Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Bespoke anatomy education without coding: a proof of concept for AI-driven simulation generation

Ghazal Ershadi-Oskoui
University of Oxford
Helen Christian
University of Oxford
Sharmila Saran Rajendran
sharmila.rajendran@dpag.ox.ac.uk
University of Oxford

Short Report

Keywords: anatomy education, large language models, generative artificial intelligence, simulation, interactive learning, educational technology

Posted Date: November 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11468822/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Interactive digital resources are widely used in anatomy education, but their development typically requires programming expertise or commercial software, and pre-existing resources may not match an educator's specific teaching requirements. Large language models (LLMs) are now capable of generating functional web code from conversational prompts, creating the possibility of educators producing their own interactive simulations without programming experience. Using LLMs to generate simulations for anatomical education, rather than as tutoring or assessment aids, remains under-investigated. This brief report is a proof-of-concept workflow in which three LLM platforms (OpenAI Codex, Google Gemini, and Anthropic Claude) were provided with reviewed anatomy content and prompted to generate interactive, browser-based simulations. Ten single-file Hypertext Markup Language (HTML) simulations were produced covering cranial nerve lesion testing, the pupillary light and corneal reflexes, extraocular muscle testing, visual field defects, spinal cord syndromes, the circle of willis, and the urinary sphincter function system. The simulations incorporate AI (Artificial Intelligence)-generated anatomical images, adjustable lesion parameters, and scenario-based clinical reasoning. Iterative refinement was necessary to achieve accurate results, as the AI repeatedly made errors, including confusion between left and right orientations, inaccurate positioning of facial features, and the occasional generation of images instead of modifying the code as instructed. These challenges required multiple rounds of correction and careful human oversight to ensure the final output met the intended requirements. Preliminary simulation assessment and comments from Medicine Year 2 colleagues, Oxford AI Build club, learning technologist, Teaching demonstrators were encouraging regarding clarity and engagement, although no formal student evaluation of learning outcomes has yet been conducted. We conclude that LLMs can translate structured anatomical knowledge into functional interactive anatomy simulations and describe a replicable workflow, while emphasising that expert review remains indispensable.

Introduction

Anatomy education has evolved over the past two decades by digital resources, a shift which accelerated sharply during the COVID-19 pandemic when rapid digital upskilling became essential across medical and health science schools [1]. Educators now have access to a wide variety of technologies such as virtual dissection tables, three-dimensional modelling software, and virtual and augmented reality platforms, each with distinct advantages, limitations, and costs [1]. Within this landscape, sophisticated digital anatomy platforms have demonstrated the educational value of interactivity: Complete Anatomy (3D4Medical/Elsevier), for example, provides interactive three-dimensional models that medical students can manipulate, dissect, and use for self-directed study, and its integration into curricula has been positively evaluated by students and educators [2].

In contrast, bespoke simulations offer greater flexibility and design control, allowing educators to tailor content, features, and learning activities to better align with curriculum requirements and specific teaching and learning needs. Nevertheless, developing such customised resources has traditionally

required programming expertise and considerable time and resources [3]. Recent work in physiology and physics education has shown that this barrier is falling: instructors with little or no programming experience can now generate browser-based interactive teaching aids using generative AI and accessible developmental tools [3]. Comparable prompt-based workflows have been used to build validated physics simulations without programming expertise [4].

In parallel, LLMs have attracted extensive attention in medical and anatomy education, principally as tutoring aids, assessment generators, and sources of personalised learning material [5, 6]. Customised applications such as AnatomyGPT have been configured as intelligent tutoring systems for anatomical education [7]. Evaluations of ChatGPT in anatomy education have documented both useful capabilities in explaining anatomical concepts and generating quiz questions, and significant limitations in accuracy and image generation [8]. Systematic reviews of AI in anatomy teaching have identified applications including virtual dissection simulations, adaptive feedback, and gamified learning, while noting concerns about over-reliance and diminished human interaction [9].

In contrast, using LLMs to generate interactive simulations, where the model produces the functioning educational resource itself rather than explanations or assessment items, has received much less attention [10]. This is a crucial distinction: generating a functional simulation requires the model to translate structured anatomical knowledge into working interactive code with correct clinical logic, a task considerably more demanding than generating prose.

This article describes a proof-of-concept workflow in which three LLM platforms were used to generate ten interactive, browser-based anatomy simulations, documents the practical challenges encountered, and reports preliminary informal feedback. Our aim is to provide anatomy educators with a transparent, replicable account of what this approach can and cannot currently achieve. The goal was not to replace anatomy educators, but to assess whether LLMs can function as a useful development tool under expert direction.

