WORK IN PROGRESS

Feel the Connection: Haptic Enhanced Interaction with an AI Agent

SOHEIL KIANZAD, Huawei Technologies Co., Ltd., Shenzhen, Guangdong, China

YINAN LI, School of Computing and Augmented Intelligence, Tempe, AZ, United States

HASTI SEIFI, School of Computing and Augmented Intelligence, Tempe, AZ, United States

Open Access Support provided by:

Huawei Technologies Co., Ltd.

School of Computing and Augmented Intelligence

Published: 26 April 2025

Citation in BibTeX format

CHI EA '25: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems
April 26 - May 1, 2025
Yokohama, Japan

Conference Sponsors:
SIGCHI

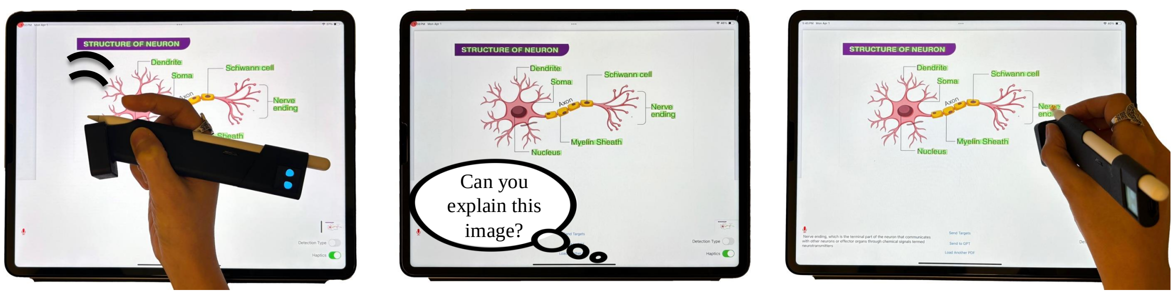
CHI EA '25: Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (April 2025)
https://doi.org/10.1145/3706599.3720173
ISBN: 9798400713958

Feel the Connection: Haptic Enhanced Interaction with an AI Agent

Soheil Kianzad
Huawei
Toronto, Ontario, Canada
soheil.kianzad@huawei.com

Yinan Li
School of Computing and Augmented Intelligence
Arizona State University
Tempe, Arizona, USA
yinanli2@asu.edu

Hasti Seifi
School of Computing and Augmented Intelligence
Arizona State University
Tempe, Arizona, USA
hasti.seifi@asu.edu



(a) (b) (c)

Figure 1: Example interaction with SenseiPen: (a) A user starts a voice interaction by double tapping the pen, (b) the system captures and processes verbal prompts with image and text coordinates via its vision system, sending them to a large language model. Next, (c) the platform's synthesizer crafts a multimodal response, combining verbal explanations with force feedback and visual cues.

Abstract

Advances in large language models (LLMs) offer new possibilities for multimodal information retrieval. Yet, the concept of transforming a language model into a haptic agent remains underexplored. We introduce a novel platform -SenseiPen- that enriches LLM interactions via voice, pen input, and haptic force feedback. SenseiPen is a handheld autonomous haptic agent that takes user queries via voice and pen input on a tablet, understands the input using GPT-4V (vision-preview), and responds with voice, visual cues, and force feedback. We investigated SenseiPen's performance and user experience in a lab-based study with 16 users. The study examined the efficacy of force feedback in guiding users toward target points and engaging users in question-answering with images and diagrams. Our research demonstrates that integrating pen-based sketching and force feedback with LLMs can improve user engagement and reward and foster intuitive physical interactions with an AI agent.

CHI EA '25, Yokohama, Japan

ACM ISBN 979-8-4007-1395-8/25/04
https://doi.org/10.1145/3706599.3720173

CCS Concepts

• **Human-centered computing → Human computer interaction (HCI)**; *Haptic devices*; User studies.

Keywords

Pen-based UIs, Force Feedback, Haptics, Large Language Models

ACM Reference Format:
Soheil Kianzad, Yinan Li, and Hasti Seifi. 2025. Feel the Connection: Haptic Enhanced Interaction with an AI Agent. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3706599.3720173

1 Introduction

Large Language Models (LLMs) offer new possibilities for multimodal interactions in various applications. These applications leverage and extend the natural language understanding and reasoning capabilities of LLMs to support users in various real-world scenarios such as creative writing [48] and programming [23], creating or editing images and animations [8, 13], playing games [9], and seeking mental health support from conversational agents [25]. Multimodal image and audio-based input and output capabilities of LLMs such as ChatGPT-4 [31] and Google Gemini [11] have marked significant steps towards human-centered computing.

Despite advancements in LLM-based interactions, a gap exists in leveraging LLMs for spatial and physical information exchanges.

Existing interfaces primarily focus on text-based interactions, with some using visual or audio modalities, but they cannot provide physical feedback. Haptic devices can enhance user task performance and enjoyment by guiding the user's hand (e.g., while drawing) or simulating real-world physical interactions [19, 29]. Yet, existing haptic devices need to be programmed for a given task, which limits their application for ad-hoc information-seeking scenarios. Integrating haptic devices and LLMs can bridge this gap to enable physical interactions with images and diagrams for tasks requiring spatial understanding and manipulation. This integration of haptic feedback can make digital interactions more intuitive and enhance the utility of LLMs in visual and spatial interaction tasks.

