

A Virtual Reality Gearbox Experiment Platform Based on Hand Tracking and Natural Language Processing

Shuyue Hui, Yuanyuan Zhang *

University of Science and Technology Liaoning, Anshan Liaoning, 114001 China

* Corresponding author: Yuanyuan Zhang

Abstract: In manufacturing engineering education, the gearbox is an essential mechanical component and a fundamental skill students must master. Recent data suggests that insufficient understanding of its load capacity and exceeding operational limits can damage internal components, potentially leading to severe consequences like fires. In order to address the challenges faced by physical laboratories and existing traditional virtual labs in the new engineering disciplines, including the lack of insufficient experimental safety, operational realism, insufficient personalized learning support, and the difficulty in creating immersive learning environments, we introduce a Virtual Reality Gearbox Experiment Platform Based on Hand Tracking and Natural Language Processing (VRGEPHTNLP). This platform aims to enhance the realism of detailed interactions, provide personalized learning experiences, and create immersive environments, thereby improving students' learning efficiency and practical skills. In order to enhance the teaching of gearboxes, we have developed the VRGEPHTNLP by integrating hand-tracking technology, the GPT3.5 large language model (GPT3.5 LLM), and Azure technology. This application allows for comprehensive observation and assembly-disassembly teaching functionalities of parts. We offer an interaction experience rich in detail and environmental immersion, alongside personalized feedback and guidance in response to student inquiries. Compared to traditional computer-based virtual labs and conventional VR laboratories, our VRGEP-HTNLP exhibits significant advantages in improving user interaction realism, enhancing personalized learning support, and optimizing the immersion of the learning environment.

Keywords: Virtual Reality; Oculus Quest 2; Hand Tracking; Natural Language Processing; Large Language Models.

1. Introduction

In the new engineering disciplines, fostering students' practical abilities has always been a focus of instruction. Traditional physical laboratories face practical issues such as high equipment costs, dangerous operations, and experimental conditions limited by space and resources. Thus, addressing these challenges becomes essential. While traditional computer-based virtual labs have served as functional extensions of physical laboratories, reducing equipment costs and ensuring safety during experiments, we have identified their limitations in realistically simulating actual operations, providing personalized learning support, and creating immersive experiences, failing to meet students' basic experimental needs[1].

Traditional computer-based virtual labs mainly rely on mouse and keyboard interactions, where clicking and dragging input methods limit the true simulation of experimental operations, failing to replicate various operational details in experiments. This is especially true in experiments requiring precise operations, such as assembling gearbox components, where the input methods of computer-based platforms cannot simulate the detailed interaction experience of actual operations. As a result, students find it difficult to invoke non-visual learning methods, such as muscle memory, as learning aids through computer-based virtual experiment platforms, thus gaining effective operational experience becomes challenging.

Most traditional virtual experiment platforms have yet to integrate advanced natural language processing(NLP) and LLM, thus lacking the capability to respond to students' personalized learning needs[2, 3]. The uniform content and interaction options regarding experimental equipment provided by traditional platforms cannot offer immediate

feedback to different students upon encountering problems, failing to fully utilize the advantages that virtual experiment platforms can offer over physical laboratories.

The immersive experience offered by computer-based virtual labs is limited by the physical boundaries of the screen, unable to provide the immersive visual experience akin to that of physical laboratories. The lack of sensory elements that can be invoked reduces students' interest in learning and affects their efficiency in absorbing knowledge. In view of these challenges, we introduce a VRGEP-HTNLP to overcome the limitations of traditional physical and virtual labs. We built a virtual reality framework using the Unity engine[4, 5]. We incorporate hand-tracking technology[6]. This enables more realistic operations in detail. Additionally, we use the GPT-3.5 LLM[7]. We also utilize Azure technology[8]. Together, these provide a personalized question-and-answer mechanism. This includes functionalities like exploded view observations (EVO), close-up part inspections (CUPI), assembly and disassembly teaching (ADT), and an AI assistant (AIA). Compared to traditional computer-based virtual labs and existing VR labs, our platform shows significant advantages in enhancing the realism of detail operations, personalized learning support, and the immersion of the learning environment. Our work is summarized as follows:

- 1) We developed the VRGEP-HTNLP, providing a highly immersive and interactive learning experience, enhancing students' learning efficiency and practical skills.
- 2) Our project incorporates hand-tracking technology to capture and respond to students' precise operations during experiments.
- 3) We introduced GPT-3.5 LLM and Azure technologies, offering instant question-answering services.
- 4) While traditional virtual experiment platforms have contributed to solving issues related to high equipment costs,

spatiotemporal limitations of resources, and operational safety, our VR platform shows significant advantages over these issues in detail interaction experience, personalized learning support, and immersive environment, compared to traditional approaches.

2. Related Work

In recent years, with the rapid development of educational technology, virtual laboratories have become one of the effective ways to integrate theory and practice in higher education and vocational training. Particularly in specialized fields such as electromechanical engineering in vocational education, medical laboratory, lathe operation, and maintenance of electrical equipment in coal mines, the application of virtual laboratory technology has become increasingly widespread. The virtual laboratory technology addresses the issues of high equipment costs, high operational risks, and resource limitations by time and space associated with traditional physical laboratories. Although existing virtual experiment platforms have improved teaching quality and learning efficiency to some extent, they still face limitations in simulating detailed operations, providing personalized learning support, and enhancing the immersion of the learning environment.