Methods

Design goals and educational context

The project aimed to produce interactive simulations suitable for undergraduate medical anatomy teaching, with an initial focus on cranial nerves. Cranial nerves were chosen as the demonstration case because they are examined routinely in neurological assessment and because lesions to individual nerves produce predictable, testable clinical signs. This creates an ideal educational scenario in which students can connect structure with clinical presentation and directly explore cause-and-effect relationships instead of memorising isolated facts.

There were four design principles which guided development. Firstly, each simulation was built as a single self-contained HTML file with embedded styling and scripts so that it runs in any standard web browser on any device without installation, following the approach described for browser-based teaching

aids in physiology education [3]. Second, every simulation placed the learner in an active role such as selecting lesions, dragging examination targets, adjusting compression or severity sliders, and observing resulting clinical signs. Third, clinical relevance was spotlighted: each simulation displays the underlying neural pathway, the simulated examination finding, and an examiner-style teaching note. Fourth, there is an on-screen disclaimer, where relevant, identifying it as an educational simplification and not a diagnostic assessment tool.

The key technical development was the use of an embedded image-state architecture for simulations in which neurological findings depended on complex changes in anatomical appearance. This was particularly effective for the facial and hypoglossal nerve simulations, where reproducing facial expression, eye closure, forehead movement or tongue deviation through simple vector transformations risked creating unrealistic findings. Instead, separate clinical illustrations were prepared for each examination state. In the facial nerve simulation, learners could examine four patients at rest and during eyebrow elevation, eye closure, smiling and cheek puffing. The hypoglossal simulation used the same principle for tongue appearance at rest and during protrusion (Fig. 1).

Each image could therefore be individually reviewed and corrected before incorporation. Approved images were encoded directly within the HTML rather than referenced as external JPEG or PNG files. This eliminated dependence on accompanying image folders and produced a single, portable, standalone HTML resource. JavaScript mapped each patient and examination selection to the appropriate embedded image. Superimposed image layers and opacity transitions enabled smooth switching between states without reloading the page or changing the position of the anatomical model. Particular attention was required to maintain image scale, patient orientation and clinical laterality.

Vector-Based Animation and Interactive Anatomical Behaviour

A second technical approach involved vector-based animation, used when the educational objective depended on movement, physiological change or direct manipulation rather than switching between fixed clinical images. This was particularly suitable for the H-test, pupillary light reflex and corneal reflex and gag reflex simulations. Anatomical components were represented as separate scalable vector elements, allowing selected structures to move or change while the surrounding anatomy remained stable.

In the H-test, for example, moving a target produces corresponding eye movements, letting learners explore normal ocular motility and patterns associated with different cranial nerve lesions (Fig. 2). The pupillary light reflex similarly uses dynamically controlled pupil size: applying a light stimulus to either eye changes the relevant pupil elements according to the selected normal or lesioned condition. The corneal reflex uses vector-controlled eyelid movement to demonstrate bilateral or abnormal blink responses (Fig. 3).

Unlike image-state switching, vector animation allowed **c**ontinuous, responsive changes in response to learner actions. However, this approach introduced different challenges. Movement direction, speed control, and coordination between structures, where not to camouflage other anatomical structures, required repeated adjustment; technically functional animation could still communicate an incorrect clinical response if these parameters were wrong. Consequently, vector animation required validation of anatomical, physiological and interaction fidelity.

Accessibility, Adaptability and Inclusive Interface Design

Another design consideration was ensuring the simulations could accommodate learners using different devices, display conditions, and interaction preferences. Several simulations incorporated light and dark display modes, allowing contrast and background appearance to be adjusted without altering the underlying educational content. However, the development process also demonstrated that optional interface features should be retained only where they improve accessibility or usability; otherwise, additional controls can compete with the principal learning task.

Responsive scaling was especially challenging where multiple visual components occupied the same screen, such as the spinal cord cross-section and accompanying body maps.

Scalable font sizes, concise instructions, clear visual hierarchy, and sufficiently large interactive targets supported readability. Patient laterality was explicitly indicated where orientation was important for clinical interpretation.

Teaching and Testing Modes

A further theme emerging from the simulations was the distinction between teaching/exploration and testing/localisation. In teaching mode, the neurological condition is identified in advance, allowing learners to manipulate the examination and observe how a known lesion alters function. Examples include selecting a CN II or CN III lesion in the pupillary light reflex and comparing the resulting direct and consensual blink responses (Fig. 3). This supports exploration of cause-and-effect relationships without requiring an immediate diagnosis.

