AI's Spatial Intelligence: Evaluating AI's Understanding of Spatial Transformations in PSVT:R and Augmented Reality

Uttamasha Monjoree and Wei Yan

Texas A&M University

Abstract

Spatial intelligence is important in many fields such as Architecture, Engineering, and Construction (AEC), Science, Technology, Engineering, and Mathematics (STEM), and Medicine. Understanding three-dimensional (3D) spatial rotations can involve verbal descriptions and visual or interactive examples, illustrating how objects move and change orientation in 3D space. Recent studies show Artificial Intelligence (AI) with language and vision capabilities still face limitations in spatial reasoning. In this paper, we have studied advanced generative AI's spatial capabilities of understanding rotations of objects in 3D space utilizing its image processing and language processing features. We examined the spatial intelligence of the GPT-4 model with vision in understanding spatial rotation process with spatial rotation diagrams based on the Revised Purdue Spatial Visualization Test: Visualization of Rotations (Revised PSVT:R). Furthermore, we incorporated an added layer of a coordinate system axes on Revised PSVT:R to study the variations in GPT-4's performance. We additionally examined GPT-4's understanding of 3D rotations in Augmented Reality (AR) scene images that visualize spatial rotations of a physical object in 3D space and observed increased accuracy of GPT-4's understanding of the rotations by adding supplementary textual information depicting the rotation process or mathematical representations of the rotation (e.g., matrices) superimposed on the object. The results indicate that while GPT-4 as a major current Generative AI model lacks the understanding of a spatial rotation process, it has the potential to understand the rotation process with additional information that can be provided by methods such as AR. AR can superimpose textual information or mathematical representations of the rotations on spatial transformation diagrams or images and create a more intelligible input for AI to comprehend or for training AI's spatial intelligence. Furthermore, by combining the potentials in spatial intelligence of AI with AR's interactive visualization abilities, we expect to offer enhanced guidance for students' spatial learning activities. Such spatial guidance can greatly benefit understanding spatial transformations and additionally support processes like assembly, construction, fabrication, manufacturing, as well as learning in AEC, STEM, and Medicine that require precise 3D spatial understanding and instruction.

1. Introduction

Spatial visualization skills refer to the ability to mentally rotate, manipulate, twist, or invert 3D objects [1], [2]. Education in many STEM disciplines demands instructional approaches that offer experimental practices and tasks with high representational accuracy and realistic simulations to enhance the learning experience [3]. Many first-year undergraduate STEM students struggle with visualization and drawing tasks, even though they have access to various computer assisted learning resources for geometry, spatial transformations, and related mathematical concepts [4]. These difficulties include understanding the symbolic notation of linear algebra, generalizing geometric reasoning, and transitioning to matrix representations of transformations, which often weakens their intuitive grasp of the concepts [5], [6].

Despite studying drawing in secondary school, many students have not sufficiently developed their spatial abilities, leading to significant challenges in mentally manipulating figures in space [7].

Personalized STEM education using interactive technologies can enhance learning experiences. It can promote engagement through hands-on simulations and immersive environments and provide real-time feedback to improve understanding of complex concepts [3]. By leveraging tools such as augmented reality (AR), virtual labs, and Artificial Intelligence (AI)-driven assistance, students can explore STEM topics at their own pace and develop critical problem-solving skills in an interactive, engaging way [8].

Generative AI, encompassing technologies that can create new content such as text, images, 3D models, and audio, have rapidly evolved in recent years, driven by advancements in deep learning and neural network architecture. Large language models (LLMs) such as GPT-4 and its successors have demonstrated remarkable proficiency in generating coherent and contextually relevant text. GPT-4 with Vision - GPT-4V(ision) - has capabilities to analyze user-input images [9],[10]. In an extensive study of GPT-4V(ision), evaluations based on many samples demonstrate GPT-4V's unprecedented ability in understanding visual inputs, spatial relationships of humans and objects, and the genericity of its capabilities [11].

Augmented Reality (AR) can overlay digital information on the real world and create a more perceivable representation for AI. The use of AR-based technologies can reduce the cognitive load of users required for spatial learning by superimposing virtual instruction on the real world. AR has three distinguishing features: combining real and virtual; interactive in real-time; and registered in three dimensions [12]. Using AR in education has several positive impacts: increased content understanding, learning spatial structures, language associations, long-term memory retention, improved collaboration, and motivation [13].

GPT-4's visual capabilities are highly relevant in AR, virtual design, and other applications that require real-time understanding of visual input [14]. For instance, it can assist in designing objects in 3D space[15], guide users through augmented experiences, or help in manipulating visual content interactively [16]. Combined textual and graphical information superimposed in real world utilizing AR technology has the potential to generate a more intelligible input for GPT-4's visual capabilities.

Using conversational AI in education is one of the major approaches to enhancing and promoting a more personalized learning experience [17]. Personalized interactive assistance of generative conversational AI that has visual capabilities along with spatial qualities of AR can help students achieve better learning outcomes. Understanding AI's ability to perceive rotation in AR is expected to have a significant impact on learning spatial visualization skills.

Spatial reasoning is a key area in psychometrics, and several datasets exist for evaluating these skills, such as the Mental Rotation Tests [18], the Purdue Spatial Visualization Test (PSVT)/The Purdue Visualization of Rotations Test [19], and the Revised Purdue Spatial Visualization Test: Visualization of Rotations (Revised PSVT:R) [20]. The Revised PSVT:R is considered one of the most effective tools for evaluating spatial reasoning skills. This test is a widely used assessment designed to measure an individual's ability to visualize and mentally manipulate 3D objects in space, particularly in terms of rotating these objects.

We can use both a verbal description and some visual or interactive examples to explain spatial rotations and convey how objects move or change orientations in 3D space [21]. Combining the visual and verbal abilities of conversational AI and visual and interactive abilities of AR has the potential to provide enhanced guidance in spatial rotation activities [8]. This type of spatial guidance can help mathematical education as well as govern assembly, construction, and fabrication process that requires spatial instructions. The goal of this study is to evaluate AI's capability of understanding 3D spatial rotations by performing a series of

tests based on Revised PSVT:R and an interactive AR application (AR-Classroom) that can visualize rotation along different axes [22]. The research questions that guide this study are as follows:

1. Can GPT-4 answer questions in Revised PSVT:R and how its performance compare with human test takers?
2. Can GPT-4 perceive a spatial rotation process created based on Revised PSVT:R and describe the process in terms of axis of rotation, angle of rotation, and (clockwise or counterclockwise) direction of rotation?
3. Can GPT-4 perceive a spatial rotation process visualized in AR and describe the process in terms of axis of rotation, angle of rotation, and (clockwise or counterclockwise) direction of rotation?