To bridge this gap, we present SenseiPen, an LLM-based handheld haptic agent (i.e., AI agent) that provides new pen-based spatial and interactive possibilities by integrating verbal, visual, and haptic modalities (Figure 1). SenseiPen's hardware is a handheld robotic device with a small footprint, designed to augment Apple Pencil and iPad with haptic sensing (orientation, acceleration) and output (force, vibration). It takes the user's verbal input, images and sketches, handwriting, and touch gestures, understands the user's intention and command by connecting to GPT-4V (vision preview), and produces force feedback and visual hints synchronized with a verbal response. It can make visual diagrams interactive and simulate the feel of physical phenomena (e.g., gravity forces). Thereby, SenseiPen introduces a tangible layer to communications with AI, effectively blending the virtual with the physical.

We ran a user study with 16 participants to assess SenseiPen's user experience. We evaluated the system's performance in guiding the user's hand to various locations on a tablet and investigated how users perceived and interacted with the LLM-based haptic pen and what challenges they faced in using it. Our results showed the capability of the system to dynamically generate movement and force feedback based on user queries and suggested that haptic cues can enhance user engagement during question-answering interactions with images. We discuss the implications of our work for haptics research and outline the remaining challenges for our ongoing and future work. Our contributions are:

- SenseiPen, a 2D haptic device and software system that integrates force feedback technology with an LLM to support multimodal interactions with digital pens and tablets.
- Data on the device's position accuracy and speed in reaching targets on a tablet when held by 16 users.
- Results on the impact of force feedback on user engagement with images, demonstrating the potential of haptics to enhance the interaction quality with LLMs.

2 Related Work

Interfaces for Large Language Models. Interacting with computers through natural language has been a focus of research for decades, with early systems like Put-that-there [3] and Quickset [6] enabling users to issue voice commands combined with pen and gesture inputs. More recently, the surge in LLMs has inspired the development of graphical interfaces to enhance user interactions. For example, Graphologue generates real-time interactive graph visualizations from LLM outputs for question-answering [15], and VISAR integrates a text editor with an interactive canvas to support creative writing [48]. These interfaces and visualizations augment LLM outputs, improving user sense-making and creativity. Similarly, systems like DirectGPT combine LLMs with direct manipulation principles to enable text and image editing [27], while tools like Instruct-nerf2nerf [13] allow users to modify 3D scenes using text prompts. For interactive behavior generation, LLMR uses GPT-4 to plan and add animations and interactivity to virtual reality scenes [8]. Beyond visualization, researchers have extended LLMs to create autonomous behavior. Generative agents use LLMs to observe environments and plan actions [33], while others program embodied agents, such as robots, where LLMs translate verbal suggestions into expressive robot behaviors [26]. Building on this, our research focuses on generating haptic behaviors synchronized with other modalities to enhance user engagement and sense-making.

Pen-based User Interfaces. Pen-based systems have long supported sketching, designing, and sensemaking. Early innovations like SketchPad [40] pioneered graphical human-machine interaction through a light pen. Subsequent interfaces, such as Sim-U-Sketch [18] and VibroSketch [17], helped users learn circuits and vibrations by recognizing hand-drawn diagrams, while tools like MathPad2 [21] and Hands-on Math [44] recognized handwritten formulas to create animations or solve equations. Recent work has incorporated AI to enhance pen functionality and user experience. Augmented Math [5] uses optical character recognition (OCR) and computer vision with a digital pen to transform static math into dynamic, interactive documents, while SMath [49] leverages recurrent neural networks (RNNs) for intelligent, pen-centric interface for manipulating mathematical expressions. Many pen-based systems combine sketches [17, 18] and pen gestures [14, 21, 22, 44–46], with visual feedback and some also provide haptic feedback (e.g., force, vibration). Our work integrates visual, verbal, and haptic inputs and outputs in SenseiPen to enable novel pen-based interactions.

Haptic Pens. Haptic pens have been developed for various GUI applications, with early devices like the Phantom robotic arm providing force feedback to simulate virtual objects and textures via a stylus [28, 37]. For pen-and-paper sketching, haptic offers both passive constraints and active guidance. Comp*Pass [30] allows semi-active drawing on non-digital surfaces, while I-Draw [10] enables smooth transitions between guided and freehand drawing for creative expression. Muscle-Plotter [24] and dePENd [42] use force feedback to guide users in drawing precise shapes. Kianzad et al. [19] introduced a force feedback pen with a ballpoint drive mechanism that assists in following predefined paths while allowing creative deviations. Besides developing haptic hardware, creating haptic content is challenging. Becoming a haptic expert takes years [38], and novices often struggle with programming force feedback devices and synchronizing haptics with visual and audio feedback [39]. Recent studies show that LLMs can effectively produce and explain executable codes. For example, CodeHelp uses LLMs to provide scalable, on-demand programming support [23]. Our work explores LLMs' capabilities in generating haptic force feedback synchronized with visual and audio output for question-answering on any input image without the need for user programming.