In testing mode, diagnostic information is deliberately concealed. The facial and hypoglossal simulations, for example, use Patient A-D rather than named lesions. Learners perform the examination, identify abnormal findings and use these observations to determine lesion side and localisation. The development process showed that these modes serve different educational purposes and should not necessarily be forced into a single format. Teaching mode supports concept acquisition, whereas testing mode supports retrieval, application and clinical reasoning. A simulation could potentially offer both sequentially: learners first explore known lesions and subsequently encounter unknown patient cases.

The simulations

Table 1 summarises the ten completed simulations, including one broader visceral system (the urinary tract) was included to test whether the workflow generalised beyond neuroanatomy.

Table 1
The ten simulations developed in this project.

Simulation	Educational focus	Key observations
Bitemporal hemianopia OSCE simulator	Visual pathway anatomy; chiasmal compression	Six-stage objective structured clinical examination (OSCE)-style station: history, inspection and acuity, confrontation visual fields with two compression scenarios, adjustable patient's-eye-view simulation of chiasmal compression, clinical reasoning, and debrief stages
Pupillary light reflex	Afferent and efferent pupillary pathways	Selection of optic or oculomotor nerve damage on either side; torch applied to either eye; observation of direct and consensual responses
Corneal reflex	Trigeminal afferent and facial efferent reflex arc	Dragging a cotton wisp onto either cornea; selection of afferent or efferent lesion side; comparison of direct and consensual blink responses
H test (extraocular muscles)	Oculomotor, trochlear, and abducens nerve function	Dragging a target along the H pattern through the six cardinal positions of gaze; selection of nerve palsy side and type; observation of resulting gaze deviation
Facial nerve lesion	Upper and lower motor neuron facial weakness	Selection of lesion type (upper versus lower motor neuron), side, facial expression, and weakness severity; side-specific deficit overlays on a photographic face
Spinal cord syndromes	Lesion localisation (Brown-Séquard, anterior, central, and posterior cord, conus medullaris, cauda equina syndromes)	Selection of spinal level, side, and lesion severity; pathway overlays on a spinal cord cross-section; live readouts of motor, dorsal column, pain/temperature, and autonomic function
Circle of Willis	Cerebral arterial circle structure and flow	Clickable vessel identification; animated blood flow with adjustable speed and particle density; display of each vessel's role
Urinary system	Urinary tract anatomy, microstructure, and micturition control	Clickable organ identification; adjustable detrusor activity and sphincter tone; estimated flow and voiding simulation
Hypoglossal nerve examination	CN XII motor pathway; upper and lower motor neuron lesion localisation	Selection of four unknown patient cases; inspection of the tongue at rest and during protrusion; observation of wasting, fasciculations, asymmetry and deviation to determine lesion type and laterality
Gag reflex	Afferent CN IX and efferent CN X pathways	Selection of normal or unilateral CN IX/CN X damage; stimulation of either side of the posterior pharynx; observation of reflex presence or absence, palatal elevation and asymmetry to distinguish sensory from motor lesions

Generation workflow

Stage 1: Content curation. For each topic, structured anatomy knowledge was prepared in advance which included normal anatomy, neural pathways, and the specific clinical signs produced by defined lesions, together with clearly defined educational objectives for the simulation. This input was treated as an essential quality control step rather than an optional refinement. Following content planning, a standardised desktop-based layout was designed using Microsoft PowerPoint to provide the large language model (LLM) with a visual template indicating the intended placement of simulations and their associated assessment questions within a single, wide-screen interface. This design aimed to minimise excessive vertical scrolling by enabling learners to view and interact with the simulations and related questions simultaneously, thereby reducing unnecessary cognitive load and facilitating a more coherent learning experience.

Stage 2: Platform selection. Three LLM platforms were used: OpenAI Codex, Google Gemini, and Anthropic Claude. All three were capable of generating functional simulation code, and the final set of simulations drew on more than one platform. Differences between platforms related mainly to code structure, the level of refinement, and sensitivity to prompt phrasing rather than to overall feasibility.

Stage 3: Asset generation and embedding. Anatomical illustrations for the circle of Willis, urinary system, and spinal cord simulations were generated using Gemini's image generation tool, Nano Banana. The image files were then Base64-encoded using a free online conversion tool and embedded inline into the HTML source as data Uniform Resource Identifier's (URIs) (RFC 2397), so that each simulation renders its images directly from the embedded data without external image requests. The dark-mode toggle used across the suite was also generated using Gemini. This embedding approach produced fully self-contained, single-file HTML documents.

