.

Analysis: The first example accurately describes a triangle with all elements correctly reported. The Model perform well on plane figures, particularly with textual and numerical values present. The **pyramid example** reveals context's importance: without it, o4-mini misinterpreted measurements (incorrectly calculating height as 26 cm); with context, the description became accurate despite minor terminological imprecision, calling the apothem an "oblique edge." The third example shows clearer difficulties with limited spatial understanding: the model incorrectly described plane α as a parallelogram (visually a rectangle), missed perpendicularity between planes, and misplaced point P outside plane π, compromising the entire description. These examples demonstrate that visual clarity is essential for accurate descriptions. Complex three-dimensional diagrams require abstraction that challenges artificial vision systems, unlike human perception.

Diagrams: The single-call pipeline effectively processes these simple, text-supported representations that align with o4-mini's prior knowledge, demonstrating good robustness.

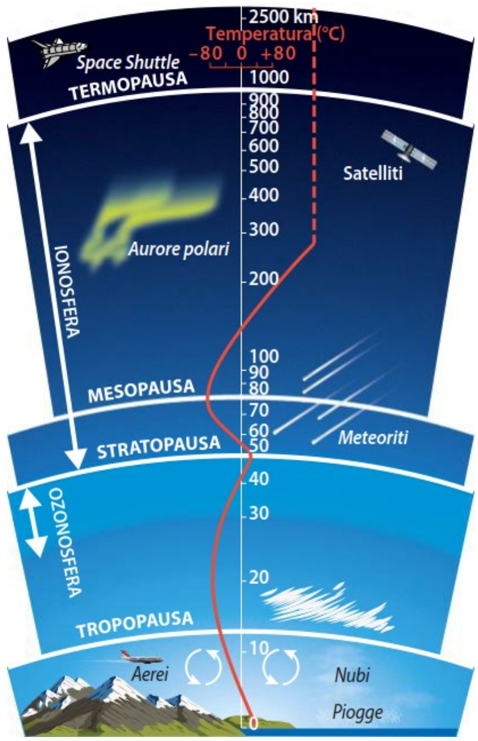

Fig. 13. Diagram showing the layers composing the atmosphere.

Table 18.

Alt Text	**Extended Description**
Diagram of the Earth's atmosphere showing the main layers, the temperature profile, and typical phenomena such as clouds, auroras, meteors, and satellites.	Diagram. Vertical representation of the Earth's atmospheric layers from 0 to 2,500 km, showing the temperature profile and main phenomena and objects present at each level. The vertical axis shows heights in kilometers: from 0 to 100 km with ticks at 10, 20, 30 … up to 100 km, then 200, 300 … up to 1,000 km, with an upper indication up to 2,500 km. The temperature scale on the right shows values in degrees Celsius: –80, 0, +80, with a curved line starting from a positive value at the ground, decreasing to the tropopause, rising in the stratosphere, dropping again in the mesosphere to about –80 °C at the mesopause, then rising above +80 °C in the upper thermosphere. The atmospheric layers and pauses include: troposphere (0–12 km), tropopause (10–12 km), stratosphere (12–50 km), stratopause (50 km), mesosphere (50–80 km), mesopause (80 km), thermosphere (80–1,000 km), thermopause (around 1,000 km), with the ionosphere indicated by a lateral arrow including mesosphere

Alt Text	**Extended Description**
Diagram of the Earth's atmosphere (cont'd)	and thermosphere, and the ozonosphere indicated by a lateral arrow in the stratosphere (30–50 km). Phenomena and objects illustrated are: at ground level (0 km) a mountain range; in the troposphere, clouds, rain, convective currents (circular arrows), and an airplane around 10 km; stratosphere has no particular objects; mesosphere shows falling meteors; lower thermosphere (100–300 km) shows polar auroras; middle thermosphere, a satellite in orbit; and thermopause/upper thermosphere, the Space Shuttle. **Context**: FIGURE 30 The injuries inflicted on the Earth by human activities. [NASTCO / ISTOCK PHOTO] BEFORE CONTINUING The Earth is surrounded by a layer of gaseous substances called the atmosphere, held by gravity. The density of the atmosphere decreases with increasing distance from the Earth. In the first 100 km there is nitrogen (N_2 -- 78%), oxygen (O_2 -- 21%), argon (Ar), carbon dioxide (CO_2), and other gases. Water vapor and dust are present only in the lower part of the atmosphere. The atmosphere is divided into layers separated by pauses. From bottom to top: troposphere (0–12 km), contains 75% of the air and hosts weather phenomena; tropopause (10–12 km); stratosphere (12–50 km), where gases are more rarefied and temperature increases due to the ozone layer (30–50 km); stratopause (50 km); mesosphere (50–80 km), which hosts ionized gases from solar radiation; the temperature decreases from the ozone layer; mesopause (80 km); thermosphere (80–1,000 km), in which gases are increasingly rarefied and have high temperatures; exosphere (from 1,000 km), containing hydrogen (H_2) and helium (He), merging into interstellar space. Due to the presence of ionized gases, the mesosphere and thermosphere constitute the ionosphere.