2. Literature Review

Dual-coding theory

The dual-coding theory is defined as a theory that posits the existence of two interconnected subsystems in the human cognitive system: a verbal system and an imagery system [23]. This dual capacity means the language system functions in both communication and symbolic interpretation. Human cognition uniquely integrates both language and nonverbal objects and events, enabling us to process linguistic input (like speech or writing) alongside symbolic representations of nonverbal aspects of the world [24]. Therefore, spatial representation should integrate these two systems, where language serves both direct communication and a framework for understanding nonverbal phenomena. Integrating sophisticated language and image processing and response generation of Artificial Intelligence (AI) with immersive three-dimensional capabilities of AR can facilitate a more engaging spatial learning experience [25]. Including Augmented Reality (AR) in the learning process can improve spatial abilities in the domain of architecture education [26].

AI Vision Model

The latest development of GPT-4, a large-scale, multimodal model, can accept image and text inputs and produce text outputs. While less capable than humans in many real-world scenarios, GPT-4 Vision can analyze images, provide detailed descriptions, identify objects, and offer explanations of scenes [9]. For instance, it can describe an uploaded photo or explain elements within the image, such as identifying objects like plants, vehicles, or people, and providing contextual information like plant care advice or understanding specific items in complex scenes [27]. AI with vision and spatial understanding have broad applications, for example, in STEM and Architecture/Engineering/Construction (AEC). Integrating GPT AI models into construction safety management enhances risk assessment, real-time insights, incident reporting, and compliance evaluation [28], e.g. the Personal Protective Equipment (PPE) or helmet usage compliance [11]. Experiments have proven GPT-4V to be highly effective in detecting and explaining global and fine-grained semantic patterns in anomaly or risk detection [29], [11].

AI's Spatial and Mathematical Abilities

Spatial intelligence is described as a skillset that helps us comprehend visual-spatial tasks and spatial relations. As a result, we gain a better orientation of objects in space [30]. In recent years there have been several studies analyzing spatial and mathematical capabilities of AI. GPT-4 model "hallucinates", wherein the model returned inaccurate information. The model was unable to accurately return bounding boxes for object detection, suggesting it is not completely fit for computer vision related usage [31]. The SPARE3D dataset introduces a series of challenging 2D-3D spatial reasoning tasks, revealing that while convolutional neural networks (CNNs) excel in visual learning, their performance on spatial reasoning remains below human average, highlighting the need for advancements in artificial neural network design to improve 3D spatial reasoning for intelligent robots [32]. A recent study has examined the perceived complexity of 3D

shapes from both human and ChatGPT perspectives, revealing that ChatGPT can effectively capture human consensus on shape complexity. The findings suggest that large generative models like ChatGPT could be used to automatically generate 3D content with varying complexities, which can be enhanced for spatial visualization skills in VR applications [14]. This research explored how the visual features of some of the 3D shapes used in the Purdue Spatial Visualization Test are perceived by ChatGPT.

[33] introduces a novel pipeline for enhancing the geometric capabilities of Multimodal Large Language Models (MLLMs) and Wolfram (a math tool) by generating a dataset of 4.9K simplified, image-text-matching geometric problems combined with 19K open-source data to create a uniformly challenging dataset, GeoGPT4V. Training models on GeoGPT4V has demonstrated improved geometric performance on MathVista and MathVision benchmarks.

[34] conducted a pilot study to explore the potential of using ChatGPT to create thematic and mental maps by utilizing both public geospatial data and textual descriptions of geographic space. They evaluated ChatGPT's performance on a GIS (Geographic Information System) exam, comparing GPT-3.5 and GPT-4, to assess their strengths and limitations in understanding geospatial concepts and discovered challenges in teaching spatial reasoning to AI models. While ChatGPT shows promise in lowering the barrier for map production, enhancing efficiency, and demonstrating spatial thinking capabilities, it still faces limitations such as unequal user benefits and reliance on user intervention for maintaining quality control [35].

ChatGPT-4 and ChatGPT 4o struggle more with geometry and complex math problems compared to algebra, suggesting that while AI can support algebra instruction, teacher intervention is needed in areas where AI is less effective, such as spatial intelligence and higher-difficulty tasks [36]. After an elaborate study, [37] observed LLMs resemble sophisticated pattern matching more than true logical reasoning. LLM performance varies highly on different versions of the same question. LLMs' performance drops substantially with a minor increase in difficulty, and their sensitivity to inconsequential information indicates that their reasoning is fragile.

Despite many positive media reports highlighting GPT-4 and ChatGPT's exam-solving abilities, possibly influenced by a selection bias, their overall mathematical performance falls significantly short of a graduate student's level [38]. Educators can leverage AI to create engaging and effective learning environments, cultivating a generation skilled in mathematics and prepared to navigate the complexities of the modern world by addressing the associated challenges and adopting a refined approach [39].

AR and AI in Education

In recent years, AR and AI are demonstrating the potential to reshape education by enhancing interactivity, personalization, and accessibility. The use of AR technology and speech recognition to support productive vocabulary among students has shown user satisfaction when parents and teachers were interviewed [40]. [41] introduced a novel early education platform that combines smart voice commands and AR technology. The platform is embedded in a storage box, making it convenient for use in homes and kindergartens where preschool education is provided. This project enhances the integration of these technologies in educational settings.

Using AR-powered embodied learning for spatial transformations and representation of corresponding matrices has already shown significant improvement in the participants' math scores [42]. The project AR-Classroom uses embodied learning and novel AR features to visualize spatial rotations and their mathematical representations, and the usability tests that were completed using the application showed promising results. The benchmark test and updated usability test done on 24 participants have shown significant improvement in usability scores [32]. The AR-Classroom application integrates AR and AI-based multi-sensory assistance to help undergraduate students better understand spatial transformations and their mathematical representations. A usability study identified four key themes: ease of use, the need for

more detailed AI responses, a desire for visual information, and improved understanding, highlighting the potential for further AI-guided visual cues in AR learning environments [8].

3. Methodology

When explaining rotation, we use references like direction, angle, and axis verbally and visually. According to the dual-coding theory, two cognitive subsystems, verbal and imagery, work together to process language and nonverbal content. Integrating these systems has the potential to enhance spatial understanding, where language aids in communication and interpreting nonverbal cues.