Stage 4: Iterative refinement. Generated code was reviewed, rendered in a browser, tested against the curated clinical content, and corrected through repeated prompts. Refinement addressed both technical behaviour and anatomical-clinical correctness. Three recurring problems are worth documenting because they were systematic rather than random. First, laterality: the models frequently confused patient-side and viewer-side left and right, producing mislabelled structures in the circle of Willis simulation and incorrect side assignment in the facial nerve lesion simulation; correcting this required explicit instructions distinguishing the patient's right from the viewer's left. Second, spatial placement: positioning facial features such as eyebrows, eyes, and mouth in the facial nerve simulation required pixel-level directional instructions (for example, moving a feature up, down, left, or right by a specified amount). Third, modality confusion: after requesting adjustments, the model would sometimes generate a new image rather than modifying the simulation code; explicitly instructing it to "update the HTML accordingly" was needed to keep the conversation in code-modification mode.

Stage 5: Deployment. Each finished simulation was saved as a standalone HTML file and shared through the institution's cloud storage, where it can be opened directly in a browser. This deployment route

requires no learning management system integration and no technical support.

Discussion

Principal findings

This proof-of-concept project demonstrates that current LLMs can translate structured, expert-curated anatomy knowledge into functional, interactive, browser-based simulations with no need for prior programming expertise from the educator. All three platforms evaluated generated working simulation code, consistent with recent reports in adjacent disciplines: LLM-assisted coding has enabled physiology educators to produce interactive teaching graphs and apps from small numbers of prompts [3], and prompt-based workflows have generated validated physics simulations covering systems of differing complexity [4]. To our knowledge, however, a documented multi-simulation anatomy workflow of this kind, with an explicit account of failure modes, has not previously been described and recent reviews have identified the generation of educational resources by LLMs as an under-investigated application [10, 6].

The centrality of expert oversight

The most significant finding of this work lies not in what the models could achieve independently, but in what they could accomplish through sustained expert guidance and iterative refinement. Every simulation in this suite depended on three forms of expert input: structured anatomical content supplied before generation, iterative clinical review of the rendered output, and correction of systematic errors. The recurring laterality problem deserves particular emphasis. In lesion-localisation teaching, the relationship between the side of a lesion and the side of a deficit is precisely the learning objective; a simulation that mislabels the patient's right as the viewer's left does not merely look wrong, it teaches an incorrect localisation rule. This mistake would not be obvious to a learner, which makes pre-deployment expert review non-negotiable. This conclusion aligns with the broader literature on LLMs in anatomy education, in which factual inaccuracy and the necessity of expert verification are consistently reported [8, 5]. It also parallels the distinction drawn in physics education between technical validation, testing whether the simulation behaves correctly, and domain validation, which compares outputs against established knowledge [4]. In anatomy, domain validation requires an anatomist as it relies on expert knowledge rather than an analytical solution.

The modality confusion observed here, in which the model produced a new image when a code modification was requested, demonstrates that productive use of these tools is a skill. Outputs were highly sensitive to prompt phrasing, and effective prompting required anticipating how the model would interpret an instruction. Educators adopting this workflow should expect a learning curve in prompt craftsmanship, comparable to that reported in other LLM-assisted development contexts [3].

Implications for anatomy educators

There are three practical implications which follow. Firstly, the workflow described here is low-cost and scalable. Beyond subscription access to the LLM platforms, no software, server, or development budget was required, and each simulation is a single portable file that can be distributed through a virtual learning environment or cloud storage. This positions LLM-generated simulations as a complement to commercial digital anatomy resources, whose selection and implementation educators already navigate with caution [1]. Second, the approach allows educators to tailor simulations precisely to their own curriculum and teaching style, in contrast to adapting pre-made content. A further design decision was the deliberate inclusion of light and dark display-mode toggles across the simulations. Evidence on display modes is mixed: light mode has been associated with lower subjective mental effort and reading preference, while dark mode reduces visual fatigue in low-light conditions but increases pupil dilation, which can make focusing on text more effortful; recommendations therefore suggest matching the display mode to ambient room lighting and to the specific learning task [11, 12]. Within this project, both the technical and academic reviewers regarded the toggles as a minor addition rather than an essential feature, and future versions could reasonably omit them or retain them only where they support the learning context. Third, the workflow is transparent and auditable since the output is human-readable HTML meaning educators can inspect and correct any element of the resource, an advantage over commercial platforms whose underlying content and functionality cannot be inspected directly. In practice, the single-file simulations can be uploaded to a virtual learning environment such as Canvas and linked to existing teaching materials, so that they sit within the curricular content students already access rather than forming a separate resource. As standalone HTML files cannot themselves track usage, informal mechanisms such as a short feedback form would be needed to monitor engagement; alternatively, where assessment questions do not depend on the exact simulation output, embedding the simulations alongside virtual learning environment quiz or H5P activities would allow learner responses to be tracked within the platform.