Traditional computer-based virtual labs mainly rely on the mouse and keyboard as interaction tools. This mode of interaction, while simple and easy to operate, often fails to adequately simulate the complex movements and details involved in actual operations, lacking immersion and realism. For instance, the virtual experiment scheme in the medical field proposed by Xinyu Xiao[9] is limited in simulating detailed operations. Precise hand movements and operational feedback are hard to achieve through simple clicking and dragging, limiting students' understanding of experimental operation details and mastery of skills.

Furthermore, most existing virtual experiment platforms lack effective mechanisms for personalized learning support. For instance, the coal mine electromechanical equipment maintenance simulation system based on virtual reality technology by Long Meng[10] and the VR geography laboratory by A. Abd-Elrahman[11] typically offer static, uniform experimental operation guides and feedback. They cannot provide customized learning resources and immediate feedback according to specific student needs and learning progress. This "one-size-fits-all" approach overlooks the differences in knowledge acquisition, learning styles, and progress among students, thereby affecting the effectiveness of learning and student motivation. Although these platforms were developed using virtual reality technology, they all use controllers as the primary input method, failing to address the issue of detailed interaction.

Additionally, computer-based virtual labs, such as the vocational electromechanical engineering virtual experiment system platform proposed by He Yikui[12], despite offering an innovative teaching method, still have limitations in creating a fully immersive learning environment. Such platforms usually rely on traditional monitors for visual output, limiting users' visual perception to the screen size and resolution. This reliance lacks three-dimensional spatial awareness, thus making it difficult for students to achieve the spatial navigation and operation experience comparable to physical laboratories.

In addressing the challenges faced by traditional virtual reality setups, we introduce VRGEP-HTNLP, a model

leveraging the Unity engine and Oculus Quest 2, designed to transcend spatial constraints[13]. The integration of the Unity Interaction Toolkit(UIT) and the Oculus Integration SDK (OISDK) enables VEP-VR to create a highly interactive environment, simulating intricate experimental operations. Central to VEP-VR's auditory and linguistic functions is WaveNet, DeepMind's advanced deep neural network[14-17]. This technology generates realistic speech waveforms with dilated convolutions within its convolutional neural network (CNN) architecture, facilitating a customized question-and-answer system[18, 19]. This system transforms student questions from speech to text, processes this input, and delivers coherent, detail-rich responses instantaneously. VEP-VR employs neural network technology, including activation functions such as the rectified linear unit (ReLU) and the Softmax function for classification tasks[20-22]. These computational models are essential for VEP-VR's pattern learning and recognition, fostering complex interaction and feedback mechanisms. Word vector embeddings (WVEs) and their evaluation through Cosine Similarity are instrumental in analyzing the semantic content of student inputs, ensuring responses are both relevant and contextually fitting[23, 24]. Enhancing the interactive experience further, VEP-VR incorporates Do Tween for crafting interactive animations, such as gearbox manipulation[25]. This feature enables students to explore complex machinery from every perspective, promoting a comprehensive understanding of gearbox structures via precise hand-tracking gestures. This approach ensures VEP-VR not only overcomes the limitations of conventional VR laboratories but also enhances educational outcomes by providing an immersive, interactive learning experience.

3. Method

To address the issues of lack of operational realism, insufficient personalized learning support, and the absence of immersive learning environments in traditional computer-based virtual labs, we developed a VRGEP-HTNLP leveraging hand-tracking and NLP technologies. The platform facilitates EVO, CUPI, ADT, and AIA functionalities. By incorporating hand-tracking technology, the GPT-3.5 LLM, and Azure technology, we have achieved a realistic sense of detail operations, personalized learning support, and immersive learning environments. Furthermore, the development was carried out using the Oculus Quest 2 VR unit, based on the OISDK, employing two main entities, Interactor and Interactable, as the foundation for rich interactive settings. This distinction from traditional virtual experiment platforms adds an element of fun to learning. Utilizing the aforementioned technologies, our platform overcomes the existing issues of traditional virtual labs while fully leveraging the freedom and immersion offered by VR, thus creating a teaching aid platform with superior practical experience.

Our platform is divided into three main parts: the interaction module, the hand-tracking module, and the AIA that incorporates the GPT-3.5 LLM and NLP.

3.1. Hand Tracking

In the OISDK, the process for implementing hand-tracking begins with acquiring hand data tracked by the headset through the *FromOVRHandDataSource* component. This data is then passed to the *HandFilter* component, which implements the *IHand* interface and can perform operations

such as debouncing. In certain scenarios, restricting the virtual hand's movements during interactions becomes necessary, such as when pressing a button in VR. If pressed fully, the hand in reality cannot move further down, preventing the occurrence of clipping through objects. However, since the real hand can continue to move downwards, the VR hand's movement needs to be restricted to synchronize with the real hand's position. The *SyntheticHand* component further processes the data on hand joints, thereby restricting the virtual hand from moving further down. The processed hand data by *SyntheticHand* is rendered by the *HandVisual* component, displaying the posture of the virtual hand.

3.2. Interaction Module

The interaction module serves as the core of the VRGEP-HTNLP, encompassing the platform's primary functionalities: EVO, CUPI, and ADT.

(1)The menu interface module: Designing the main menu interface for selecting gearboxes, integrating hand-tracking as the trigger mechanism. Setting the palm-up position of the left hand as the menu activation condition and implementing swipe gestures to enable switching between gearbox types.