Haptics in Learning and Education. Haptic technology improves learning by leveraging embodied cognition reducing cognitive load [43]. It reinforces visual cues while introducing a new sensory channel, such as weight perception, improving comprehension. Umetsu and Kashihara [41] show that pseudo-haptic feedback aids narrative understanding, while Crandall and Karadogan [7] highlight its role in making abstract concepts more tangible, especially in fields like physics and engineering. Additionally, Kaimoto et al. [16] demonstrate how bi-directional sketching interactions in Sketched Reality integrate haptic and visual feedback, bridging virtual and physical environments for enhanced learning experiences.

3 SenseiPen's Interaction Framework

We designed SenseiPen as an AI agent with verbal, visual, and haptic interaction capabilities. The core novelty of our work lies in incorporating haptics as a primary interaction modality with a language model, while the visual and verbal modalities work simultaneously to facilitate free-form communication between the user and the system.

Verbal Interaction. SenseiPen allows users to ask questions, issue commands, or provide explanations with other modalities. It synchronizes LLM's verbal output with visual and haptic cues to enhance user comprehension.

Visual Interaction. SenseiPen can process an input image to understand its content. It also allows users to make rough sketches with a digital pen directly on a tablet interface and write text to specify objects in their drawings. This functionality enables users to clarify instructions and any misunderstandings in real-time. Additionally, it can display visual hints like colored markers or text highlights that dynamically move to reference objects or concepts it describes.

Haptic Interaction. SenseiPen interprets touch gestures like tapping and strokes as haptic inputs to the system and outputs force and vibration cues. User actions, such as tapping on the pen or making strokes or other gestures with SenseiPen, can convey commands like requesting for speedup or encouraging the agent. Resistance by the user against SenseiPen's movements or moving the pen in a direction that encounters resistance is also interpreted as input, with the system actively resisting the user's movements to guide them towards or away from specific regions. SenseiPen can move the user's hand to assist with tasks such as drawing a straight line, following a curve, or learning characters in a different alphabet, offering a kinesthetic learning experience by adjusting its force feedback in real time based on an analysis of the user's movements. This force output not only gives feedback on user actions but also maintains user engagement. Additionally, the system can generate vibrations to communicate system delays and task progress. To achieve this, the system integrates sensors for detecting force and motion with actuators to provide physical feedback.

4 SenseiPen's Hardware

Our platform (Figure 2) features a low-cost, handheld haptic device designed to work with a digital pen and tablet, using the sensing and computing capabilities of the iPad and Apple Pencil. The entire hardware system costs under $55.

Haptic Device. The core of SenseiPen is the ballpoint drive mechanism (Figure 2a, adopted from Kianzad et al. [19, 20]), which generates 2D force feedback via a rolling ball. To ensure compatibility with standard-sized tablets, the ball size was reduced by 50%. The system uses four coreless gear motors, each rated at 3.7V, with a maximum speed of 1200 rpm, a no-load current of 60 mA, and a stall current of up to 200 mA. The motors, fitted with custom 6 mm metal gearheads, interface with the rubber ball to exert up to 0.65 N of force under non-skidding conditions. Other components include sensors, a linear resonance actuator (LRA), and a Nordic nRF52840 chip with Bluetooth Low Energy (BLE) capabilities. This chip enables communication with the tablet for receiving force feedback commands and processing data from integrated sensors, including a PDM microphone, an inertial measurement unit (IMU), and Near Field Communication (NFC). A 3.7V, 700mAh LiPo rechargeable battery powers the system for about two hours of continuous operation.

Key Modifications. SenseiPen's innovation lies in seamless integration with Apple Pencil and iPad. The Apple Pencil captures position, pressure, altitude, and azimuth at 240 Hz, enabling precise tracking without extra sensors. The iPad processes data in real-time and transmits force feedback commands to SenseiPen via Bluetooth. Unlike traditional haptic systems that require extensive onboard processing, our design reduces complexity by relying on the iPad for data processing. Additionally, the ballpoint drive's compact redesign ensures compatibility with standard tablet dimensions, improving usability and portability. These innovations enable SenseiPen to deliver advanced haptic feedback at a fraction of the cost of traditional systems.

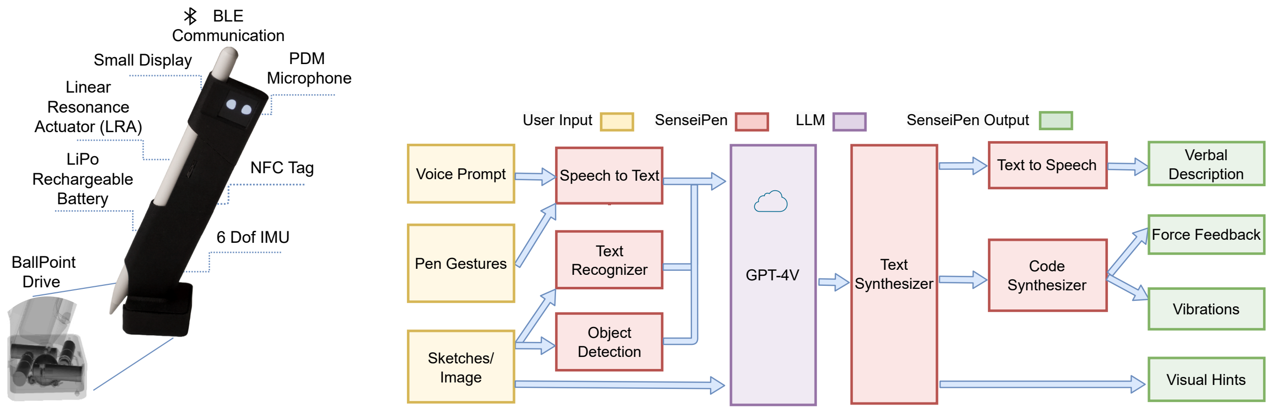
5 Connecting Haptic Feedback with LLM

SenseiPen has software functionalities for (1) vision perception and rendering, (2) speech perception and synthesis, and (3) haptic perception and control. The system's unique contribution lies in its ability to integrate haptic feedback with LLMs like GPT-4V, transforming user input into GPT-4V prompts and providing multimodal output (Figure 2b).