LLMs rely on descriptive datasets and visual cues, which can enhance spatial descriptions in vision models. Using visual cues to communicate a rotation process could improve interactions between AR and AI systems, providing a more intuitive tool for conveying spatial information, thus facilitating spatial learning. This approach could bridge the gap between text-based AI and spatially oriented AR, making explanations of spatial transformations more accessible and effective. Combined AR and AI have the potential of becoming instrumental in assembly and fabrication, offering real-time guidance, error reduction, and efficiency improvements. In this paper, we examine generative AI's vision and language capacity of understanding and describing spatial rotation using standard test (Revised PSVT:R) vs AR.

Our evaluation focused on the AI's capacity to comprehend key aspects of spatial rotations, including the axis, angle, and direction of each rotation, as well as its ability to process sequences of multiple rotations. We integrated multiple layers of graphical and textual features into the Revised PSVT:R to analyze their impact on the performance of GPT-4 (ChatGPT 4o or GPT-4o). We also tested GPT-4's capability in an AR application that visualizes the spatial rotation of a hand-held object in 3D space and tested the accuracy with the addition of supplementary textual information (angle, textual information, rotation matrix equation etc.)

4. Experiments and Results

Experiment 1: Evaluate GPT-4 Using Revised PSVT:R (original and with added coordinate axes)

The Revised PSVT:R is a widely used test for assessing spatial visualization skills, consisting of rotation problems. In this test, participants are shown diagrams of 3D objects in both a starting and finishing orientation after an unknown rotation along one or more axes. The task is to identify the correct diagram of a second 3D object that has undergone the same rotations as the first. Participants are given the second 3D object in its starting orientation and must choose the correct finishing orientation from five possible options.

Test Using Revised PSVT:R (Test 1a)

Using ChatGPT 4o (with a vision capability) from June 9, 2024 through Fall 2024, we conducted various tests. In the first test, we provided GPT-4 with all the 30 problems from Revised Purdue Spatial Visualization Test: Visualization of Rotations (Revised PSVT:R) as a standard test to study its ability to analyze input images and spatial rotation capabilities (Figure 1a). After getting the input (a screenshot of the test questions) uploaded, GPT-4 responded with answering the questions by choosing from (A, B, C, D, and E options). Comparing GPT-4's results with the correct results on Revised PSVT:R, we found that GPT-4 only got 5 out of 30 answers correct. The 5 correct answers are for Question 10, 15, 16, 19, and 30. (For copyright of Revised PSVT:R and keeping the test results confidential for future tests with participants, the 30 questions and results are not revealed in this paper. An example question in the DIRECTIONS section of the test that is allowed to publish is shown in Figure 1.) See Appendix A for sample interaction between authors and GPT-4 in Test 1a.

After getting unsatisfactory results in the first test, we explored more of its spatial analyzing capabilities by adding a coordinate system with x- y- z-axes onto the Revised PSVT:R question images. The motivation behind this is incorporating more visual and textual cues for analyzing images and enhancing communication with AI. We superimposed the coordinate system on 30 question images of Revised PSVT:R where x-, y-, and z-axes are indicated (Figure 1b). See Appendix B for sample interaction between authors and GPT-4 in Test 1b.
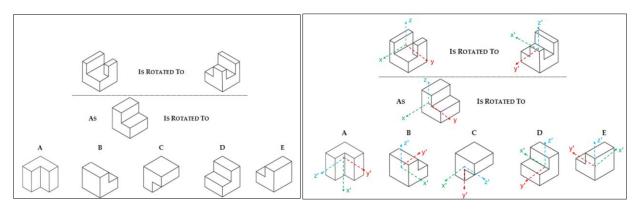
Figure 1: (a) The example question from Revised PSVT:R (left), (b) the example question from Revised PSVT:R with a coordinate system added (right). In both questions, D is the correct answer.

In Experiment 1, both Test 1a and 1b, GPT-4 had the most complete task to identify the correct option from the five 3D object orientations (A to E). To execute the whole task, like a human test-taker, GPT-4 needs to identify the rotation process in Step 1 and then compare it to the five rotation processes in Step 2, in order to choose the correct option (Figure 2). For example, in Figure 2, to answer correctly, one should identify the similarities between the axis systems between Step 1 and Step 2, and choose option D. In Test 1a, GPT-4 gave correct answers only 5 out of 30 questions which reflects 17% accuracy (Figure 3). In Test 1b, GPT-4 had a similar accuracy, which indicates that it is unable to identify and compare rotation processes correctly even with the visual cues (axes) added.
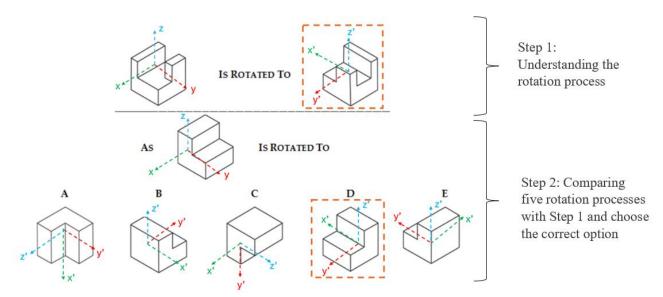
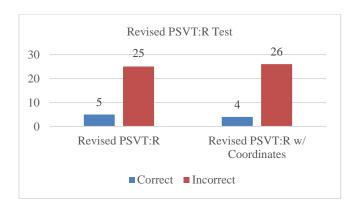Figure 2: Example question in Revised PSVT:R with a coordinates system. D is the correct answer.

Figure 3: Accuracy of GPT-4 in Experiment 1 using Revised PSVT:R: 5 correct out of 30 answers, and Revised PSVT:R with a coordinate system added: 4 correct out of 30 answers.

Experiment 2: Evaluate GPT-4 Using Step 1 of Revised PSVT:R with a Coordinate System (Test 2)

GPT-4 showed an unsatisfactory result in answering the questions of Revised PSVT:R in Experiment 1. To simplify the tasks for AI, we continued testing with only Step 1 of the questions of Revised PSVT:R (Figure 3) to see if AI can identify the rotations of one object in each question. We superimposed a coordinate system with axes on Step 1 of each question of Revised PSVT:R, and asked GPT-4 to explain the rotation process with minimum rotation steps (Figure 4). The resulting orientation in 14 questions of Revised PSVT:R can be achieved by rotating along a single axis and in the rest 16 questions can be achieved by rotating along multiple axes (e.g., rotating around one axis and then around another axis). As the resulting orientation of a 3D object can be achieved using different sequences of multiple rotations, we asked GPT-4 to describe the change of orientation using rotations along a minimum number of axes. We used the prompt, *"Explain the rotation process using minimum steps of rotation in terms of axis of rotation, angle of rotation, and direction of rotation."* GPT-4 performed better in case of the Step 1 rotation process than

in Experiment 1 with both Step 1 and 2. See Appendix C for sample interaction between authors and GPT-4 in Test 2.
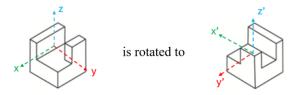

Figure 4: Diagram of Step 1 of Revised PSVT:R with a coordinate system added.