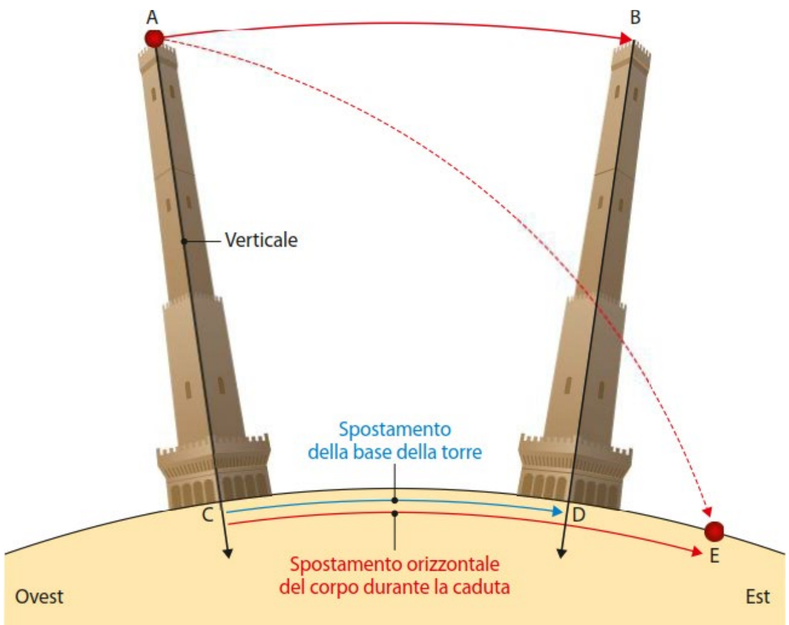

Fig. 14. Diagram showing the effects of Earth's rotation on the trajectory of a body falling from the height of a tower.

Table 19.

Alt Text	Extended Description
Diagram showing the fall of an object from two positions of an inclined tower, highlighting horizontal displacement and Earth's curvature.	Diagram. Two inclined towers are shown, standing on a portion of the Earth's surface represented with exaggerated curvature. On the left tower, the top point A is marked; on the right tower, the top point B is marked. From A, a continuous arc follows the Earth's curvature to B, indicating the angular displacement of the top of the tower during Earth's rotation. From A, a descending dashed arc goes down to the right and ends at point E, representing the actual trajectory of a falling ball. Inside each tower, a straight vertical line is drawn, labeled "Vertical". At the base of the left tower, at the vertical line, there is point C; at the base of the right tower, at the vertical line, there is point D. On the curved surface, between C and D, a horizontal segment with an arrow toward D is drawn, labeled "Displacement of the tower base". Below, between C and a point under E, a second horizontal segment is drawn, labeled "Horizontal displacement of the object during the fall". At the lower edges of the surface, "West" appears on the left and "East" on the right.

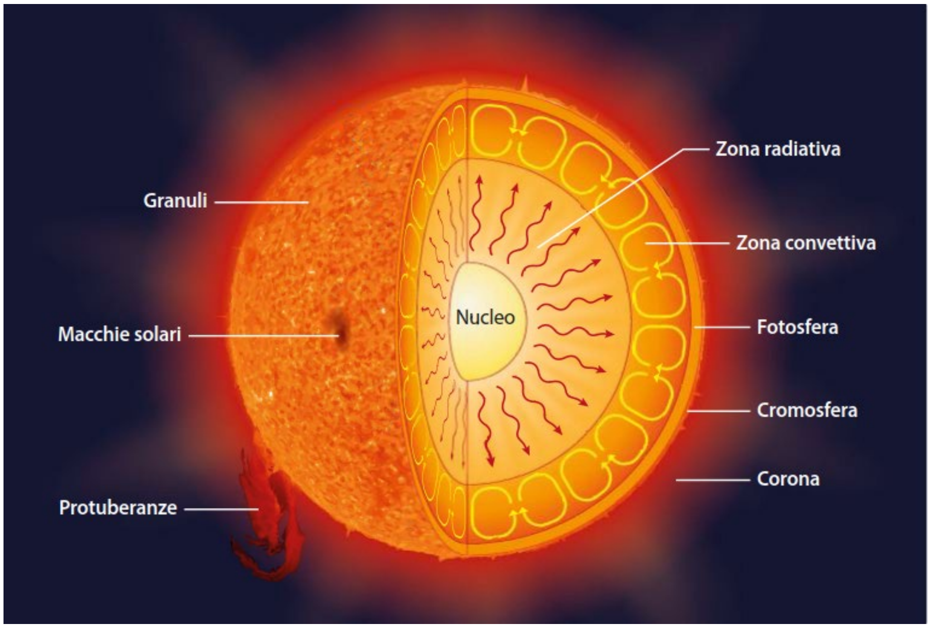

Fig. 15. Diagram showing the regions that make up the Sun.

Table 20.

Alt Text	**Extended Description**
Diagram of a cross-section of the Sun showing the main internal and external regions, from the core to the solar corona.	Diagram. The drawing shows the Sun divided into two vertical halves. Left hemisphere (external view) has a dotted surface labeled "Granules", darker areas labeled "Sunspots", and a prominence extending outward labeled "Prominences". Right hemisphere (internal section) shows at the center a circular area labeled "Core", around it the "Radiative Zone" represented by wavy arrows pointing outward, immediately outside the "Convective Zone" indicated by circular arrows representing matter motions, beyond the convective zone a reference line for the "Photosphere", further out the "Chromosphere", and finally on the outside the "Corona".