Vision System. The current implementation of multimodal interactions with GPT-4V allows for image input but cannot specify object locations. To overcome this, we use Apple's Core ML for text recognition and YOLOv8 for object detection [35], enabling the system to identify and locate objects and text in images and sketches for haptic guidance.

SenseiPen monitors visual streams (e.g., images, drawings, handwriting) and sends GPT information about detected objects and changes. Screenshots are periodically captured and processed in a dedicated thread, ensuring real-time updates to the visual context. The system also provides visual hints to the user about detected objects. Detected objects and text, along with their positions, are sent to GPT upon user request. The initial prompt includes a screenshot for context, while subsequent prompts omit it unless the user modifies the visual context through sketching or drawing.

Speech Perception and Synthesis. We employ Apple's Speech-to-Text and Text-to-Speech technologies to enable verbal interactions with SenseiPen. Users can ask questions or describe drawings.



(a) SenseiPen's hardware. **(b) Overview of SenseiPen's data processing pipeline from the user input to SenseiPen output.**

Figure 2: SenseiPen's compact ballpoint drive delivers untethered 2D force feedback, integrating sensors and actuators with a multimodal data processing software pipeline for pen-tablet interactions.

Speech-to-text outputs are combined with spatial data from machine vision to create context-aware prompts integrating verbal and visual elements.

SenseiPen uses a custom text synthesizer to handle GPT responses. This synthesizer queues sentences for spoken output and controls the flow of speech based on user input and system-generated haptic commands. When a sentence is queued for speech synthesis, the synthesizer checks if it includes a command, such as directing the user to a specific location. If a command is detected, the system parses the command to execute the corresponding action, such as activating haptic feedback to guide the user towards a target synchronized with the spoken output. The synthesizer also interacts with the haptic perception and control module to monitor the user's progress towards a target position. If the target is not reached within a set time, the system prompts the user to place the pen on the tablet or change their grip. This design ensures that the user receives timely and relevant feedback.

Haptic Perception and Control. The system manages SenseiPen's position and force feedback. When a SenseiPen device is detected, the controller establishes a connection and identifies the required communication parameters. It transmits pulse width modulation (PWM) commands to control force feedback on the tablet. The controller tracks touch gestures, IMU data, and Apple Pencil movements to adapt force feedback. It adjusts PWM values to align with the user's hand orientation. For precise navigation, a PID controller fine-tunes the force feedback based on the target position.

This module makes user drawings and annotations interactive by generating force feedback through Swift's dynamic execution of JavaScript. The platform lets developers script interactions in TypeScript without compile-time development. For example, users can sketch a spring and mass system and ask SenseiPen to make it interactive. SenseiPen parses the request and uses GPT to generate JavaScript code, defining haptic feedback from visual data and speech input. Scripts run in real-time, using SenseiPen's setForce() and getPosition() to adapt force feedback to actions like extending or compressing the spring. SenseiPen continuously evaluates the script to ensure feedback adapts dynamically.

GPT Error Analysis and Correction. Incorporating responses generated by GPT-4V into our system frequently introduced errors, primarily due to the unpredictable nature of the model's output. Within the context of question answering on images, a significant portion of these errors originated from the language model's attempts to integrate force feedback instructions into haptic navigation commands, often resulting in the omission or inclusion of special characters. We built SenseiPen's text synthesizer to identify and address such cases, thereby avoiding any failed force feedback integrations.

6 System Evaluation

We evaluated SenseiPen in a preliminary user study with two goals: (1) assessing the precision and speed of the force feedback mechanism in guiding user hand movements and (2) testing user engagement with SenseiPen against a baseline with only verbal and visual feedback. The study was approved by the university's ethics review board.

Participants. We recruited 16 participants (11 male, 5 female) with a mean age of 23.75 years (std = 1.89). Five participants had prior experience in haptics, including technologies such as vibrotactile feedback in phones and VR controllers (n=3), mid-air ultrasound (n=1), and force feedback devices (n=1). Nine participants had experience with LLMs; four were daily ChatGPT users, and five had been involved in projects utilizing language models. Five of the participants were daily tablet users.

Study Procedure. Each session took 60-75 minutes and participants received $15 Amazon gift cards as compensation. After a background questionnaire and a SenseiPen demonstration, participants were asked to complete two tasks.

The first task is the Navigation Task. Participants used the *SenseiNavigation* app, which presented random red dots on the screen. They were instructed to hold the pen stationary on the tablet, allowing SenseiPen to guide their hand toward a target dot. Each participant completed two practice trials, followed by 10 main trials, emphasizing response speed and interaction fluidity. A condition without tactile feedback was not included in the Navigation Task because prior study by Richard et al. [36] has demonstrated that 2D force feedback not only provides more efficient guidance in navigation tasks compared to vibrotactile feedback but also imposes a lower cognitive load on users.