In this test, GPT-4 performed better in case of rotation along a single axis than rotations along multiple axes (Figure 5). In case of a single-step rotation, it was correct 5 out of 14 times reflecting 35% accuracy, in case of rotations along multiple axes, it was correct 2 out of 16 times showing 12.5% accuracy. The combined accuracy of this test is 23.33%. The performance in case of rotation along a single axis is almost 2 times more accurate than the rotations along multiple axes, though in both cases performance by GPT-4 is unsatisfactory. This experiment results indicate that even though GPT-4 cannot compare similar rotation processes, it is somewhat capable of understanding one rotation process. The information we provided is not sufficient for GPT-4 to understand the rotation process. However, there may still be a possibility of achieving better results by adding supplementary information and context.
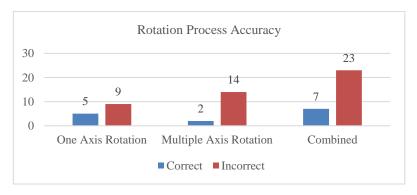

Figure 5: Rotation process accuracy described by GPT-4 using Step 1 of Revised PSVT:R with axes of a coordinate system

Experiment 3: Evaluate GPT-4 on Understanding Rotations with Supplementary Information through an AR Application

AR can superimpose 3D models, data, or text over real-world objects, enhancing the user's understanding of their surroundings and the physical world. An AR application named *AR-Classroom* provides an interactive learning environment with better context and visualization for understanding rotations (Figure 6) [22]. The app aims to run on an AR device, such as an AR-enabled tablet, an AR headset, or a computer with a camera. A user can hold and manipulate a 3D physical model to learn spatial transformations by observing the corresponding mathematical representation and visualization that augment the 3D physical model with a graphical user interface (GUI) in AR.
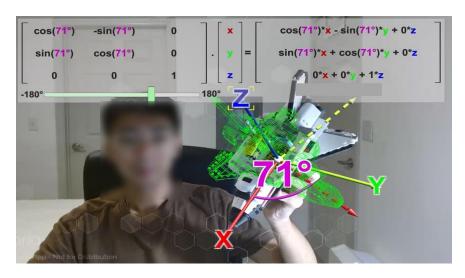
Figure 6: AR-Classroom GUI – augmenting the 3D physical space shuttle model with a 3D virtual space shuttle model, rotation axes, rotation arc and angle, and the matrix representation of the rotation.

The X-, Y-, and Z-axes of the models' coordinate systems and corresponding matrix equations are displayed in the GUI. In Figure 6, the following AR scene components are shown.

1. a physical model, called pre-image of a transformation, showing the state before rotation
2. a 3D reference frame coordinate with red, green, and blue solid arrows for X-, Y-, and Z-axis, attached to the physical model
3. a 3D wireframe model, called image of a transformation, showing the state after rotation
4. a 3D body frame coordinate with red, green, and blue dotted arrows for X-, Y-, and Z-axes, attached to the 3D wireframe model

Test Using AR with a Coordinate System (Test 3a): We uploaded to GPT-4 with 12 screenshots of the AR environment visualizing only the physical and virtual models and the coordinate system. The angle of rotation and corresponding rotation matrix equation were hidden (Figure 7a). All screenshots varied in axis of rotation, angle of rotation, and direction (clockwise or counterclockwise) of rotation. We provided the prompt, "*Explain the rotation process with axis of rotation, angle of rotation and direction. Note: All red lines represent X-axis, all yellow-green lines represent Y-axis, all blue lines represent Z-axis. The dotted line is the orientation of axis after rotation. The axis of rotation does not have a dotted counterpart of the same color. Follow right hand rule for the direction of rotation*."

Test Using AR with a Coordinate System and Angle of Rotation (Test 3b): We uploaded to GPT-4 with 12 screenshots of the AR environment visualizing the physical and virtual models, the coordinate system, and angle of rotation. The corresponding rotation matrix equation was hidden (Figure 7b). All screenshots varied in axis of rotation, angle of rotation, and direction of rotation. We provided the same prompt as in Test 3a.

Test Using AR with Coordinate System, Angle of Rotation, and Rotation Equation (Test 3c): We uploaded to GPT-4 with 12 screenshots of AR-Classroom where the coordinate system, rotation angle, and corresponding rotation matrix equation are all visible (Figure 7c). We provided the same prompt as in Test 3a and GPT-4 gave very satisfactory results each time we tested, as explained below. Furthermore, even with a simpler prompt "*Explain the rotation process using minimum steps of rotation in terms of axis of rotation, angle of rotation, and direction of rotation*", the results are as accurate as the more detailed prompt in Test 3c.

is rotated to

(a)

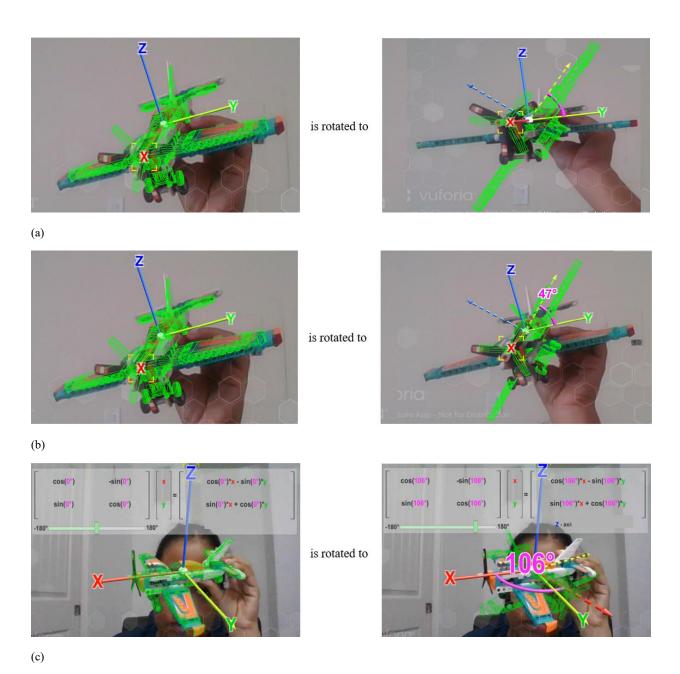
is rotated to

(b)

is rotated to

(c)

Figure 7: (a) AR rotation visualization with a coordinate system, (b) AR rotation visualization with a coordinate system and angle, (c) AR rotation visualization with a coordinate system, angle, and a rotation equation

In this experiment, where we tested different versions of AR-Classroom images, we observed that the performance of GPT-4 gets better when more descriptive texts are added to the visual input. AR rotation with only the coordinate system has 25% accuracy (Test 3a), and AR rotation with a coordinate system and an angle of rotation (Test 3b) has 75% accuracy. When we provided the whole UI of AR-Classroom (AR with a coordinate system, an angle, and rotation equation) as input, the additional information related to the rotation process helped GPT-4 gain 100% accuracy (Test 3c). Here, GPT-4 could identify and connect the rotation matrix equation with the rotation activity, and it described the rotation process correctly. We can see from Table 1 or Figure 8 that making the angle of rotation and the rotation matrix equation visible in

the AR scene each has increased the accuracy significantly. See Appendix D for sample interaction between authors and GPT-4 in Test 3c.