(2)The exploded view module(EVM): Designing an EVM for gearbox components to clearly show the structural relationship between parts, aiding students in comprehensively understanding the gearbox's construction and function. See Figure 1 for the design flowchart.

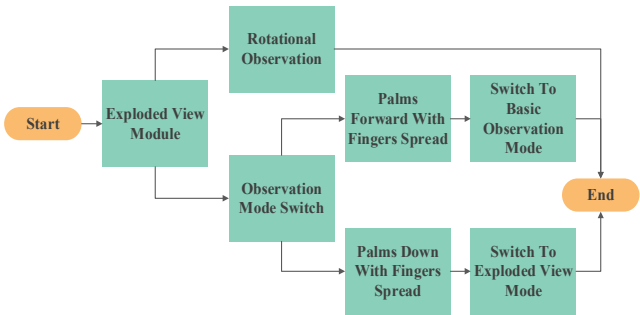


Fig 1. EVM Flowchart

We establish a logical parent-child hierarchy and use the *MoveObjectWithArrows* script to control the real-time transformations of the model's local coordinates through coding, thus achieving the effect of rotating the gearbox on a turntable for omnidirectional observation. Additionally, we incorporate OISDK-provided scripts for hand-tracking, close-up detail grabbing, and event triggering as the foundational methods for interaction within the platform.

The pseudocode for the *MoveObjectWithArrows* script is as follows table 1.

(3) The Close-up Observation Module (COM): This Module allows students to meticulously inspect each part individually, with a design flowchart available in Figure 2. We aim to provide a more intuitive and engaging mode for students to deeply understand the gearbox's construction. This part incorporates OISDK-provided scripts for remote grabbing and two-handed grabbing as the basis for interaction methods.

Table 1. Algorithm 1 *MoveObjectWithArrows*

Input:
gameObjects: array of objects to be manipulated in the scene;

Initialize:
1: Record the initial positions of each gameObject in the scene;

ExplodeView:
2: **for** each gameObject in gameObjects **do**
3: Displace gameObject from its initial position along the x-axis;
4: **end for**

ResetView:
5: **for** each gameObject in gameObjects **do**
6: Move gameObject back to its recorded initial position;
7: **end for**

end Algorithm

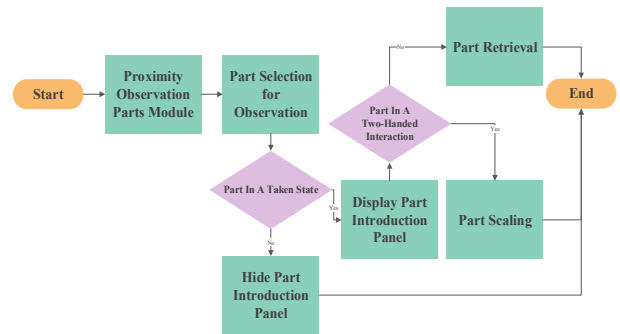


Fig 2. COM Flowchart

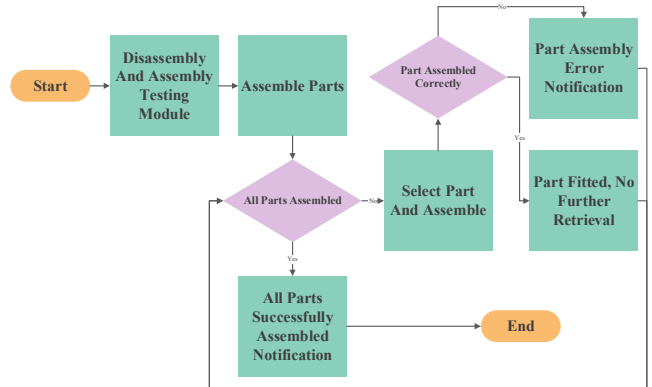


Fig 3. ADTM Flowchart.

(4) The Assembly and Disassembly Teaching Module (ADTM): ADTM offers students immersive, first-person perspective teaching on the sequence of part assembly and disassembly through interactions like one-handed grabbing and raycasting. We designed this module to help students master correct part assembly skills based on a hands-on feeling, using OISDK-provided UI raycasting scripts for interaction. Through the Toggle Control script, we display a selected part on the pickup station. By assigning grabbable attributes to parts, we designed the Parts Disassemble script to show phantom cues of the next part and check if the part is correctly placed. ADTM also provides a testing platform for students on the sequence of gearbox part assembly and disassembly, with a design flowchart seen in Figure 3. Building on the assembly teaching, this feature adds collision detection around the part phantom cues and designs the Part Order Detector script, which checks if the current assembly

sequence is correct while displaying corresponding hint panels.

The pseudocode for the PartOrderDetector script is as follows:

Table 2. Algorithm 2 PartOrderDetection

Input:
 partsOrder: sequence of parts to be assembled;
 feedbackPanel: UI for displaying assembly feedback;

Initialize:
 1: Record each part's original parent position;

PartInteraction:
 2: **if** the part is next in assembly order **then**
 3: Attach part to the correct position;
 4: Update feedbackPanel to indicate success;
 5: **else**
 6: Reset part to original position;
 7: Update feedbackPanel to show error;

FeedbackDisplay:
 8: Fade out the feedbackPanel after a set duration;

end Algorithm

Additionally, we primarily utilize the input data and interaction settings provided by the OISDK technology.