The second task, the Question-Answering Task, was designed to explore our hypothesis that adding force feedback enhances the guided learning dialogue, making interactions more engaging, rewarding, and effective in maintaining user attention. Participants used the *SenseiInsight* interface (Figure 1) to ask questions about two diagrams: a neuron's structure and the solar system. Each participant experienced two conditions: (A) a baseline condition with verbal and visual responses, where the language model highlighted and narrated the diagram elements, and (B) a SenseiPen condition that added haptic feedback, physically guiding the user's hand to the specific parts of the diagram being described. Each participant completed both scenarios (one per diagram) under each condition, resulting in four trials in total. The order of conditions and diagrams was counterbalanced to mitigate learning effects. The questions posed to the LLM were controlled across participants to ensure consistency in content and difficulty level.

In the final interview, we asked how their interactions with the tool differed from their previous LLM experience, its usefulness in supporting their queries, any challenges they faced, and potential applications for SenseiPen.

Data Collection. In the first task, we logged the time and trajectory of the pen guiding the participant to the target dot, along with the user's hand pressure on the pen. For the second task, user engagement was measured using the User Engagement Scale (UES-SF) [32], capturing ratings after each experimental condition. We also recorded participants' questions, interaction times, and overall usability via the System Usability Scale (SUS) [4] after finishing all tasks. We also made video recordings of user interactions and their interview responses.

7 Results of User Evaluation

SenseiPen's Navigation Efficiency. Navigation between two random dots averaged 1.93 seconds (std=0.75). The total average distance covered was 66.01 mm (std=13.10 mm) across trials. The average velocity was 37.89 mm/s (std=12.58 mm/s). The average pen tip pressure was 47% of the full-scale pressure (std=12%). The Pearson correlation coefficient of (r=0.18, p=0.52) indicates a weak non-significant linear relationship between pressure and average velocity, suggesting minimal association.

The average deviation between the SenseiPen's trajectory and the optimal path was 2.38 mm (std=0.92 mm), outperforming the 6.78 mm error (std=4.88 mm) reported for the Phasking Pen [19] and 4.07 mm (std=3.03 mm) reported for the Muscle-Plotter [24] in similar position control tasks when following a trajectory. When considering the direct line between the initial and target dots, the average deviation could reach up to 20% of the distance to the target. The largest deviation often occurred at the beginning of the navigation task, perhaps due to the users adjusting their grip to balance their own applied pressure with the platform's generated force. Figure 3 visualizes the velocity, pressure, and position errors

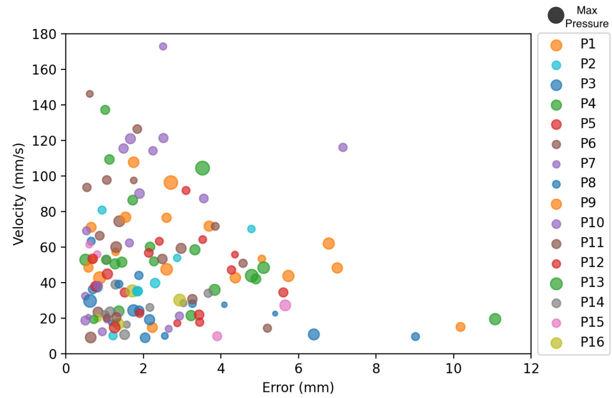


Figure 3: Velocity and positional error measurements in the navigation task. Participants are color-coded. Each circle shows one trial, with circle size indicating pressure exerted; larger circles show higher pressure.

for all participants and trials, showing that 86.25% of trials have errors smaller than 4 mm. Some participants achieved speeds 3 to 4 times faster than the average while maintaining near-average error levels.

Results of the Question-Answering Task. The average rating on the System Usability Scale (SUS) was 72.06 with a standard deviation of 13.12, denoting an acceptable usability according to Bangor et al. [2]. The UES-SF scores are based on ratings on multiple statements and thus are treated as interval variables. We used the paired sample t-test to compare the ratings for SenseiPen and baseline versions. Results were reported at a 0.05 significance level using a two-tailed p-value.

UES-SF Factor	t	P	η_p^2
Aesthetics Appeal	.909	.378	.052
Focused Attention	1.801	.092	.178
Perceived Usability	.915	.375	.053
Reward	**3.521**	**.003**	**.453**
Overall Score	**3.110**	**.007**	**.392**

Table 1: Results of the paired sample t-test for UES-SF factors and the averaged overall score. Test results for Reward and Overall Score showed a significant difference between the SenseiPen and baseline versions at $P < .05$. The df is 15 for all factors.

Table 1 summarizes the paired sample t-test results. The test showed significant differences in the participant ratings for Reward and Overall Score, with large effect sizes. Based on the ratings, SenseiPen provided a more rewarding experience to users after interactions (mean=4.40, std=0.49), compared to the baseline version (mean=3.75, std=0.88). The Overall Score was also significantly higher for SenseiPen (mean=3.96, std=0.41) than the baseline (mean=3.58, std=0.59).