Limitations

Several limitations must be acknowledged. First, this is a proof-of-concept report, not an effectiveness study. No learning outcomes were measured, and the informal feedback collected only concerned two of the ten simulations, from small and self-selected groups. Second, formal expert anatomical review of all ten simulations is in progress; the accuracy of the simulations cannot yet be assumed, and any AI-generated educational content requires expert review before use in education. Third, LLM outputs are variable: regenerating a simulation from the same prompt may yield different code, so the workflow as described is not deterministic and each artifact requires individual review. Fourth, AI-generated anatomical images are themselves subject to error and were reviewed alongside the code; image quality and anatomical fidelity varied. As a related consideration, Base64 encoding increases file size by approximately one third relative to the original image files; the trade-off was accepted in favour of single-file portability. Fifth, the simulations deliberately simplify physiology, which is why disclaimers are necessary on-screen. Finally, early technologist review of the simulations identified several interface refinements still outstanding. These include reducing unnecessary on-screen text such as prominent titles and author details, improving use of screen space when pages are resized (for example in the H-

test simulation), simplifying status messages such as automatic condition-change notifications, and adopting a unified visual theme with consistent question types across all simulations. More broadly, accessibility has not yet been formally addressed; the simulations are strongly visual so future versions should incorporate screen-reader compatibility and adaptable presentation, following the accessibility considerations raised for browser-based teaching tools [3].

Future work

Planned next steps follow directly from these limitations. A formal evaluation with undergraduate students, including usability measures and learning outcomes, is planned, alongside systematic comparison of the three platforms on identical prompts. The improvements suggested by questionnaire respondents, including brief explanatory content and links to further reading will be incorporated. We also intend to explore whether students can safely use this workflow to build their own revision resources under supervision, an application foreshadowed in the physiology education literature [3].

Conclusion

The preliminary work presented here indicates that large language models can translate structured anatomy knowledge into functional interactive simulations, and it documents a replicable workflow for educators without prior programming experience. Importantly, AI should be viewed as an educational development tool rather than a replacement for educators. Expert oversight is essential, and anatomical accuracy cannot be assumed at this stage. Within those constraints, this approach offers a scalable, cost-efficient route to interactive learning resources that are tailored to an educator's own curriculum, and it contributes to the responsible and innovative integration of AI into anatomy education.

Declarations

Acknowledgements

Conflicts of Interest

The authors declare no conflicts of interest.

Use of Generative Artificial Intelligence

We used AI to polish the language, generating images and vector-based illustration for the simulation purposes.

Ethics Statement

Not Applicable

Data Availability

The simulation HTML files are available from the corresponding author upon reasonable request.

Funding

Our Proof of concept work was funded by the Anatomical Society, UK. Undergraduate Summer Vacation Research Scholarships. Anatomical Society Undergraduate Awards 2025/26

References

1. Aland RC, Hugo HJ, Battle AR, et al. A plethora of choices: an anatomists' practical perspectives for the selection of digital anatomy resources. Smart Learn Environ. 2023;10:66. https://doi.org/10.1186/s40561-023-00285-3.
2. Steinborn A, Berger R, Brunk I. Evaluation der App Complete Anatomy durch Studierende der Zahn- und Humanmedizin und Lehrende der Anatomie an der Charité – Universitätsmedizin Berlin. GMS German Med Sci. 2022;20:Doc05. https://doi.org/10.3205/mbi000530.
3. Samways DSK, Yousef L, Sheffield SA. Creating accessible browser-based interactive content for teaching physiology and pharmacology: options for nonprogrammers. Adv Physiol Educ. 2025;49:1102–8. https://doi.org/10.1152/advan.00127.2025.
4. Ben-Zion Y, Einhorn Zarzecki R, Glazer J, Finkelstein ND. Leveraging AI for rapid generation of physics simulations in education: building your own virtual lab. arXiv preprint. 2024. https://doi.org/10.48550/arXiv.2412.07482. arXiv:2412.07482.
5. Leng L. Challenge, integration, and change: ChatGPT and future anatomical education. Med Educ Online. 2024;29(1):2304973. https://doi.org/10.1080/10872981.2024.2304973.
6. Xu X, Chen Y, Miao J. Opportunities, challenges, and future directions of large language models, including ChatGPT in medical education: a systematic scoping review. J Educational Evaluation Health Professions. 2024;21:6. https://doi.org/10.3352/jeehp.2024.21.6.
7. Collins BR, Black EW, Rarey KE. Introducing AnatomyGPT: a customized artificial intelligence application for anatomical sciences education. Clin Anat. 2024;37(6):661–9. https://doi.org/10.1002/ca.24178.
8. Saluja S, Tigga SR. Capabilities and limitations of ChatGPT in anatomy education: an interaction with ChatGPT. Cureus. 2024;16(9):e69000. https://doi.org/10.7759/cureus.69000.
9. Joseph T, Gowrie S, Montalbano MJ, et al. The roles of artificial intelligence in teaching anatomy: a systematic review. Clin Anat. 2025. https://doi.org/10.1002/ca.24272.
10. Li R, Wu T. Delving into the practical applications and pitfalls of large language models in medical education: narrative review. Adv Med Educ Pract. 2025;16:625–36. https://doi.org/10.2147/AMEP.S497020.
11. Xie X, Song F, Liu Y, Wang S, Yu D. Study on the effects of display color mode and luminance contrast on visual fatigue. IEEE Access. 2021;9:82125–35. https://doi.org/10.1109/ACCESS.2021.3061770.