Input Data:

In the context of virtual reality projects, the primary input methods for users are divided into controller use and hand-tracking. When using a controller, the virtual environment's hands synchronize with the controller's position and rotation in the real world. However, with hand-tracking, the virtual hands align with the real-world hand posture.

The OISDK employs a series of *from OVR...Source* components to acquire tracking data. This data is then sent to various processing interfaces, namely *IController*, *IHand*, *IHmd*, and *IBody*, for further processing and packaging, thus fitting the components within the OISDK. Consequently, this data can be utilized by these components to execute corresponding interactive functions, serving as the foundation for diverse interactions. This method of hand-tracking input offers students a natural interaction feel and detailed operation detection that traditional computer-based virtual labs, with their click and drag interactions, cannot achieve.

Interactor And Interactable:

The OISDK sets up various interactions through two main objects: the Interactor and the Interactable. The Interactor initiates interactions, typically representing the user's virtual hands or controllers, whereas the Interactable is an object that can respond to interactions, such as various objects in the virtual world. The Interactor component is usually attached to the virtual object representing the user's hands or controllers. Once the system acquires position and rotation information about the hands or controllers through device tracking, the Interactor starts searching for Interactable objects nearby.

During interaction, the Interactor and Interactable objects go through several different states:

Normal: This is the default state with no interaction taking place.

Hover: The Interactor enters the Hover state when approaching an Interactable object under specific conditions.

Select: The Interactor transitions from Hover to Select state upon completing an interaction, indicating the completion of the interaction.

Disabled: In this state, both the Interactor and Interactable cannot interact.

The transition between these states facilitates a complete interaction process, allowing users to interact with the virtual world in a manner consistent with real-life habits. Also, based on these two objects, we have expanded the functionalities of the VRGEP-HTNLP, implementing gamified interaction designs.

3.3. GPT-3.5

In the development of the VRGEP-HTNLP, LLM technology plays a pivotal role, with WVEs being one of its foundational elements. WVEs transform vocabulary into mathematical vectors, quantifying and conveying the semantic content of words. The similarity between these vectors, measured by Cosine Similarity, reveals the semantic closeness between words. Cosine Similarity calculates the cosine of the angle between two non-zero vectors in the inner product space, serving as an effective method to assess the similarity in vector direction. The formula is as follows:

$$\text{CosineSimilarity}(A, B) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (1)$$

Where, A and B represent the vector embeddings of two words, respectively. This calculation method aids in identifying the semantic similarity between two words.

The platform also incorporates technologies based on neural networks, which are complex computational systems that mimic the structure of the human brain for learning and pattern recognition. These networks consist of numerous neurons organized in layers and connected by weights. The activation functions in neural networks define the relationship between a node's output and the node's input, playing a crucial role in nonlinear mapping. For instance, the rectified linear unit (ReLU) function is a commonly used activation function. The formula is as follows:

$$\text{ReLU}(x) = \max(0, x) \quad (2)$$

The ReLU function remains unchanged when the input is greater than 0, and outputs 0 otherwise. This characteristic of the ReLU function makes it very effective in reducing the problem of gradient vanishing and speeding up the network training process.

For NLP and classification tasks, the output layer of neural networks often uses the Softmax function, which converts the network's raw output into a probability distribution. The Softmax function is particularly suitable for multi-class classification problems, as the function can calculate the probability of each category. The formula is as follows:

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}} \quad (3)$$

Where, z_i represents the network's raw output for the i th class, and the denominator is the sum of the exponentials of the outputs for all classes. Through the Softmax function, outputs can be interpreted as a probability distribution, thereby enabling precise categorization and interpretation of model predictions in NLP tasks. The integration of these technologies provides the VRGEP-HTNLP with powerful language understanding and interaction capabilities, forming the theoretical foundation for implementing personalized teaching modules.

3.4. The Natural Language Processing module

WaveNet is a deep neural network architecture introduced by DeepMind, specifically designed for generating high-quality speech waveforms. WaveNet employs a special type of CNN that directly generates the probability distribution of sound waves by learning patterns in audio data. The key innovation of the WaveNet model lies in the model's use of dilated convolutions, which enable the network to efficiently process long-range sequence data without sacrificing model complexity or increasing computational costs. The formula is as follows:

$$p(\mathbf{x}_t | \mathbf{x}_{1:t-1}) = \text{WaveNet}(\mathbf{x}_{1:t-1}) \quad (4)$$

In the WaveNet model, the generation of each audio sample depends on the history of all previous samples. The model's output, $p(\mathbf{x}_t | \mathbf{x}_{1:t-1})$, represents the probability distribution of the audio sample \mathbf{x}_t at the current time point t , given the history of all previous audio samples $\mathbf{x}_{1:t-1}$. In this way, WaveNet can construct the entire audio waveform step by step, with each step based on the output of the previous steps.

The VRGEP-HTNLP implements an AI voice assistant feature by converting voice signals into text, a process achieved through speech recognition using deep learning models. Speech recognition technology based on deep learning primarily predicts text sequences by maximizing the conditional probability of given audio feature sequences, with its mathematical expression being:

$$p(Y | X) = \prod_{i=1}^T p(y_i | y_{1:i-1}, X) \quad (5)$$

Where, X represents the input audio feature sequence, while Y is the corresponding output text sequence, where y_i refers to the i th character in the sequence. This approach allows the model to consider not only the current audio features but also enhance the prediction accuracy of subsequent characters based on previously predicted ones. Through this sequence dependency, deep learning models can effectively transform continuous speech signals into corresponding text expressions, thereby providing robust voice recognition support for user interactions on the VRGEP-HTNLP.