Analysis: The second and third diagram examples produced complete, satisfactory descriptions. However, the first example shows a significant omission: the final atmospheric layer, though present in context, was not mentioned. This likely stems from the non-deterministic nature of text generation, which sometimes omits details despite their presence in context or images. While such omissions often involve secondary information, missing the final atmospheric region reduces completeness when describing the image's full structure. Additional issues include a simplified temperature trend description lacking detail across atmospheric layers, suggesting the model extracted only limited information. Despite instructions not to repeat contextual information, the model reported atmospheric layer altitudes—unexpected behavior, adding unnecessary redundancy.

Chemistry figures: The single-call pipeline handles chemistry figures effectively, as these images present relatively low difficulty. o4-mini handles text and technical notation well, thanks to its training, and performs satisfactorily, as demonstrated in the examples.

Analysis: Both images were accurately described without difficulty. This illustration type—featuring clear structure, textual references, and organized visuals—presents minimal challenges and produces good descriptions.

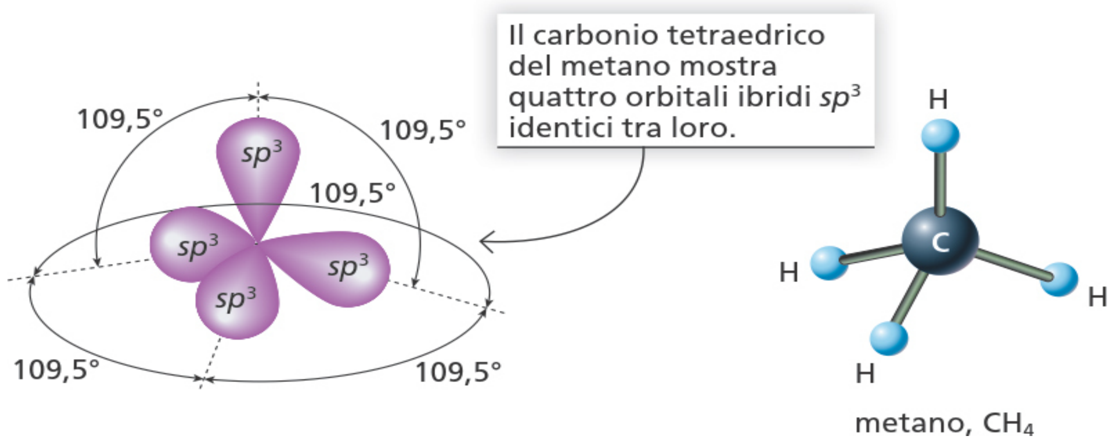

Fig. 16. Chemistry figure showing the graphical representation of the composition of two molecules.

Table 21.

Alt Text	**Extended Description**
Chemistry figure: representation of the sp³ hybrid orbitals of carbon in methane and a ball-and-stick model of CH_4, highlighting the tetrahedral angles of 109.5°.	Chemistry figure. The image consists of two adjacent parts. Left: Diagram of the four sp³ hybrid orbitals of carbon in methane, showing four identical lobes each labeled "sp³", arranged tetrahedrally with one lobe pointing upwards and three on a plane indicated by a circle; between each pair of lobes the angle "109.5°" is indicated with arrows, and a text box states "The tetrahedral carbon in methane shows four identical sp³ hybrid orbitals." Right: Ball-and-stick model of methane with a central atom labeled "C" and four peripheral atoms labeled "H" connected via sticks; below the model, the caption reads "methane, CH_4".

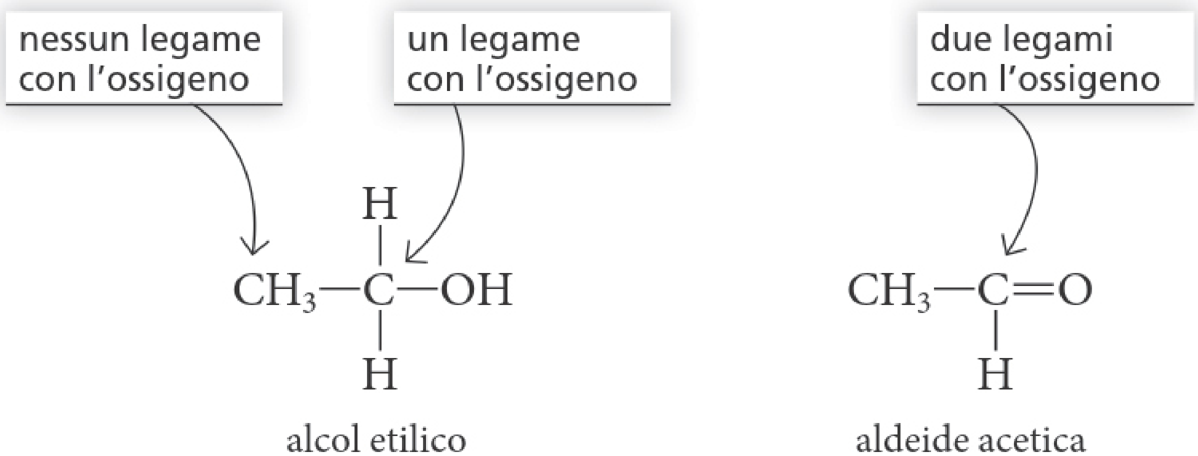

Fig. 17. Chemistry figure showing the bonds between two types of substances.