Table 1: Accuracy of identifying rotation process in Test 3a, Test 3b, and Test 3c

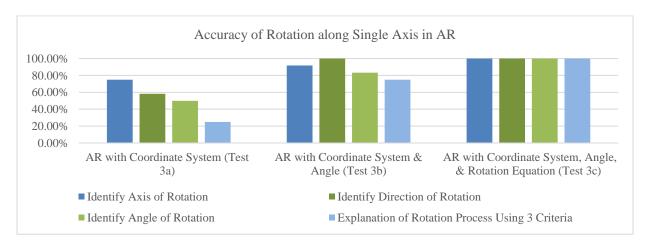|	AR with Coordinate System (Test 3a)	AR with Coordinate System & Angle (Test 3b)	AR with Coordinate System, Angle, & Rotation Equation (Test 3c)
Identify Axis of Rotation	75.0%	91.7%	100.0%
Identify Direction of Rotation	58.3%	100.0%	100.0%
Identify Angle of Rotation	50.0%	83.3%	100.0%
Explanation of Rotation Process Correctly Using all the above Criteria	25.0%	75.0%	100.0%

Figure 8: Accuracy of identifying rotation process in Test 3a, Test 3b, and Test 3c

5. Discussions

In GPT-4, the language model tends to be more accurate for purely text-based tasks, while the vision model is better suited for multimodal tasks involving both images and text [43]. GPT-4V integrates image understanding with text, making it powerful for multimodal tasks like interpreting images alongside text descriptions. This model performs well in visual question answering, image-to-text tasks, and recognizing objects in images. However, its accuracy for purely image-based tasks, like object detection or more complex image analysis, may not match specialized vision models and still relies heavily on textual inputs to guide its performance. This supports what we observed in our experiments: GPT-4 gave inadequate performance in case of Revised PSVT:R (Test 1a) with a score of 17% (5 out of 30). This indicates that it cannot identify complex rotation tasks and cannot identify the similarities between two identical rotation processes with different objects. The cases when it got the right answers could be coincidental. As a comparison, in 1993, a PSVT:R was administered to 535 first-year engineering students at Michigan Tech during orientation. Approximately 20% of the students (n = 96) scored 60% or less on the PSVT:R, while 80% of the students scored above 60% accuracy [44]. In another study [20] using Revised PSVT:R, the mean score of 1,022 participants was 19.08 out of 30 with a standard deviation of 6.16. Scores ranged from 3 to 30, and 13 participants (1.27%) achieved a perfect score.

Furthermore, providing a 3D coordinate system with the original Revised PSVT:R did not show significant improvement (Test 1b).

Test 2 (Experiment 2) showed that there is an improvement of GPT-4 in recognizing Step 1 only in Revised PSVT:R or a simplified rotation process.

In the AR environment, we provided better context than the previous experiments. GPT-4 could see pre-image before rotation and image after the rotation in the AR environment (Test 3a). By increasing textual input step by step in the AR environment, we observed improved performance by GPT-4 (Figure 8). Just including the angle of rotation with the coordinate system showed significant improvement in result (Test 3b). Providing a rotation matrix equation as additional information enabled GPT-4 to identify the rotation process correctly (Test 3c). Even with a less detailed prompt and same image as Test 3c, it identified the rotation process correctly. The use of the AR environment is for the convenience of adding accurate rotation information to the testing for GPT-4. A drawing style of PSVT:R may be used in a similar evaluation process. However, the rendering mode (realistic 3D physical model, wireframe virtual model, colors of rotation arc and angle values, etc.) may have affected the understanding of rotations by GPT-4.

Overall, this result points out that GPT-4 shows limited spatial intelligence in terms of understanding 3D rotations with 2D image inputs. For applications, it is very promising for future development of AR applications in the education and STEM sectors, where GPT-4 can guide a user on task performance by taking real-world data and relevant, superimposed textual and graphical data as input. In an AR setting, users can view physical objects, diagrams, or virtual models while AI provides additional information layered over the scene. For example, in a geometry lesson, a 3D model of a shape could be displayed, and GPT-4 could guide the student by offering real-time feedback on angles, dimensions, or rotation processes, enabling more effective learning. This combination of visual and textual data allows GPT-4 to better understand spatial relationships and provide meaningful insights.

The result also indicates that the current generative AI doesn't perform well regarding spatial intelligence without additional context information (e.g., rotation axis, angle, and equation). The context information can facilitate AI for understanding spatial transformations and can help train AI to improve its spatial intelligence, so that in the future the context information may not be needed any more.

A major limitation of the research is that the sample size (number of each test) is small as this is a very early evaluation of spatial intelligence of AI in terms of 3D rotations.

6. Conclusions and Future Work

Spatial knowledge is critical in fields like architecture, computer graphics, civil engineering, and other STEM disciplines. While generative AI-powered tools can assist students in learning, conversational AI agents still struggle with spatial abilities. AR can enhance these systems by overlaying textual information on spatial diagrams, making the data more comprehensible for AI. This paper investigates GPT-4's spatial capabilities, particularly its proficiency in interpreting 3D rotations, as tested through Revised PSVT:R and rotation images from an AR application. We improvised Revised PSVT:R with additional graphical and textual features to assess if GPT-4's performance changes with supplementary information. GPT-4 has difficulty recognizing complex rotation tasks and identifying the similarities when two rotations are essentially the same but for different objects, as tested in the Revised PSVT:R.