We categorized participant questions and found no notable differences in the type and quantity between the SenseiPen and baseline systems. In total, 158 questions were asked with 86 directed at the SenseiPen version and 72 at the baseline version. On average, they posed about five questions per diagram.

Qualitative Results from Interviews. Most participants (n=11) reported that with the integration of haptic feedback and visual cues, SenseiPen felt more interactive and engaging than previous experiences with ChatGPT. They appreciated the ability to use voice commands, handwriting, and the inclusion of PDFs, which provided different ways of input beyond traditional typing and text input. Some (n=4) highlighted the system's ability to recognize and respond to visual elements, such as diagrams or drawings, as a significant improvement. This feature allowed users to focus on specific points of interest and receive targeted explanations. Some mentioned drawbacks for SenseiPen, such as response delays (n=3) and the inability to interrupt or redirect the system's responses (n=1). Three participants wanted it to be thinner for holding and drawing, with one participant mentioning the pen obstructed their view.

Participants imagined using SenseiPen for various applications. Most (n=11) saw its potential for education, noting how haptic features and visual aids could enhance learning in academic work, scientific discussions, e-textbooks, online classrooms, or skill acquisition. For example, two users mentioned the suitability for children's education and one participant expressed interest in car maintenance, where the system would provide step-by-step guidance on how to fix a car based on a picture they uploaded. Some participants saw the potential for the system to increase engagement in gaming applications (n=4), or to support individuals with visual impairments in understanding the spatial elements in an image (n=3). Others proposed using the system to facilitate code generation from UI drawings, display eye gaze sequences on a page, or guide the user's hand in document signing or UI navigation.

8 Discussion

The results of the Navigation Task demonstrate efficient target acquisition with minimal deviation, exceeding comparable systems. In the Question-Answering Task, higher scores for reward and overall experience suggest that force feedback enhances the guided learning dialogue. Qualitative feedback reinforces this, although response delays and lack of interruption mechanisms were noted.

While YOLOv8 could enable more natural target detection, the vision system combining Apple's Core ML and text-based detection proved more adaptable for interacting with diagrams and handwritten content, leading to more precise and flexible haptic guidance.

Ongoing Work: Interactive Physics Simulation of User Sketches. In addition to guiding user's hand, we are exploring how combining haptic feedback with LLMs could render user-sketched diagrams interactive. We have developed *SenseiPhysics* application, enabling users to draw physics concepts (e.g., a spring-mass system) and describe them to GPT. The system generates code based on these descriptions, and SenseiPen simulates physical sensations, such as the force exerted by a spring. We have pilot tested the application where users sketched various physics-related scenarios, such as pendulums, pulley and mass systems, ball free falls or rolling down a hill, object collisions, spring and mass systems, parabolic ball throw. In about half of the scenarios, *SenseiPhysics* successfully generated code with haptic feedback for the sketch.

We believe that this modest success rate is primarily due to three reasons. First, the novelty effect on users and the lack of experience in designing haptic interactions can lead to prompts that vary in clarity and detail which impacts the effectiveness of the generated simulations. Second, the absence of a specialized haptic compiler for haptic rendering adds a layer of complexity to the task. Translating user's drawing and their descriptions into executable code that accurately simulates the haptic feedback presents a significant challenge. Third, the static nature of user drawings requires a level of interpretation to determine the object's dynamic behaviors and movements. The absence of movement in the drawings means assumptions have to be made about how objects react under various forces, leading to a disconnect between an intended simulation and generated haptic feedback. A potential direction is to leverage a physics engine like RealityKit for more sophisticated and interactive simulations [1]. RealityKit allows for runtime scene changes without the need for pre-compilation, supporting dynamic content integration such as loading 3D models or adding and removing entities based on user interactions.

Limitations and Future Work. This work has several limitations that can be addressed in future research. The first limitation is the SenseiPen's response delay (5-10 seconds), primarily due to the time required to retrieve responses from the GPT-4V API, and the lack of a mechanism to promptly interrupt or redirect system responses. To address these aspects, we plan to integrate on-device LLMs, such as Google's Gemma 2B [12], which requires only 2.5 GB of RAM usage, making it highly efficient for integration on the iPad. Second, our study explores short-term interactions with SenseiPen in a controlled setting. Future work can assess its utility in personalized and long-term tasks to reveal how user strategies and experiences evolve over time. Finally, we did not investigate SenseiPen's design for emotional communication. This study focused on the integration of LLM and force feedback, excluding vibration to isolate the unique contribution of our SenseiPen. Future research could investigate the role of vibration in directing attention to the LLM agent. Future research could design personality profiles [34] for SenseiPen that will make the pen more or less physically active or lead to different force, motion, and vibration profiles synchronized with non-verbal sounds [47]. Furthermore, including long-term memory for SenseiPen similar to generative agents by Park et al. [33] could turn the haptic agent into a low-cost handheld companion for lifelong learning.

9 Conclusion

We designed SenseiPen to enable new user interactions with language models through the integration of multimodal and physical feedback. Our study suggests that haptic feedback can improve user engagement and promote a greater sense of interactivity with LLMs. Our ongoing work focuses on on-demand force feedback haptic content creation for user-drawn physics simulations. As language models gain popularity and usage among the public, we aim to facilitate tailored physical experiences for all users.