Table 22.

Alt Text	Extended Description
Chemistry figure: comparison between ethanol and acetaldehyde, highlighting the number of carbon-oxygen bonds.	Chemistry figure. On the left, the structure of ethanol is shown: a central carbon atom connected on the left to a CH_3 group (ethyl) indicated by an arrow with the text "no bond with oxygen"; above and below to hydrogen atoms (H); and on the right to an OH group (hydroxyl) indicated by an arrow with the text "one bond with oxygen". On the right, the structure of acetaldehyde is shown: a CH_3 group bonded to a carbon atom that is bonded below to a hydrogen atom (H) and double-bonded on the right to an oxygen atom (C=O), indicated by an arrow with the text "two bonds with oxygen".

Visual category–based overview: Category-level analysis shows highest error incidence in mathematically dense visuals: math figures report an error rate of 88% with mean quality 67%, reflecting misread coordinates and quantitative reasoning failures. Charts and function graphs remain sensitive to absent numeric anchors, with error rates of 23% and 33% yet relatively high-quality scores (88% and 85%), indicating trend capture with value approximations. Illustrated diagrams and chemistry figures attain strong mean quality (83% and 84%) despite moderate errors, benefiting from clear labeling and structured layouts. Photographs show low errors (16%) and high quality (94%), while logos, maps, and screenshots present 0% errors and 100% quality.

Table 23.

Categories	TOT Omissions	TOT Hallucinations	TOT Inaccuracies	TOT Wrong Categories	TOT Records	TOT Errors	Error rate (%)	Quality Score Mean (%)
Tables	0	0	0	1	1	1	100.0	75.0
Concept Map	0	0	0	35	36	35	97.22	75.69
Math Figure	6	3	3	22	26	23	88.46	67.31
Complex Image	1	0	1	31	39	31	79.49	78.85
Icon	0	0	0	2	3	2	66.67	83.33
Text Image	1	1	1	9	14	9	64.29	78.57
Illustrated Diagram	4	0	3	44	77	47	61.04	83.12
Chemistry Figure	0	3	1	28	53	32	60.38	84.91
Diagram	1	1	1	12	25	12	48.0	85.0
Function Graph	1	1	1	2	12	4	33.33	85.42
Illustration	1	0	0	1	7	2	28.57	92.86
Chart	0	3	2	0	13	3	23.08	88.46
Photograph	2	1	1	3	30	5	16.67	94.17
Logo	0	0	0	0	2	0	0.0	100.0
Map	0	0	0	0	1	0	0.0	100.0

Categories	TOT Omissions	TOT Hallucinations	TOT Inaccuracies	TOT Wrong Categories	TOT Records	TOT Errors	Error rate (%)	Quality Score Mean (%)
Screenshot	0	0	0	0	1	0	0.0	100.0
Total	**17**	**13**	**14**	**190**	**340**	**206**	**60.59**	**82.50**

During the experimental phase, a domain expert evaluated the alt text and extended descriptions following the Risi (2025a) guidelines, with particular attention to STEM images and their long descriptions, an aspect that constitutes the starting point of the present study. An accessibility-specialized organization also reviewed alt texts and extended descriptions and provided further feedback to refine the process. The guidelines were subsequently shared with the digital editors responsible for quality control. To date, no hallucinations or major issues have been detected in the alt texts, and the feedback has been consistently positive: editors report satisfaction with the quality of the results, and they have also expressed a positive attitude toward using LLMs for additional experiments within the publishing house. This workflow ensures that all images are accompanied by both alt text and extended descriptions. A user-testing phase involving participants with visual impairments is also planned, in order to assess whether the descriptions effectively meet their accessibility needs.

Model Comparison

The comparative evaluation of six multimodal large language models. o4-mini, gpt-4o-latest, o3, gpt-4o, gpt-4.1, and Claude Opus 4, reveals significant differences in their capacity to interpret visual content and generate accurate STEM descriptions. While all models successfully captured the overall trend of test functions, only the most advanced models demonstrated precise identification of critical features such as intersection points and extrema coordinates. To provide clear insight into these performance variations, we present one challenging function plot example from our comprehensive testing across multiple samples and diverse prompting techniques.

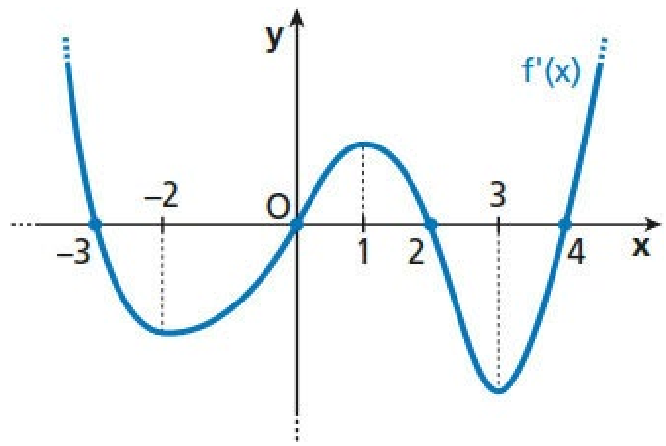

Fig. 18. Graph of a continuous, undefined polynomial function.