In the AR images, where objects are visualized in 3D space, adding supplementary textual and math information significantly improved GPT-4's accuracy. For example, explicitly showing the angle of rotation

increased accuracy from 25% to 75% in one of the tests. Providing the math equation of rotation in the AR application, along with detailed text, allowed GPT-4 to reach 100% accuracy. This suggests that while GPT-4V performs well in multimodal tasks, it relies heavily on textual cues for better performance. Additional information, as visualized in AR applications, has the potential to enhance the performance of GPT-4. These results indicate that in an AR setting, users can view physical objects, diagrams, or virtual models' additional information layered over the scene while AI can provide real-time feedback observing the user's interaction process.

The future work will include more thorough testing with an increased sample size and with different AI tools, and enhancing the spatial intelligence of AI, so that AI can understand spatial transformations including rotations, translations, scales, perspectives, etc. based on various styles (e.g., real-world photo or drawing styles) of image or video input of arbitrary 3D physical objects, with or without context information of axes, text, math equations, etc.

Acknowledgement

References

[1]   D. Ben-Chaim, G. Lappan, and R. T. Houang, "The Effect of Instruction on Spatial Visualization Skills of Middle School Boys and Girls," *American Educational Research Journal*, vol. 25, no. 1, pp. 51–71, 1988, doi: 10.2307/1163159.

[2]   R. Fleisig, A. Robertson, and A. Spence, "Improving the Spatial Visualization Skills of First Year Engineering Students," *Proceedings of the Canadian Engineering Education Association*, Aug. 2011, doi: 10.24908/pceea.v0i0.4050.

[3]   N. Pellas, A. Dengel, and A. Christopoulos, "A Scoping Review of Immersive Virtual Reality in STEM Education," *IEEE Transactions on Learning Technologies*, vol. 13, no. 4, pp. 748–761, Oct. 2020, doi: 10.1109/TLT.2020.3019405.

[4]   M. Garmendia, J. Guisasola, and E. Sierra, "First-year engineering students' difficulties in visualization and drawing tasks," *European Journal of Engineering Education*, vol. 32, no. 3, pp. 315–323, Jun. 2007, doi: 10.1080/03043790701276874.

[5]   C. Andrews-Larson, M. Wawro, and M. Zandieh, "A hypothetical learning trajectory for conceptualizing matrices as linear transformations," *International Journal of Mathematical Education in Science and Technology*, vol. 48, no. 6, pp. 809–829, Aug. 2017, doi: 10.1080/0020739X.2016.1276225.

[6]   J. Dick and M. Childrey, "Enhancing Understanding of Transformation Matrices," *Mathematics Teacher*, vol. 105, no. 8, pp. 622–626, Apr. 2012.

[7]   C. Potter and E. MERWE, "Perception, imagery, visualization and engineering graphics," *European Journal of Engineering Education*, vol. 28, pp. 117–133, Mar. 2003, doi: 10.1080/0304379031000065216.

[8]   U. Monjoree *et al.*, "AR-Classroom: Integrating Conversational Artificial Intelligence with Augmented Reality Technology for Learning Spatial Transformations and Their Matrix Representation," in *2023 IEEE Frontiers in Education Conference (FIE)*, Oct. 2024.

[9]   OpenAI *et al.*, "GPT-4 Technical Report," Mar. 04, 2024, *arXiv*: arXiv:2303.08774. doi: 10.48550/arXiv.2303.08774.

[10]   "OpenAI Platform." Accessed: Apr. 23, 2024. [Online]. Available: https://platform.openai.com

[11] Z. Yang *et al.*, "The Dawn of LMMs: Preliminary Explorations with GPT-4V(ision)," 2023, doi: 10.48550/ARXIV.2309.17421.

[12] R. T. Azuma, "A Survey of Augmented Reality," *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, pp. 355–385, Aug. 1997, doi: 10.1162/pres.1997.6.4.355.

[13] I. Radu, "Augmented reality in education: a meta-review and cross-media analysis," *Personal Ubiquitous Comput.*, vol. 18, no. 6, pp. 1533–1543, Aug. 2014, doi: 10.1007/s00779-013-0747-y.

[14] L. McShane and C. Lopez, "Perceived Complexity of 3D Shapes for Spatial Visualization Tasks: Humans vs Generative Models," doi: 10.1115/DETC2023-115081.

[15] B. Regassa Hunde and A. Debebe Woldeyohannes, "Future prospects of computer-aided design (CAD) – A review from the perspective of artificial intelligence (AI), extended reality, and 3D printing," *Results in Engineering*, vol. 14, p. 100478, Jun. 2022, doi: 10.1016/j.rineng.2022.100478.

[16] A. Rychert, M. L. Ganuza, and M. N. Selzer, "Integrating GPT as an Assistant for Low-Cost Virtual Reality Escape-Room Games," *IEEE Computer Graphics and Applications*, vol. 44, no. 4, pp. 14–25, Jul. 2024, doi: 10.1109/MCG.2024.3426314.

[17] S. Cunningham-Nelson, W. Boles, L. Trouton, and E. Margerison, "A review of chatbots in education: Practical steps forward," in *30th Annual Conference for the Australasian Association for Engineering Education (AAEE 2019): Educators Becoming Agents of Change: Innovate, Integrate, Motivate*, Australia: Engineers Australia, 2019, pp. 299–306. Accessed: Feb. 29, 2024. [Online]. Available: https://search.informit.com.au/documentSummary;dn=068364390172788;res=IELENG

[18] S. G. Vandenberg and A. R. Kuse, "Mental rotations, a group test of three-dimensional spatial visualization," *Percept Mot Skills*, vol. 47, no. 2, pp. 599–604, Oct. 1978, doi: 10.2466/pms.1978.47.2.599.

[19] G. M. BODNER and R. B. GUAY, "The Purdue Visualization of Rotations Test," *Chem. Educator*, vol. 2, no. 4, pp. 1–17, Oct. 1997, doi: 10.1007/s00897970138a.

[20] S. Y. Yoon, "Psychometric properties of the Revised Purdue Spatial Visualization Tests: Visualization of Rotations (the Revised PSVT:R)," *Theses and Dissertations Available from ProQuest*, pp. 1–180, Jan. 2011.

[21] J. M. Loomis, R. L. Klatzky, and N. A. Giudice, "Representing 3D Space in Working Memory: Spatial Images from Vision, Hearing, Touch, and Language," in *Multisensory Imagery*, S. Lacey and R. Lawson, Eds., New York, NY: Springer, 2013, pp. 131–155. doi: 10.1007/978-1-4614-5879-1_8.

[22] S.-H. Yeh *et al.*, "AR-Classroom: Augmented Reality Technology for Learning 3D Spatial Transformations and Their Matrix Representation," in *2023 IEEE Frontiers in Education Conference (FIE)*, Oct. 2023, pp. 1–8. doi: 10.1109/FIE58773.2023.10343049.