Acknowledgments

We thank the anonymous reviewers, our colleagues, and the study participants for their input on this project. This work is supported by research grants from VILLUM FONDEN (VIL50296) and the National Science Foundation (#2339707).

References

[1] Inc Apple. 2023. RealityKit Simulate and render 3D content for use in your augmented reality apps. https://developer.apple.com/documentation/realitykit/. Accessed: (2024/04/01).

[2] Aaron Bangor, Philip T Kortum, and James T Miller. 2008. An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction* 24, 6 (2008), 574–594.

[3] Richard A Bolt. 1980. "Put-that-there" Voice and gesture at the graphics interface. In *Proceedings of the 7th annual conference on Computer graphics and interactive techniques*. 262–270.

[4] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[5] Neil Chulpongsatorn, Mille Skovhus Lunding, Nishan Soni, and Ryo Suzuki. 2023. Augmented Math: Authoring AR-Based Explorable Explanations by Augmenting Static Math Textbooks. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–16.

[6] Philip R Cohen, Michael Johnston, David McGee, Sharon Oviatt, Jay Pittman, Ira Smith, Liang Chen, and Josh Clow. 1997. Quickset: Multimodal interaction for distributed applications. In *Proceedings of the fifth ACM international conference on Multimedia*. 31–40.

[7] Riley Crandall and Ernur Karadoğan. 2021. Designing pedagogically effective haptic systems for learning: a review. *Applied sciences* 11, 14 (2021), 6245.

[8] Fernanda De La Torre, Cathy Mengying Fang, Han Huang, Andrzej Banburski-Fahey, Judith Amores Fernandez, and Jaron Lanier. 2024. LLMR: Real-time prompting of interactive worlds using large language models. *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems* (2024).

[9] Meta Fundamental AI Research Diplomacy Team (FAIR)†, Anton Bakhtin, Noam Brown, Emily Dinan, Gabriele Farina, Colin Flaherty, Daniel Fried, Andrew Goff, Jonathan Gray, Hengyuan Hu, et al. 2022. Human-level play in the game of Diplomacy by combining language models with strategic reasoning. *Science* 378, 6624 (2022), 1067–1074.

[10] Piyum Fernando, Roshan Lalintha Peiris, and Suranga Nanayakkara. 2014. I-Draw: towards a freehand drawing assistant. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: the Future of Design*. 208–211.

[11] Inc Google. 2024. Gemini: Supercharge your creativity and productivity. https://gemini.google.com/. Accessed: (2024/04/01).

[12] Inc Google. 2024. Gemma: Introducing new state-of-the-art open models. https://blog.google/technology/developers/gemma-open-models/. Accessed: (2024/04/01).

[13] Ayaan Haque, Matthew Tancik, Alexei A Efros, Aleksander Holynski, and Angjoo Kanazawa. 2023. Instruct-nerf2nerf: Editing 3d scenes with instructions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 19740–19750.

[14] Ken Hinckley, Koji Yatani, Michel Pahud, Nicole Coddington, Jenny Rodenhouse, Andy Wilson, Hrvoje Benko, and Bill Buxton. 2010. Pen+ touch= new tools. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. 27–36.

[15] Peiling Jiang, Jude Rayan, Steven P Dow, and Haijun Xia. 2023. Graphologue: Exploring large language model responses with interactive diagrams. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–20.

[16] Hiroki Kaimoto, Kyzyl Monteiro, Mehrad Faridan, Jiatong Li, Samin Farajian, Yasuaki Kakehi, Ken Nakagaki, and Ryo Suzuki. 2022. Sketched reality: Sketching bi-directional interactions between virtual and physical worlds with ar and actuated tangible ui. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–12.

[17] Levent Burak Kara. 2004. *Automatic parsing and recognition of hand-drawn sketches for pen-based computer interfaces*. Ph. D. Dissertation. carnegie mellon university.

[18] Levent Burak Kara and Thomas F Stahovich. 2004. Sim-U-Sketch: A sketch-based interface for SimuLink. In *Proceedings of the working conference on Advanced visual interfaces*. 354–357.

[19] Soheil Kianzad, Yuxiang Huang, Robert Xiao, and Karon E MacLean. 2020. Phasking on paper: Accessing a continuum of physically assisted sketching. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.

[20] Soheil Kianzad and Karon E MacLean. 2018. Harold's purple crayon rendered in haptics: Large-stroke, handheld ballpoint force feedback. In *2018 IEEE Haptics Symposium (HAPTICS)*. IEEE, 106–111.

[21] Joseph J LaViola Jr and Robert C Zeleznik. 2006. Mathpad2: a system for the creation and exploration of mathematical sketches. In *ACM SIGGRAPH 2006 Courses*. 33–es.

[22] Chunyuan Liao, François Guimbretière, Ken Hinckley, and Jim Hollan. 2008. Papiercraft: A gesture-based command system for interactive paper. *ACM Transactions on Computer-Human Interaction (TOCHI)* 14, 4 (2008), 1–27.

[23] Mark Liffiton, Brad E Sheese, Jaromir Savelka, and Paul Denny. 2023. Codehelp: Using large language models with guardrails for scalable support in programming classes. In *Proceedings of the 23rd Koli Calling International Conference on Computing Education Research*. 1–11.