Table 24.

Model	Model Response
o4-mini	The figure shows the graph of a function labeled f'(x) (in blue), plotted on a Cartesian plane with horizontal axis x and vertical axis y. The origin is marked as O; axes have arrows at their ends. Certain Elements: * *Axes and Labels:* x-axis marked at -3, -2, 1, 2, 3, 4; y-axis with arrow upwards, no numerical values shown. * *Curve f'(x):* continuous blue in central portion; dashed at left (x ≈ -3) and right (x ≈ 4) ends to indicate continuation. Label "f'(x)" at top-right. * *Notable Points:* local minimum around x=-2; intersection with x-axis at x=0; local maximum between x=1 and x=2; second x-axis intersection around x=2; second local minimum at x=3; curve continues beyond x ≈ 4 (dashed). * *Vertical Dashed Lines:* at x=-2 and x=3, highlighting local minima of f'(x). Ambiguous or Uncertain Elements: * *Vertical dashed lines:* possibly indicate critical x-values for f (e.g., inflection points) or simply guide the eye to minima of f'; not asymptotes. * *Exact values of f'(x):* y-axis lacks graduations, precise maxima and minima unknown. * *Symmetry:* graph not symmetric; apparent wave-like repetition is not mirrored, so no true symmetry. * *Horizontal dashed lines:* none significant besides the dashed curve itself; x-axis continuous. Conclusions: Vertical dashed lines likely indicate minima of f'(x), points where f''(x)=0 and possible inflections of f(x). Graph shows zeros of f' (x=0 and x≈ 2), two minima, and one local maximum. No evidence of asymptotes or perfect symmetry; dashed edges indicate general extension of the graph.

Model	Model Response
gpt-4o	The image shows the graph of the derivative f'(x) of a function f(x), plotted with respect to the axes x (horizontal) and y (vertical). Key elements of the graph: 1. Ordered points: * The graph intersects the x-axis at x = -2, x = 0, x = 2, and x = 3, which are the points where f'(x) = 0. These may correspond to critical points of f(x) (maxima, minima, or horizontal inflection points). * The defined extremes of the graph are approximately x = -3 on the left and x = 4 on the right. 2. Behavior of the graph: * Between x = -3 and x = -2, f'(x) > 0; the slope of f(x) is positive (increasing). * Between x = -2 and x = 0, f'(x) < 0; f(x) is decreasing. * Between x = 0 and x = 2, f'(x) > 0; f(x) increases again. * Between x = 2 and x = 3, f'(x) < 0; f(x) is decreasing. * Between x = 3 and x = 4, f'(x) > 0; f(x) increases again. 3. Values and shapes of the derivative f'(x): * f'(x) takes both positive and negative values, and equals 0 at x = -2, 0, 2, 3. * The graph shows two points where f'(x) reaches a local maximum or minimum: one positive before x = 0 and one negative just before x = 3. 4. Symmetry and regularity: * The graph does not appear to have any clear symmetry with respect to the axes or other points. * The overall shape of the graph is an oscillating curve between positive and negative values, without noticeable irregularities. 5. Analytical hypotheses: * The zeros of the graph (f'(x) = 0) indicate the critical points of the function f(x), i.e., where maxima, minima, or inflection points may occur. Conclusion: The graph represents a continuous and regular behavior of f'(x). The analysis of the signs and zeros provides information on the increasing and decreasing behavior of the original function f(x).
gpt-4.1	Evident and certain elements: 1. Function graph: * A continuous and regular curve is depicted, labeled "f'(x)" in light blue, indicating that this is the graph of the first derivative of a function f(x). 2. Cartesian axes: * The x and y axes are clearly indicated with proper notation. * The axes intersect at the origin (0,0). 3. Notable points: * The curve intersects the x-axis at three visible points: x = 0, x = 2, x = 4. Hence, f'(0) = 0, f'(2) = 0, f'(4) = 0. 4. Curve behavior: * For x between -3 and -2, the curve decreases (negative values), reaches a minimum, then rises (positive values), reaches a maximum, and descends again crossing the x-axis. * A second local minimum occurs between x = 2 and x = 4. 5. Subdivision of the x-axis: * The x-axis shows marks and vertical gray dashed lines at x = -2, 0, 2, 3, which seem to serve as references for characteristic points of the graph. Ambiguous aspects: * No true symmetry of the graph is observed, although some portions visually resemble each other.