[23] J. M. Clark and A. Paivio, "Dual coding theory and education," *Educ Psychol Rev*, vol. 3, no. 3, pp. 149–210, Sep. 1991, doi: 10.1007/BF01320076.

[24] A. Paivio, *Mental Representations: A Dual Coding Approach*. in Oxford Psychology Series. Oxford, New York: Oxford University Press, 1990.

[25] J. S. Devagiri, S. Paheding, Q. Niyaz, X. Yang, and S. Smith, "Augmented Reality and Artificial Intelligence in industry: Trends, tools, and future challenges," *Expert Systems with Applications*, vol. 207, p. 118002, Nov. 2022, doi: 10.1016/j.eswa.2022.118002.

[26] A. Dünser, K. Steinbügl, H. Kaufmann, and J. Glück, "Virtual and augmented reality as spatial ability training tools," in *Proceedings of the 7th ACM SIGCHI New Zealand chapter's international conference on Computer-human interaction: design centered HCI*, in CHINZ '06. New York, NY, USA: Association for Computing Machinery, Jul. 2006, pp. 125–132. doi: 10.1145/1152760.1152776.

[27] "GPT-4." Accessed: Oct. 09, 2024. [Online]. Available: https://openai.com/index/gpt-4-research/

[28] A. Saka *et al.*, "GPT models in construction industry: Opportunities, limitations, and a use case validation," *Developments in the Built Environment*, vol. 17, p. 100300, Mar. 2024, doi: 10.1016/j.dibe.2023.100300.

[29] T. Driessen, D. Dodou, P. Bazilinskyy, and J. de Winter, "Putting ChatGPT vision (GPT-4V) to the test: risk perception in traffic images," *Royal Society Open Science*, vol. 11, no. 5, p. 231676, May 2024, doi: 10.1098/rsos.231676.

[30] Z. C. K. Hawes, K. A. Gilligan-Lee, and K. S. Mix, "Effects of spatial training on mathematics performance: A meta-analysis," *Dev Psychol*, vol. 58, no. 1, pp. 112–137, Jan. 2022, doi: 10.1037/dev0001281.

[31] "GPT-4 with Vision: Complete Guide and Evaluation," Roboflow Blog. Accessed: Oct. 09, 2024. [Online]. Available: https://blog.roboflow.com/gpt-4-vision/

[32] W. Han, S. Xiang, C. Liu, R. Wang, and C. Feng, "SPARE3D: A Dataset for SPAtial REasoning on Three-View Line Drawings," Sep. 02, 2020, *arXiv*: arXiv:2003.14034. doi: 10.48550/arXiv.2003.14034.

[33] S. Cai, K. Bao, H. Guo, J. Zhang, J. Song, and B. Zheng, "GeoGPT4V: Towards Geometric Multi-modal Large Language Models with Geometric Image Generation," Jun. 17, 2024, *arXiv*: arXiv:2406.11503. doi: 10.48550/arXiv.2406.11503.

[34] P. Mooney, W. Cui, B. Guan, and L. Juhász, "Towards Understanding the Geospatial Skills of ChatGPT: Taking a Geographic Information Systems (GIS) Exam," in *Proceedings of the 6th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*, in GeoAI '23. New York, NY, USA: Association for Computing Machinery, Nov. 2023, pp. 85–94. doi: 10.1145/3615886.3627745.

[35] R. Tao and J. Xu, "Mapping with ChatGPT," *ISPRS International Journal of Geo-Information*, vol. 12, no. 7, Art. no. 7, Jul. 2023, doi: 10.3390/ijgi12070284.

[36] X. Wei, "Evaluating chatGPT-4 and chatGPT-4o: performance insights from NAEP mathematics problem solving," *Front. Educ.*, vol. 9, Sep. 2024, doi: 10.3389/feduc.2024.1452570.

[37] I. Mirzadeh, K. Alizadeh, H. Shahrokhi, O. Tuzel, S. Bengio, and M. Farajtabar, "GSM-Symbolic: Understanding the Limitations of Mathematical Reasoning in Large Language Models," Oct. 07, 2024, *arXiv*: arXiv:2410.05229. Accessed: Oct. 27, 2024. [Online]. Available: http://arxiv.org/abs/2410.05229

[38] S. Frieder *et al.*, "Mathematical Capabilities of ChatGPT," *Advances in Neural Information Processing Systems*, vol. 36, pp. 27699–27744, Dec. 2023.

[39] N. Rane, "Enhancing Mathematical Capabilities through ChatGPT and Similar Generative Artificial Intelligence: Roles and Challenges in Solving Mathematical Problems," Aug. 20, 2023, *Rochester, NY*: 4603237. doi: 10.2139/ssrn.4603237.

[40] N. Che Hashim, N. A. Abd Majid, H. Arshad, W. Khalid Obeidy, and T. Mäkilä, "User Satisfaction for an Augmented Reality Application to Support Productive Vocabulary Using Speech Recognition," *Adv. MultiMedia*, vol. 2018, Jan. 2018, doi: 10.1155/2018/9753979.

[41] K. Xia, X. Xie, H. Fan, and H. Liu, "An Intelligent Hybrid–Integrated System Using Speech Recognition and a 3D Display for Early Childhood Education," *Electronics*, vol. 10, no. 15, Art. no. 15, Jan. 2021, doi: 10.3390/electronics10151862.

[42] Z. Shaghaghian, H. Burte, D. Song, and W. Yan, "An augmented reality application and experiment for understanding and learning spatial transformation matrices," *Virtual Reality*, vol. 28, no. 1, p. 12, Jan. 2024, doi: 10.1007/s10055-023-00904-x.

[43] "Gemini vs GPT-4V : A Preliminary Comparison and Combination of Vision-Language Models Through Qualitative Cases," ar5iv. Accessed: Oct. 10, 2024. [Online]. Available: https://ar5iv.labs.arxiv.org/html/2312.15011

[44] S. A. Sorby, "Educational Research in Developing 3-D Spatial Skills for Engineering Students," *International Journal of Science Education*, vol. 31, no. 3, pp. 459–480, Feb. 2009, doi: 10.1080/09500690802595839.

Appendices

Some conversation examples from our experiments are included here.