12. Ettling T, Steinmann D, Bektaş K, Abbad-Andaloussi A. An eye tracking study on the effects of dark and light themes on user performance and workload. ETRA. 2025. https://doi.org/10.1145/3715669.3725879.

Figures

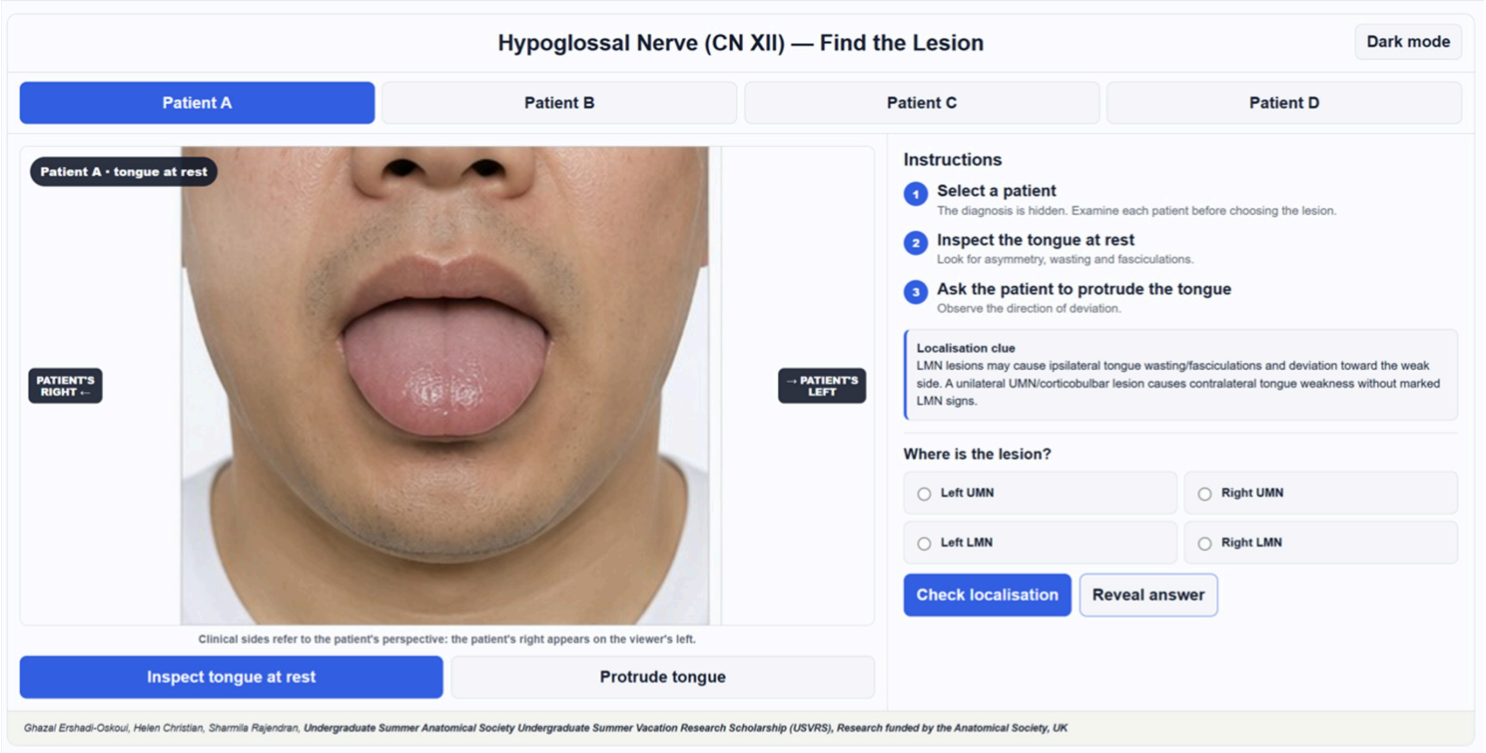


Figure 1

Hypoglossal nerve examination simulation in testing mode. The learner selects among four anonymised patient cases and examines tongue appearance at