Model	**Model Response**
chatgpt-4o	The image shows the first derivative graph f'(x) of a function f(x). Below is a detailed description, distinguishing certain observations from interpretative ones. Certain observations: * General aspect: A continuous curve crosses the x-axis four times at x = -2, 0, 2, and approximately 3.5. * Graph meaning: Where f'(x) = 0, the original function f(x) has critical points (local maxima, minima, or horizontal inflection points). Where f'(x) > 0, f(x) increases; where f'(x) < 0, f(x) decreases. * Intervals of increase/decrease: * f'(x) > 0 (increasing f(x)): x ≈ -2 to 0, x = 2 to x ≈ 3.5 * f'(x) < 0 (decreasing f(x)): x < -2, 0 < x < 2, x > 3.5 * Critical points: * x = -2: negative to positive → local minimum * x = 0: positive to negative → local maximum * x = 2: negative to positive → local minimum * x ≈ 3.5: positive to negative → local maximum Interpretative or uncertain observations: * Rightmost zero at x ≈ 3.5 is approximate, visually estimated. * Inflection points of f(x) require analyzing f''(x). Relative maxima/minima in f'(x) suggest potential inflections in f(x). Symmetry and regularity: No evident symmetry is observed; the curve is generally oscillatory but not guaranteed to be periodic. Conclusion: The graph depicts f'(x). From it, one can infer intervals of increase/decrease and local maxima/minima of f(x). Determining concavity or inflection points requires analyzing the second derivative.
o3	CERTAIN OBSERVATIONS (unambiguous) 1. Reference system * A Cartesian plane is shown, with axes labeled “x” (horizontal, arrow to the right) and “y” (vertical, arrow upward). * The origin is marked with the letter “O”. 2. Blue curve * A single continuous blue curve is labeled f'(x), thus presented as the derivative of a function f. * The curve intersects the x-axis at four points, marked with solid blue dots: x = -2, x = 0, x = 3, x = 4 (values written next to or below each dot). * On the far left and far right, the curve rises steeply towards large positive y values; the ends are cut off with short vertical dashed marks, indicating that the graph continues beyond the frame. 3. Notable points and local behavior * Between x = -2 and x = 0, the curve dips below the x-axis, reaches a minimum around y = -3, then rises again. * Between x = 0 and x = 3, the curve stays above the x-axis, reaches a maximum (approximately at x ≈ 1.5), then decreases again. * Between x = 3 and x = 4, the curve dips below the x-axis, touches a second minimum, and rises back above the axis at the root x = 4. 4. Dashed lines * Three vertical dashed segments descend from the x-axis downward at (or slightly to the left of) x = -2, x = 0, and x = 3. * A longer vertical dashed line coincides with the y-axis below the origin. 5. Numerical labels on the y-axis * The number “-3” is visible on the y-axis, corresponding to the left-hand minimum of the curve.

Model	Model Response
o3 (cont'd)	UNCERTAIN OBSERVATIONS OR HYPOTHESES (with justification) A. Exact coordinates of extrema A full scale is not shown; therefore, the abscissas of maxima/minima (other than the roots) and the ordinates (except for the indicated value -3) are only qualitative estimates. B. Meaning of the dashed lines * It is plausible that they serve only as "guides" to highlight roots and minima on the same vertical (hypothesis), not as asymptotes. Their limited extension and the absence of arrows or a distinct color support this interpretation. * The dashed line on the y-axis below the origin appears simply as a continuation of the axis, not as a marker of discontinuity. C. Symmetry At first glance, the left part (between x ≈ -3 and x = 0) and the right part (between x = 3 and x ≈ 4.5) appear similar in shape (peaks/troughs), but the intervals are not of equal width and the heights do not match: hence this is only a visual resemblance, not a true symmetry (hypothesis confirmed by the absence of an evident symmetry axis). D. "Vertical" dotted ends The short vertical dashed marks at the edges may indicate that the graph continues beyond the frame without an upper bound (hypothesis), but they do not provide information about asymptotic behavior. E. Role of the value -3 The number -3 likely indicates simply the ordinate of the left-hand minimum (probable observation, though not 100% certain since it is not accompanied by other scale marks). FINAL ASSESSMENT OF REGULARITY AND SYMMETRIES * No strict symmetry emerges (neither with respect to the y-axis, nor with respect to a point). * The curve shows a regular (continuous and smooth) behavior with changes in concavity and slope: this is consistent with the fact that it represents a derivative. * The dashed lines are best interpreted as auxiliary guides, not as asymptotes. In summary, the image depicts the graph of f'(x) with four zeros, two minima and one maximum, indicative but not numerically exact values for all extrema, and dashed lines used as simple visual references.
Claude Opus 4	The image is analyzed distinguishing between certain elements and possible interpretations. Certain and Evident Elements: * Cartesian coordinate system: * Horizontal x-axis with arrow to the right * Vertical y-axis with arrow upward * Clearly marked origin O * Function f'(x): Continuous smooth blue curve labeled "f'(x)" (first derivative of a function f) * Values on x-axis: Points marked: -3, -2, 1, 2, 3, 4 * Intersections with x-axis: Curve crosses x-axis at three points: * Near x = -2 * x = 1 (certain) * x = 3 (certain) * Curve behavior: * For x < -2: f'(x) > 0 * Between -2 and 1: f'(x) < 0 * Between 1 and 3: f'(x) > 0 * For x > 3: f'(x) < 0 Uncertain or Interpretative Elements: * Vertical dashed lines: Appear at x = -2, 1, 2, 3; likely indicate: * Critical points of the original function f(x) (f'(x) = 0) * Visual guides for significant values *