4. Platform Design

4.1. Platform Construction

The platform is built on Unity to create a high-definition, realistic virtual simulation environment. We start by creating a 3D project and incorporating the Oculus-provided developer toolkit, OISDK. We also introduce the UIT and perform basic parameter settings to establish a VR project based on the Oculus Quest 2 device. The gearbox model is exported as an fbx file from 3ds Max and imported into Unity. Finally, we construct the overall background of the laboratory and perform scene lighting baking.

4.2. Implementation of the Interaction Module

Implementation of the Exploded View Module.

In this module, by making targeted hand-tracking gestures, operations for exploding and restoring the gearbox can be executed. We also set up a turntable based on detailed grabbing interactions, establish the turntable's pivot, and

combine the turntable with scripts based on DoTween to support direct hand manipulation of the turntable's rotation, enabling students to observe the part structure from all angles.

This functionality uses *Grabbable* and *OneGrabRotateTransformer* scripts to define the turntable's rotation center rotate pivot and set the orientation of its rotation axis, achieving a rotatable interaction effect for the turntable. We design the *MoveObjectWithArrows* script, define explosion and restoration functions, and integrate the DoTween plugin for smooth movement of parts after explosion. The project creates two target gestures: stop and paper, the former for triggering the switch from exploded view mode to normal observation mode, and the latter for triggering the switch from normal observation mode to exploded view mode. The created gesture files are then passed to the *ShapeRecognizer*, and finally, the *TargetGestureTriggeringEvents* are implemented in combination with the *ActiveStateSelector* script and *SelectorUnityEventWrapper*. The actual effect of this functionality is shown in Figure 4.

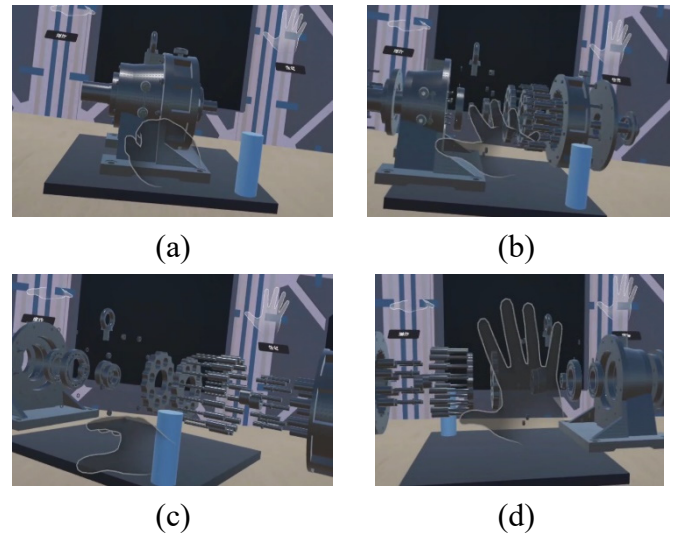


Fig 4. EVM Process. (a) represents the original state, (b) shows the gesture-triggered explosion event, (c) displays the rotation effect of the turntable and (d) illustrates the gesture-triggered restoration event.

Implementation of the Close-up Observation Module.

In the parts wall area, students select target parts for observation, supporting remote grabbing of parts. The COM allows students to closely grab parts with both hands and to scale and rotate them within a certain range, thereby achieving a clearer observation of the parts. During the grabbing process, an information panel about the part is displayed, enabling students to instantly learn and understand relevant knowledge about the part.

This functionality utilizes the *TwoGrabFreeTransformer* script to enable dual-hand interactivity of parts, providing internal scaling capabilities with appropriate scaling limit values set. Based on close-range grabbing interaction, a long-distance grabbing feature for each part is established using the *DistanceHandGrabInteractable* and *MoveTowardsTargetProvider* scripts, allowing students to select and grab parts for observation from a distance. Event trigger scripts *PointableUnityEventWrapper* are set for both close-range and long-distance grabbing interactions of parts, enabling the immediate display of the corresponding part's introduction panel when students grab parts for observation. The actual

effect of this functionality is shown in Figure 5.

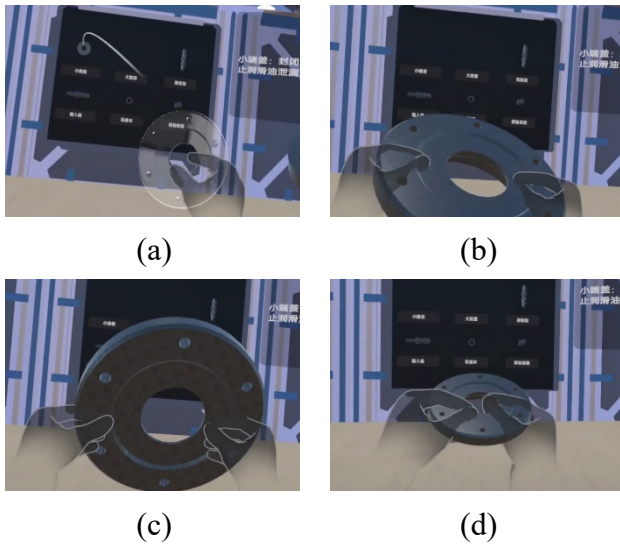


Fig 5. COM Process. (a) displays part selection,(b) shows the basic effect of picking up and observing a part,(c) demonstrates the effect of enlarging a part for observation with both hands and (d) illustrates the effect of reducing a part for observation with both hands.