A. GPT-4 Test Using Revised PSVT:R

Sample question from authors (June, 2024):

`[Revised PSVT:R Question 2 screenshot image upload to GPT-4]`

Sample answer from GPT-4, whose answer is B in the end (not necessarily to be correct):

```
To answer question 2, we need to visualize the rotation applied to the
object in the middle line and find the matching option from the bottom line.



**Step-by-Step Analysis:**

1. **Top Line**: Shows an initial object and its rotated version.

2. **Middle Line**: Shows another object labeled "As."

3. **Bottom Line**: Provides five possible rotated versions labeled A, B, C,
D, and E.



**Top Line Rotation:**

- The initial object in the top line is rotated such that the front face
(with the cutout) moves to the left, and the side face is on the top.



**Applying the Same Rotation:**

- The object labeled "As" in the middle line will undergo the same rotation.

- Visualize the rotation where the curved surface, currently facing forward,
will move to the left, and the flat surface will face upward.



**Matching the Rotated Object:**

- Option A does not match the rotation.

- **Option B** matches the rotation.
```
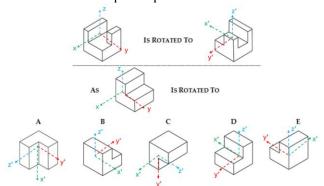

```
- Option C does not match the rotation.


- Option D does not match the rotation.


- Option E does not match the rotation.




Thus, the correct answer is **B**.
```

B. Test Using Revised PSVT:R with a coordinate system axes

Sample prompt by authors into GPT-4 with the uploaded image based on Revised PSVT:R with axes added (September, 2024):
The first line shows an object and how it is rotated. The task is to study how the object in the top line of the question is rotated. Picture in your mind what the object shown in the middle line of the question looks like when rotated in exactly the same manner. Select from among the five drawings (A, B, C, D, or E) given in the bottom line of the question the one that looks like the object rotated in the correct position. Objects in the Example: The top line shows the initial object and its rotated version. The middle line shows another object labeled "As". The bottom line provides five possible rotated versions labeled A, B, C, D, and E. To solve the example, you would visualize the rotation applied to the middle object ("As") and choose the correct rotated version from the options provided.

Sample answer from GPT-4:
To solve this problem, let's walk through the steps in detail by analyzing the rotation transformation applied to the top object and then applying it to the "As" object.
1. Analyze the Rotation in the Top Row: In the top row, we see that the initial object is rotated to achieve a new orientation. Observing the axes:
   • The x-axis (green) in the initial object is rotated to the z-axis position (upward).
   • The z-axis (blue) rotates to the x-axis position (forward).
   • The y-axis (red) stays in the same orientation.
2. Apply the Same Rotation to the "As" Object: Now, we apply this rotation to the middle object labeled "As":
   • Rotate the x-axis (green) of "As" to where the z-axis would point (upward).
   • Rotate the z-axis (blue) of "As" to the position of the x-axis (forward).
   • Keep the y-axis (red) in the same orientation.
3. Identify the Matching Option:

- Now, look at each option (A, B, C, D, E) in the bottom row to find the one that matches the rotated "As" object.
- Option C has the correct orientation of the axes:

The green x-axis points upward.

- The blue z-axis points forward.
- The red y-axis is unchanged.

Therefore, the correct answer is Option C.

C. Test of Step 1of Revised PSVT:R with a coordinate system

Sample prompt by authors into GPT-4 with the uploaded image based on Step 1 of a Revised PSVT:R question with axes added (September, 2024):
Explain the rotation process using minimum steps of rotation in terms of axis of rotation, angle of rotation, and direction of rotation.

is rotated to

Sample answer from GPT-4:
To describe the rotation process shown in this image with minimal steps, let's focus on the changes in axis orientation between the initial and rotated object.
Step-by-Step Rotation Process
  1. Identify the Rotation Needed:
- In the initial object, the x-axis (green) is pointing to the left, the y-axis (red) is pointing forward, and the z-axis (blue) is pointing upward.
- In the rotated object, the x'-axis (green) is pointing forward, the y'-axis (red) is pointing to the right, and the z'-axis (blue) remains pointing upward.

This indicates a rotation that affects the x and y axes while keeping the z-axis unchanged.
  2. Rotation Details:
- Axis of Rotation: Z-axis.
- Angle of Rotation: 90 degrees.
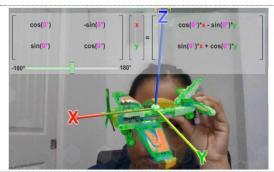- Direction of Rotation: Clockwise when viewed from above.

Explanation
By applying a 90-degree clockwise rotation around the z-axis, the x-axis moves from pointing left to pointing forward, and the y-axis moves from pointing forward to pointing right, while the z-axis remains in its original upward orientation.
This minimal rotation successfully transforms the initial object to match the rotated version shown.
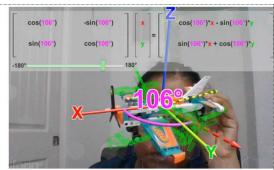
D. Test of AR Rotation with Coordinate, Angle of Rotation, Corresponding Matrix Equation

Sample prompt by authors into GPT-4 with the uploaded image based on AR-Classroom (October, 2024):
Explain the rotation process with axis of rotation, angle of rotation and direction. Note: All red lines represent X-axis, all yellow-green lines represent Y-axis, all blue lines represent Z-axis. The dotted line is the orientation of axis after rotation. The axis of rotation does not have a dotted counterpart of the same color. Follow right hand rule for the direction of rotation.

is rotated to

Sample Answer from GPT-4:

In this image, we see a rotation of the model from an initial position (left) to a rotated position (right) around the **Z-axis** (blue line) by an angle of **106°**.

Here's a detailed breakdown of the rotation process:

1. **Axis of Rotation**:

     The **Z-axis** (blue line) serves as the axis around which the model rotates. This axis remains fixed and does not have a dotted counterpart because it acts as the pivot for the rotation.

2. **Angle of Rotation**:

     The rotation angle is **106°**, as indicated in purple. This positive angle represents a counterclockwise rotation around the Z-axis.

3. **Direction of Rotation**:

     Using the **right-hand rule**, if you point your right thumb along the Z-axis, your fingers curl in the direction of the rotation. Since the angle is positive, the rotation proceeds counterclockwise when viewed from the positive direction of the Z-axis.

4. **Resulting Orientation**:

     After rotating 106° around the Z-axis, the **X-axis** (red line) and **Y-axis** (yellow-green line) shift orientation. The dotted red and yellow-green lines in the right image show the new positions of the X and Y axes, respectively, after the rotation. The X-axis rotates to a new position indicated by the dotted red line, and the Y-axis shifts to the dotted yellow-green line, showing how the model's alignment changes in 3D space.

In summary, this transformation represents a **106° rotation around the Z-axis**. The Z-axis remains as the fixed axis, while the X and Y axes change orientation, rotating counterclockwise according to the right-hand rule for a positive angle.