Model	Model Response
Claude Opus 4 (cont'd)	End behavior: Curve seems to rise for $x \to -\infty$ and $x \to +\infty$, suggesting $f'(x) \to +\infty$ * Local maximum: Around $x = 2$, $f'(x)$ appears to have a local maximum * Symmetry: No clear symmetry observed with respect to the origin or y-axis Mathematical Interpretation (Hypothesis Based on Observation): * Since this is the graph of $f'(x)$, the original function $f(x)$ can be analyzed as follows: * Stationary points (local maxima or minima) where $f'(x) = 0$: near $x = -2, 1, 3$ * $f(x)$ decreases where $f'(x) < 0$ (between -2 and 1, and for $x > 3$) * $f(x)$ increases where $f'(x) > 0$ (for $x < -2$ and between 1 and 3) * The presence of dashed lines reinforces the hypothesis that these are significant points for the analysis of $f(x)$

This single instance exemplifies broader patterns observed throughout our evaluation. The o4-mini and o3 models emerged as superior performers, with o4-mini correctly identifying the positions of local minima and detecting maximum points while assigning appropriate confidence scores to uncertain elements. Although these leading models contained some inaccuracies in coordinate specification, they crucially avoided hallucinations in recognized features and demonstrated genuine interpretative understanding of image content rather than mere pattern detection. In contrast, most other models introduced systematic hallucinations by misassigning textual coordinate references, a critical flaw in STEM contexts where precision is non-negotiable. The performance differential stems from architectural advantages: both o3 and o4-mini integrate reasoning mechanisms directly within image analysis and provide native manipulation tools such as cropping, zooming, and rotation that enhance feature extraction even from sub-optimal inputs. Given its comparable performance to o3 combined with lower computational cost and superior cost-effectiveness, o4-mini proved the optimal choice for this application and became the foundation for our extended alt text generation pipeline.

Conclusions

This study highlights that accessible STEM images, especially graphs and visualizations, are vital for epistemic equity. Scientific images convey complex data that simple alt text often misses; richer descriptions improve understanding. Visual Language Models (VLMs) show promise for automation, but have limits. Hallucinations decrease when focusing on visual details and using prompt-chaining, yet VLMs struggle with quantitative reasoning in complex graphs and tables. Thus, STEM accessibility cannot rely solely on AI. Hybrid approaches combining AI and human review offer the most reliable and ethical solution.

Acknowledgments

Thanks to Zanichelli Editore, these were available without limits, letting the study take its own shape, as art for art's sake, research should be pursued for research's sake. Special thanks go to Andrea Malfer Poma, who gave the freedom to explore, make mistakes, and learn from them. This openness allowed the seed of knowledge to be shared without restrictions, carried forward to the CSUN conference and beyond. Thanks also to the Iconographic Office of Zanichelli for their support in coordinating and granting permission for the use of the images featured. The images were sourced from the following volumes:

- Carlo Bertoni, Yeap Ban Har, Joseph Yeo, and Andrea Kang, *Pensaci!*, 2019, Zanichelli Editore.
- Beatrice Panebianco, Sara Frigato, Caterina Bubba, Roberto Cardia, Antonella Varani, *Limpida Meraviglia*, 2020, Zanichelli Editore.
- Francesco Iarrera, Giorgio Pilotti, *Paesaggi del turismo*, 2022, Zanichelli Editore.
- Alessandro Barbero, Chiara Frugoni, Sergio Luzzatto, Carla Sclarandis, *La storia racconta*, 2023, Zanichelli Editore.
- Alfonso Bosellini, *Le scienze della terra*, 2020, Zanichelli Editore.
- Giuseppe Valitutti, Patrizia Amadio, Marco Falasca, *Chimica: concetti e modelli*, 2023, Zanichelli Editore.

Works Cited

Allen Institute for AI. "HCI Alt Text Dataset." 2022, https://github.com/allenai/hci-alt-texts/blob/main/data/hci-alt-text-dataset-20220915.jsonl.

An, Wenbin, et al. "Mitigating object hallucinations in large vision-language models with assembly of global and local attention." *Proceedings of the Computer Vision and Pattern Recognition Conference*. 2025.

Anukriti, Kumar; Lucy, Lu Wang. 2024. "Uncovering the New Accessibility Crisis in Scholarly PDFs: Publishing Model and Platform Changes Contribute to Declining Scholarly Document Accessibility in the Last Decade." In Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '24). Association for Computing Machinery, New York, NY, USA, Article 36, 1–16. https://doi.org/10.1145/3663548.3675634.

Bendel, O., 2024. "How Can Generative AI Enhance the Well-being of Blind?" *Proceedings of the AAAI Symposium Series*, 3, pp. 340–347. https://doi.org/10.1609/aaaiss.v3i1.31232.

Chen, Xuweiyi, et al. "Multi-Object Hallucination in Vision Language Models." *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, edited by Amir Globersons et al., 2024, https://doi.org/10.48550/arXiv.2407.06192.

Chen, Zhe, et al. "How Far Are We to GPT-4V? Closing the Gap to Commercial Multimodal Models with Open-Source Suites." 2024, https://arxiv.org/abs/2404.16821.

Crane, Elizabeth, et al. "Alternative Text Quality in STEM Education: Challenges and Opportunities." *Journal of Accessible Technology*, vol. 15, no. 2, 2023, pp. 45–62.

European Union. "Regulation (EU) 2024/1689 of the European Parliament and of the Council on Artificial Intelligence." 2024, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689.