Implementation of the Assembly and Disassembly Teaching Module.

In this module, students can fully observe and assemble the gearbox model by manually rotating the assembly stand. The

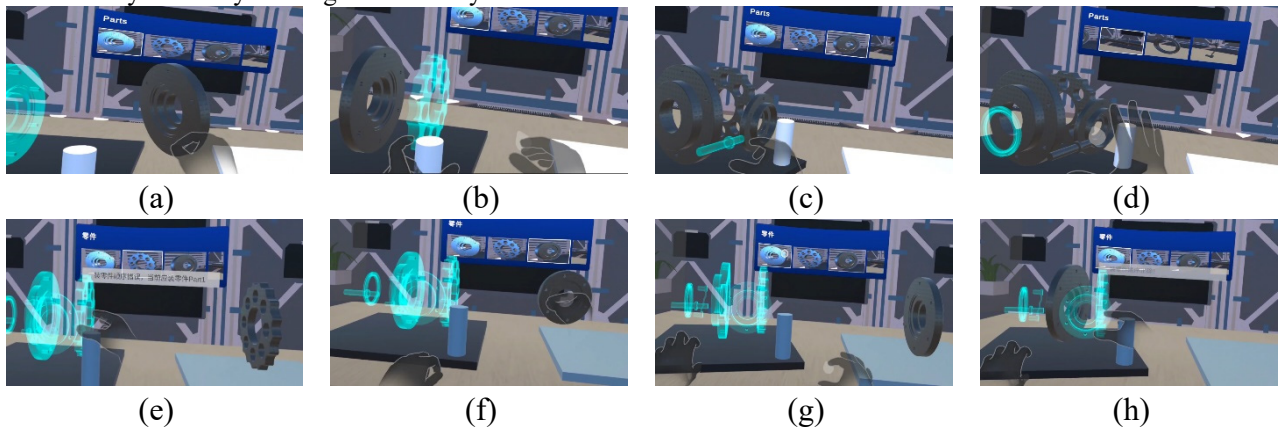


Fig 6. ADTM Process. (a) displays part selection, (b) shows the turntable rotation effect, (c) displays selecting the part assembly position,(d) successful part assembly effect,(e) shows the prompt panel when incorrectly assembling parts,(f) displays the process of picking up parts during disassembly testing,(g) represents the process of assembling parts and (h) represents the prompt panel when parts are correctly assembled.

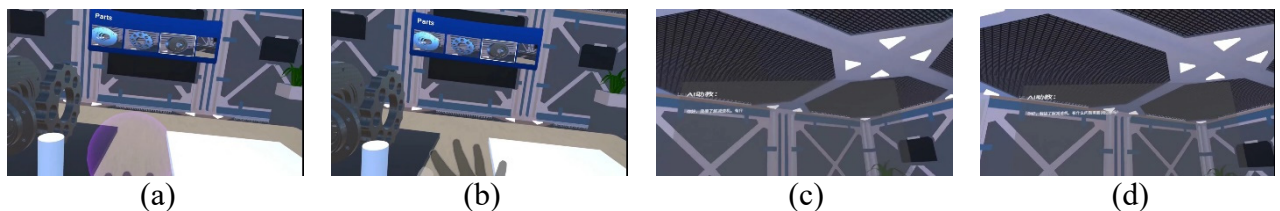


Fig 7. AIM Process. (a) shows the gesture-triggered start of questioning,(b) displays the gesture-triggered end of questioning,(c) demonstrates the AIA's response process and (d) presents the AIA's voice and text feedback

4.3. GPT-3.5 and Natural Language Processing Module

Implementation of the AI Assistant Module(AIM).

This module is integrated within the ADTM, COM and EVM . When the user's right palm faces up, the question-

assembly stand sequentially displays a preview of the next part to be assembled. Students select the currently needed part through hand-ray interaction with the part selection panel in front and grab the part from the pickup station. If students assemble the wrong part or in the incorrect order, the part will return to its original position, serving as a reminder to reassemble correctly.

We utilize the *CanvasRenderTexture* script to generate the Canvas Mesh for the part selection panel, and scripts *PointableCanvas* and *PointableCanvasMesh* are applied to both Canvas and Canvas Mesh to enable interactivity of the part selection panel. The *RayInteractable* script sets the part selection panel's interaction mode to ray interactable. In the ADTM scene, the assembly stand displays all preview images of gearbox parts. Students must think carefully, use hand-ray interaction to select the correct part from the part selection panel, and move the part to the designated position. When a part is detected within the specified range, the part locks into place and cannot be taken again, considered as successfully installed. If users assemble parts in the wrong order, the part automatically returns to the pickup station, and a prompt box pops up, indicating the incorrect sequence and which part should be selected next. If the assembly order is correct, a prompt box appears, confirming the correct sequence. Upon successful assembly of all parts in the correct order, a congratulatory prompt box appears, serving as an encouragement mechanism. The actual effect of this feature is shown in Figure 6.