rest and during protrusion before localising the lesion from findings such as wasting, fasciculations, asymmetry and deviation.

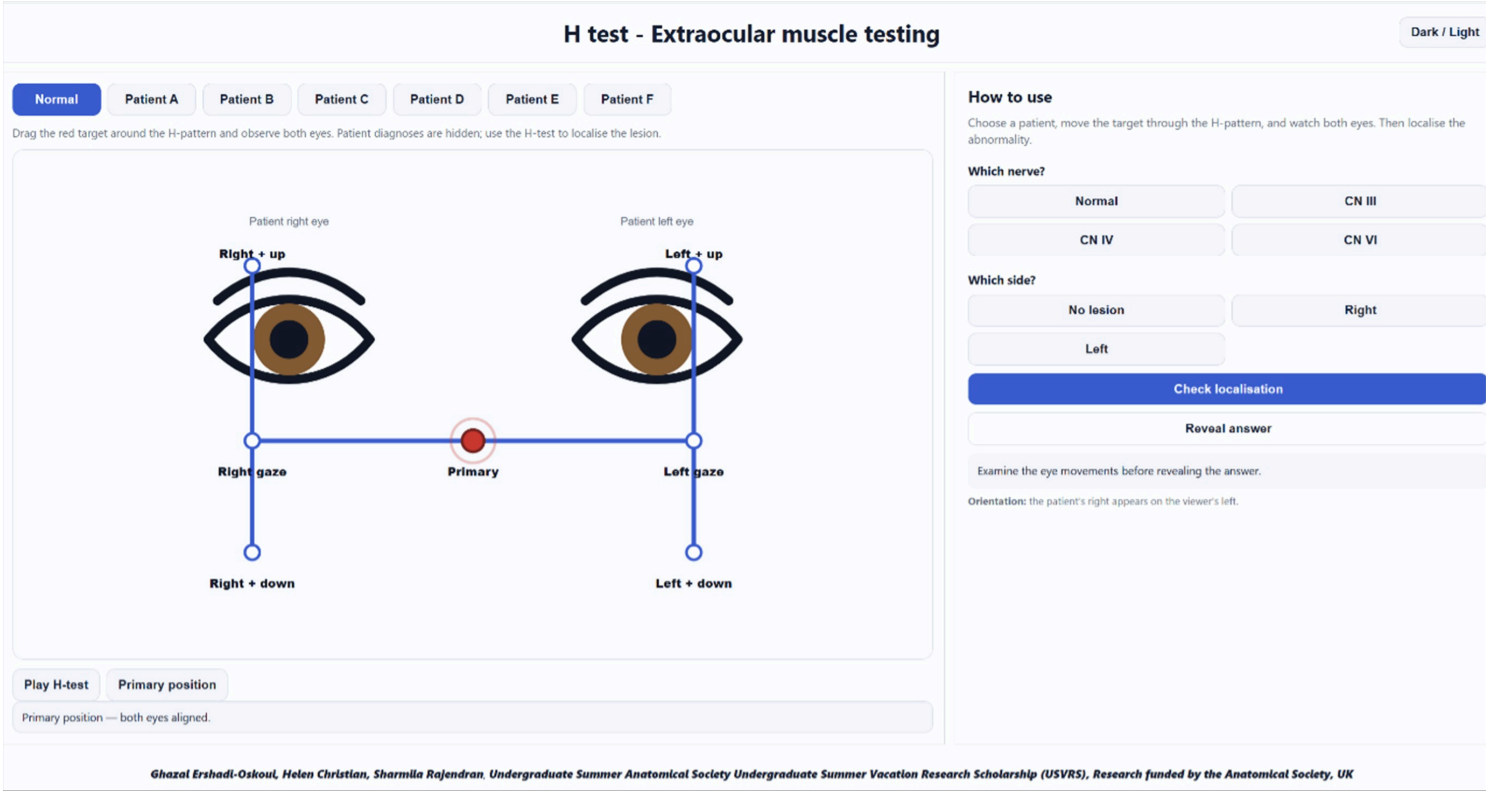


Figure 2

H-test simulation in testing mode. The learner selects a patient case, moves a target through the H-shaped path across the six cardinal positions of gaze, observes both eyes and localises the lesion while the diagnosis remains hidden.