Gao, Yifei, et al. "AIGCs Confuse AI Too: Investigating and Explaining Synthetic Image-Induced Hallucinations in Large Vision-Language Models." *CoRR*, abs/2403.08542, 2024, https://doi.org/10.48550/ARXIV.2403.08542.

GPT-4o System Card. 2024, https://arxiv.org/abs/2410.21276.

Guo, Anhong, et al. "Toward Fairness in AI for People with Disabilities: A Research Roadmap." *CoRR*, abs/1907.02227, 2019, http://arxiv.org/abs/1907.02227.

Ha, S., Monadjemi, S., and Ottley, A., 2024. "Guided By AI: Navigating Trust, Bias, and Data Exploration in AI-Guided Visual Analytics." arXiv e-prints.

Huang, Kung-Hsiang, et al. "Why Vision Language Models Struggle with Visual Arithmetic? Towards Enhanced Chart and Geometry Understanding." *arXiv preprint arXiv:2502.11492* (2025).

Jung, Crescentia, et al. "Communicating Visualizations without Visuals: Investigation of Visualization Alternative Text for People with Visual Impairments." *IEEE transactions on visualization and computer graphics* 28.1 (2021): 1095-1105.

Kalai, Adam Tauman et al. "Why Language Models Hallucinate." *ArXiv* abs/2509.04664 (2025): n. pag.

Maurizio Leotta and Marina Ribaudo. 2024. "Evaluating the Effectiveness of STEM Images Captioning." In Proceedings of the 21st International Web for All Conference (W4A '24). Association for Computing Machinery, New York, NY, USA, 150–159. https://doi.org/10.1145/3677846.3677863.

LIA, Fondazione. "Indagine Di APACE: Accessibilità Nell'editoria in Europa." 2024, https://www.fondazionelia.org/progetti/indagine-di-apace-accessibilita-editoria-europa/.

Morelli, Caterina. "Enhancing Accessibility in Publishing: Leveraging GAI for Effective Alt-Text Solutions." *Technology for Inclusion and Participation for All: Recent Achievements and Future Directions*, edited by K. Mavrou and P. Encarnação, Springer Nature Switzerland AG, 2025, pp. 383-90.

Mozilla Foundation. "Mozilla Alt-Text Validation Dataset." 2023, https://huggingface.co/datasets/Mozilla/alt-text-validation.

Othman, Achraf, et al. "Accessible Metaverse: A Theoretical Framework for Accessibility and Inclusion in the Metaverse." *Multimodal Technol. Interact.*, vol. 8, no. 3, 2024, p. 21, https://doi.org/10.3390/MTI8030021.

Perkins School for the Blind. "How to Write Alt Text and Image Descriptions for the Visually Impaired." 2023, https://www.perkins.org/resource/how-write-alt-text-and-image-descriptions-visually-impaired/.

Platzer, Michael, and Thomas Reutterer. "Holdout-Based Fidelity and Privacy Assessment of Mixed-Type Synthetic Data." *arXiv Preprint arXiv:2104.00635*, 2021, https://arxiv.org/abs/2104.00635.

Risi, Milena. (2025A). "Accessible Images: AI-Generated Alt Text in Educational Publishing: Key Criteria for Creating Alt Text to Enhance Accessibility in Educational Publishing." Logos. 36(1), 50–69. doi.org/10.1163/18784712-20240039.

Risi, Milena. (2025B). "Comparative Study on Gen-AI Models to Write Alt Text." In: Technology for Inclusion and Participation for All: Recent Achievements and Future Directions (pp. 217–223). Springer Nature. doi.org/10.1007/978-3-032-01628-7_27.

W3C. "MathML Gap Analysis and Accessibility Issues." 2023.

Web Content Accessibility Guidelines (WCAG) 2.1. 2018, https://www.w3.org/TR/WCAG21/.

Web Content Accessibility Guidelines (WCAG) 2.1 - Images Tutorial. 2018, https://www.w3.org/WAI/tutorials/images/.

W3C Publishing Maintenance Working Group. "EPUB Accessibility 1.1: Conformance and Discoverability Requirements for EPUB Publications." 2024, https://www.w3.org/TR/epub-a11y-11/.

W3C Web Accessibility Initiative (WAI). "SVG Access: Access to Scalable Vector Graphics (SVG) for Users with Disabilities." 2000, https://www.w3.org/TR/2000/NOTE-SVG-access-20000807/.

Wang, Bin et al. "Image Over Text: Transforming Formula Recognition Evaluation with Character Detection Matching." *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2024): 19681-19690.

Woo, D. J., Wang, D., Yung, T., and Guo, K., 2024. 'Effects of a Prompt Engineering Intervention on Undergraduate Students' AI Self-Efficacy, AI Knowledge and Prompt Engineering Ability: A Mixed Methods Study', arXiv e-prints.

World Health Organization. *World Report on Vision*. WHO, 2022.

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, P., Nie, J.-Y., and Wen, J.-R., 2023. "A Survey of Large Language Models', Journal of Artificial Intelligence Research, 58(4), pp. 2–3.

Zambrano, Ana M., et al. "Math for Everybody: A Sonification Module for Computer Algebra Systems Aimed at Visually Impaired People." *Technologies*, vol. 12, no. 8, 2024, https://doi.org/10.3390/technologies12080133.