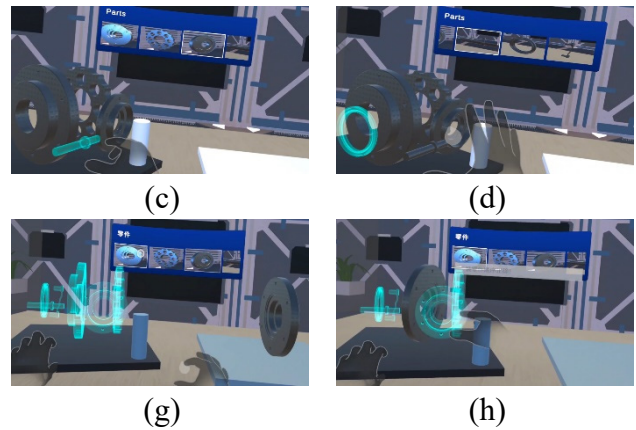


Fig 6. ADTM Process. (a) displays part selection, (b) shows the turntable rotation effect, (c) displays selecting the part assembly position,(d) successful part assembly effect,(e) shows the prompt panel when incorrectly assembling parts,(f) displays the process of picking up parts during disassembly testing,(g) represents the process of assembling parts and (h) represents the prompt panel when parts are correctly assembled.

asking mechanism is activated, displaying a purple semi-transparent sphere above the right palm, accompanied by an auditory cue signaling the user can begin asking questions. To conclude the inquiry, the user simply turns the right palm down, causing the purple sphere and sound to disappear, and the user's voice message is sent to Azure for speech

recognition, processed into text, and forwarded to GPT-3.5 LLM. GPT-3.5 LLM then returns the answer in text to Azure's speech synthesis, which conveys the answer to the student in audio form. The design of the AIA enables students to solve problems immediately and specifically.

The functionality employs the Speech Recognizer script using Azure's speech recognition to convert student voices into text form. The ChatScript and GptTurboScript scripts are utilized for establishing connections and managing the basic logic for information transmission. The AzureSpeech script converts text into speech. The actual effect of this feature is illustrated in Figure 7.

5. Conclusion

We introduce an innovative Virtual Reality Gearbox Experiment Platform Based on Hand Tracking and Natural Language Processing. This platform centers on simulating detailed operations and crafting personalized teaching mechanisms. Our main aim is to provide an auxiliary tool for engineering practice that enhances the accuracy of detailed operation simulation and personalizes the teaching process. Distinguishing our platform from existing virtual labs, we highlight several key advantages. First, integrating hand-tracking technology through OISDK, we enable detailed operational simulations for student experiments, closely mirroring real-life practices. Second, combining GPT-3.5 LLM with Azure's NLP, we craft an AIM for personalized learning support, ready to instantly respond to student inquiries verbally. Finally, utilizing Oculus Quest 2 VR's device immersive capabilities, we create a comprehensive learning environment, using visual effects to spike student interest and learning outcomes. Our platform addresses the gaps in traditional virtual labs related to detailed operation simulation, personalized teaching support, and immersive environment creation. Furthermore, the platform harnesses VR technology's benefits, offering a new, efficient tool for practical teaching modes and forging paths for future educational practice development.

Acknowledgments

This work was supported in part by innovation and entrepreneurship training program for college students of university of science and Technology Liaoning 2024.