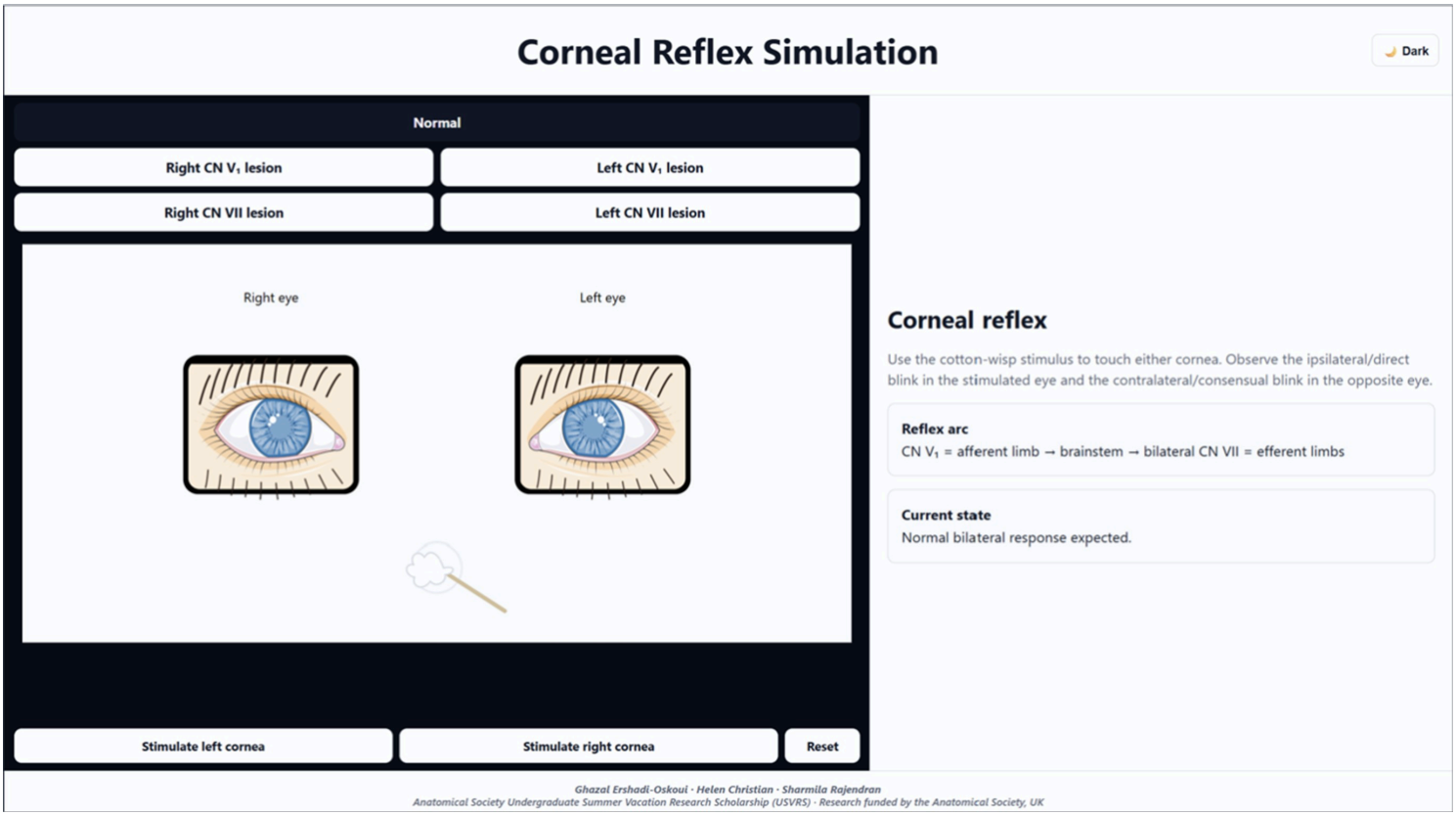


Figure 3

Corneal reflex simulation in teaching mode. A normal condition or a left- or right-sided CN V1 afferent or CN VII efferent lesion is selected before simulation. The learner stimulates either cornea and observes the predicted direct and consensual blink responses alongside the reflex pathway.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- HTestsupplementaryvideo.mp4