[24] Pedro Lopes, Doăa Yüksel, François Guimbretière, and Patrick Baudisch. 2016. Muscle-plotter: An interactive system based on electrical muscle stimulation that produces spatial output. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. 207–217.

[25] Zilin Ma, Yiyang Mei, and Zhaoyuan Su. 2023. Understanding the benefits and challenges of using large language model-based conversational agents for mental well-being support. In *AMIA Annual Symposium Proceedings*, Vol. 2023. American Medical Informatics Association, 1105.

[26] Karthik Mahadevan, Jonathan Chien, Noah Brown, Zhuo Xu, Carolina Parada, Fei Xia, Andy Zeng, Leila Takayama, and Dorsa Sadigh. 2024. Generative expressive robot behaviors using large language models. *ACM/IEEE International Conference on Human-Robot Interaction* (2024).

[27] Damien Masson, Sylvain Malacria, Géry Casiez, and Daniel Vogel. 2024. Directgpt: A direct manipulation interface to interact with large language models. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*.

[28] Marilyn Rose McGee, Philip Gray, and SA Brewster. 2002. Mixed feelings: Multimodal perception of virtual roughness. In *Proceedings of Eurohaptics*, Vol. 2002. Citeseer, 47–52.

[29] Tania Morimoto, Melisa Orta Martinez, Richard Davis, Paulo Blikstein, and Allison M Okamura. 2020. Teaching with hapkit: Enabling online haptics courses with hands-on laboratories. *IEEE Robotics & Automation Magazine* 28, 3 (2020), 79–91.

[30] Ken Nakagaki and Yasuaki Kakehi. 2014. Comp* pass: A compass-based drawing interface. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. 447–450.

[31] Inc OpenAi. 2024. GPT-4 is OpenAI's most advanced system, producing safer and more useful responses. https://openai.com/gpt-4. Accessed: (2024/04/01).

[32] Heather L O'Brien, Paul Cairns, and Mark Hall. 2018. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies* 112 (2018), 28–39.

[33] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–22.

[34] Alisha Pradhan and Amanda Lazar. 2021. Hey Google, do you have a personality? Designing personality and personas for conversational agents. In *Proceedings of the 3rd Conference on Conversational User Interfaces*. 1–4.

[35] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 779–788.

[36] Grégoire Richard, Thomas Pietrzak, Ferran Argelaguet, Anatole Lécuyer, and Géry Casiez. 2021. Studying the role of haptic feedback on virtual embodiment in a drawing task. *Frontiers in Virtual Reality* 1 (2021), 573167.

[37] J Kenneth Salisbury and Mandayam A Srinivasan. 1997. Phantom-based haptic interaction with virtual objects. *IEEE Computer Graphics and Applications* 17, 5 (1997), 6–10.

[38] Oliver Schneider, Karon MacLean, Colin Swindells, and Kellogg Booth. 2017. Haptic experience design: What hapticians do and where they need help. *International Journal of Human-Computer Studies* 107 (2017), 5–21.

[39] Hasti Seifi, Matthew Chun, Colin Gallacher, Oliver Schneider, and Karon E MacLean. 2020. How do novice hapticians design? A case study in creating haptic learning environments. *IEEE transactions on haptics* 13, 4 (2020), 791–805.

[40] Ivan E Sutherland. 1963. Sketchpad: A man-machine graphical communication system. In *Proceedings of the May 21-23, 1963, spring joint computer conference*. 329–346.

[41] Kazuaki Umetsu and Akihiro Kashihara. 2017. Pseudo-Haptic Feedback for Promoting Narrative Comprehension. *International Association for Development of the Information Society* (2017).

[42] Junichi Yamaoka and Yasuaki Kakehi. 2013. dePENd: augmented handwriting system using ferromagnetism of a ballpoint pen. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. 203–210.

[43] Zacharias C Zacharia. 2015. Examining whether touch sensory feedback is necessary for science learning through experimentation: A literature review of two different lines of research across K-16. *Educational Research Review* 16 (2015), 116–137.

[44] Robert Zeleznik, Andrew Bragdon, Ferdi Adeputra, and Hsu-Sheng Ko. 2010. Hands-on math: a page-based multi-touch and pen desktop for technical work and problem solving. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. 17–26.

[45] Robert Zeleznik and Timothy Miller. 2006. Fluid inking: augmenting the medium of free-form inking with gestures. In *Proceedings of graphics interface 2006*. 155–162.

[46] Robert C Zeleznik, Kenneth P Herndon, and John F Hughes. 2006. SKETCH: An interface for sketching 3D scenes. In *ACM SIGGRAPH 2006 Courses*. 9–es.

[47] Brian J Zhang and Naomi T Fitter. 2023. Nonverbal sound in human-robot interaction: a systematic review. *ACM Transactions on Human-Robot Interaction* 12, 4 (2023), 1–46.

[48] Zheng Zhang, Jie Gao, Ranjodh Singh Dhaliwal, and Toby Jia-Jun Li. 2023. Visar: A human-ai argumentative writing assistant with visual programming and rapid draft prototyping. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–30.

[49] Dmytro Zhelezniakov, Anastasiia Cherneha, Viktor Zaytsev, Tetiana Ignatova, Olga Radyvonenko, and Oleg Yakovchuk. 2020. Evaluating new requirements to pen-centric intelligent user interface based on end-to-end mathematical expressions recognition. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*. 212–220.