References

- [1] K. Frady, "Use of virtual labs to support demand-oriented engineering pedagogy in engineering technology and vocational education training programmes: a systematic review of the literature," *European Journal of Engineering Education*, vol. 48, no. 5, pp. 822-841, 2023.
- [2] G. L. He, C. Y. Chi, and Y. Y. Zhan, "Combining N-gram Statistical Model with Pre-trained Model to Correct Chinese Sentence Error," *Engineering Letters*, vol. 30, no. 2, pp. 476-484, 2022.
- [3] Z. Li, J. Wu, J. Miao, and X. Yu, "Improve the Response Diversity of Multi-turn Dialogue System by Combining Knowledge," *IAENG International Journal of Computer Science*, vol. 49, no. 3, 2022.
- [4] M. H. Qazi, F. Khan, J. Kim, and E. J. Rojas-Munoz, "Developing a VR-based Training Platform for Emergency Fire Handling Services Using Unity 3D," in *20th International Conference on Frontiers of Information Technology, FIT 2023*, December 11, 2023 - December 12, 2023, Islamabad, Pakistan, 2023: Institute of Electrical and Electronics Engineers Inc., pp. 102-107.
- [5] N. Rane, S. Choudhary, and J. Rane, "Enhanced product design and development using Artificial Intelligence (AI), Virtual Reality (VR), Augmented Reality (AR), 4D/5D/6D Printing, Internet of Things (IoT), and blockchain: A review," *Virtual Reality (VR), Augmented Reality (AR) D*, vol. 4, 2023.
- [6] Y. Wang, Y. Wu, S. Jung, S. Hoermann, S. Yao, and R. W. Lindeman, "Enlarging the Usable Hand Tracking Area by Using Multiple Leap Motion Controllers in VR," *IEEE Sensors Journal*, vol. 21, no. 16, pp. 17947-17961, 2021.
- [7] T. Ivanova, A. Staneva, D. Borissova, and K. Rasheva-Yordanova, "Chat GPT Performance Evaluation Model for Learning," in *15th International Conference on Interactive Mobile Communication, Technologies, and Learning, IMCL 2023*, November 9, 2023 - November 10, 2023, Thessaloniki, Greece, 2024, vol. 936 LNNS: Springer Science and Business Media Deutschland GmbH, pp. 149-157.
- [8] S. Singh, K. R. Ramkumar, and A. Kukkar, "Analysis and Implementation of Microsoft Azure Machine Learning Studio Services with Respect to Machine Learning Algorithms," in *International Conference on Modern Electronics Devices and Communication Systems, MEDCOM 2021*, October 29, 2021 - October 31, 2021, Noida, India, 2023, vol. 948: Springer Science and Business Media Deutschland GmbH, pp. 91-106.
- [9] X. Xiao, X. Liu, and Z. Xiao, "Construction and application of computer virtual simulation teaching platform for medical testing," in *2021 International Conference on Information Systems and Computer Engineering, ISCE 2021*, March 29, 2021 - March 30, 2021, Qingdao, China, 2021, vol. 1915: IOP Publishing Ltd.
- [10] J. Ma et al., "Using immersive virtual reality technology to enhance nursing education: A comparative pilot study to understand efficacy and effectiveness," *Applied Ergonomics*, vol. 115, 2024.
- [11] A. Abd-Elrahman, C. Atwill, A. Gonzalez, K. Hernandez, D. Barry, and Z. Abd-Elrahman, "Virtual Reality (VR) Assisted Geomatics lab Development," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 48, pp. 933-938, 2023.
- [12] J. Wang and W. Zhu, "Design and Implementation of Virtual Animation Based on Unity3D," in *2nd International Seminar on Artificial Intelligence, Networking and Information Technology, AINIT 2021*, October 15, 2021 - October 17, 2021, Virtual, Shanghai, China, 2021: Institute of Electrical and Electronics Engineers Inc., pp. 667-669.
- [13] S. Ho and U. Karabiyik, "Forensics Analysis of Virtual Reality Social Community Applications on Oculus Quest 2," in *6th IFIP International Conference on Internet of Things, IFIP IoT 2023*, November 2, 2023 - November 3, 2023, Denton, TX, United states, 2024, vol. 683 AICT: Springer Science and Business Media Deutschland GmbH, pp. 337-352.
- [14] Aripin, W. Agastya, and S. Huda, "Multichannel Convolutional Neural Network Model to Improve Compound Emotional Text Classification Performance," *IAENG International Journal of Computer Science*, vol. 50, no. 3, 2023.
- [15] N. Mazhar, A. Gaballa, A. K. Pandey, and J.-J. Cabibihan, "Virtual Reality Hand Tracking for Immersive Telepresence in Rehabilitative Serious Gaming," in *15th International Conference on Social Robotics, ICSR 2023*, December 3, 2023 - December 7, 2023, Doha, Qatar, 2024, vol. 14453 LNAI: Springer Science and Business Media Deutschland GmbH, pp. 3-12.
- [16] H. Park, D. Lichtman, and O. Oyekoya, "Exploring Virtual Reality Game Development as an Interactive Art Medium: A Case Study with the Community Game Development Toolkit," in *2023 Conference on Interactive Surfaces and Spaces, ISS*

- 2023, November 5, 2023 - November 8, 2023, Pittsburgh, PA, United states, 2023: Association for Computing Machinery, Inc, pp. 5-9.
- [17] J. Qian, Y. Zhang, D. Zhao, X. Zhang, Y. Xu, and Z. Tao, "Investigation of Vowel Generation Method in Low-resource Pathological Voice Database," *Engineering Letters*, vol. 31, no. 1, pp. 399-412, 2023.
- [18] T. Liu, Y. Yuan, G. Teng, and X. Meng, "Improved Deep Convolutional Neural Network-Based Method for Detecting Winter Jujube Fruit in Orchards," *Engineering Letters*, vol. 32, no. 3, pp. 569-578, 2024.
- [19] X. Zhang and Y. Tian, "Improved YOLOv5s Traffic Sign Detection," *Engineering Letters*, vol. 31, no. 4, pp. 1883-1893, 2023.
- [20] O.-S. Berot et al., "Choice of Parameters of an LSTM Network, based on a Small Experimental Dataset," *Engineering Letters*, vol. 32, no. 1, pp. 59-71, 2024.
- [21] Erwin, H. K. Putra, B. Suprihatin, and F. Ramadhini, "A Hybrid CLAHE-GAMMA Adjustment and Densely Connected U-NET for Retinal Blood Vessel Segmentation using Augmentation Data," *Engineering Letters*, vol. 30, no. 2, pp. 485-493, 2022.
- [22] Z. Hu, L. He, and H. Wu, "A Multi-feature Fusion Transformer Neural Network for Motor Imagery EEG Signal Classification," *Engineering Letters*, vol. 31, no. 4, pp. 1822-1831, 2023.
- [23] N. Khotimah and A. S. Girsang, "Indonesian News Articles Summarization Using Genetic Algorithm," *Engineering Letters*, vol. 30, no. 1, pp. 152-160, 2022.
- [24] Venkatesh, S. U. Hegde, S. B. Basapur, and Y. Nagaraju, "DistilBERT-CNN-LSTM Model with GloVe for Sentiment Analysis on Football Specific Tweets," *IAENG International Journal of Computer Science*, vol. 49, no. 2, 2022.
- [25] H. Yikui, "Design and Implementation of Virtual Experiment System Platform for Electrical and Electronic Engineering in Vocational College," in *2021 2nd International Conference on Artificial Intelligence and Information Systems, ICAIIS 2021*, May 28, 2021 - May 30, 2021, Chongqing, Virtual, China, 2021, vol. 1982: IOP Publishing Ltd.