

TEESSIDE UNIVERSITY
School of Computing, Engineering & Digital Technologies

CIS3040-N Computing Project
Final Project Proposal

An Investigation into How Rule-Based Safety Checks Can Make AI-Supported Home Exercise Guidance Safer and More Accessible for People Following a Physiotherapy Programme (the Physiq App)

Item	Detail
Student	Blessing Masukwedza
Student ID	Provided on the submitted copy
Programme	BSc (Hons) Information Technology (BCS accredited)
Supervisor	Mansha Nawaz
Second reader	
Live prototype	https://byncx.dev/
Source repository	https://github.com/Yungtaytay/Physiq
Project diary	https://byncx.dev/blog.html
Submission	Friday week 5, Semester 1 (30 October 2026), 16:00

Table of Contents

1. Working Title

An Investigation into How Rule-Based Safety Checks Can Make AI-Supported Home Exercise Guidance Safer and More Accessible for People Following a Physiotherapy Programme (the Physiq App)

2. Outline Description and Research Question

Over 10 million people in the UK live with arthritis, about one in six (Arthritis UK, 2026), and the World Health Organization (WHO, 2024) estimates that 2.4 billion people worldwide live with a condition that may benefit from rehabilitation. Much of physiotherapy depends on exercises people do alone at home, and keeping to them is hard. A systematic review of ten trials found that digital tools improved adherence to physiotherapist-prescribed home exercise in seven of them, mainly in the short term (Lang *et al.*, 2022), and another review found strong evidence of benefit for back conditions (Nagel *et al.*, 2024).

AI coaching could help, but it is risky. A systematic review found safety flaws in 14 of 24 studies of large language model (LLM) exercise advice, including exercises that were contraindicated for clinical groups (He *et al.*, 2026), and the Chartered Society of Physiotherapy (CSP, 2025) warns that chatbots can encourage over-reliance and fail to pass problems to a human therapist. My project takes **Physiq**, my existing Expo/React Native prototype, and refocuses it on supporting a home exercise programme that a physiotherapist or doctor has set. It adds a rule-based safety layer between the AI and the user, accessible design for disabled people and people with long-term conditions, and a proper evaluation.

Research question: *To what extent can a mobile app that combines rule-based safety checks with an LLM help people follow a physiotherapist-set home exercise programme in a way that is safe, accessible and useful, including for disabled adults and adults with long-term conditions?*

3. Proposed Project Artefact and Deliverables

The artefact is a cross-platform app (iOS, Android and web) with a small serverless back end, plus a public project journal website. Physiq **supports** a programme set by a physiotherapist or doctor. It does not diagnose, prescribe or replace them. I am using the MoSCoW approach and promising only the **Must have** items, and I will report honestly on what I achieve:

- a **safety layer** that filters exercises against the restrictions and avoid-lists for each condition, validates the AI's structured output and signposts red-flag symptoms to a GP or NHS 111;
- **rehabilitation support:** editable programme templates for five areas (lower back, knee, shoulder, general post-surgery and Parkinson's), a pain diary and session logging, based on the data model in the database pack shared with me (Boucekoua, 2026);

- a **secure server-side proxy** so no API keys ship in the app, and **on-device SQLite** storage with export and delete, so health data stays on the user's phone;
- **accessibility work** on the core journeys (screen-reader labels, large text and targets, aiming for WCAG 2.2 AA);
- an **evaluation**: a safety benchmark of 160 synthetic personas (40 built around the five rehabilitation areas) and 494 scenarios (already generated, all synthetic and not clinical), physiotherapist review of a sample of outputs if I can recruit them, and usability testing with the System Usability Scale (SUS);
- a **project journal website** with dated weekly entries, an ethics and safety page, a references page and a demo of the web app using synthetic data only.

Could have (stretch, not promised): a voice coach (agreed in principle with my supervisor), a summary a user could share with their physiotherapist, and extra languages. These only start once the Must have items are finished and tested.

Won't have: diagnosis, medication advice, replacing a physiotherapist, claims of clinical effectiveness, and GP-record or NHS login integration. Food and weight tracking from the prototype stays as a supporting feature but is not part of the evaluation.

Software and hardware: Expo SDK 54, React Native, TypeScript, expo-sqlite, Vercel serverless functions, Anthropic Claude Haiku 4.5, Jest and GitHub Actions, Figma, my own laptop and one Android and one iOS phone. **Costs** (prices checked 1 October 2026, \$1 = £0.75): Anthropic API credits are essential, about £60-75 (Anthropic, 2026). An Apple Developer Program membership, about £75 a year (Apple, 2026), is only needed for iPhone test builds. Hosting, Expo and the data APIs use free tiers. **Total: about £60-75, or £135-150 with Apple.**

AI-use declaration: Claude (Anthropic) helped me draft documents, generate the synthetic data and review code.

4. Rationale for Project Choice

- **Real need and a bounded setting:** home exercise is central to physiotherapy and adherence is a known problem (Lang *et al.*, 2022), while professional guidance on AI use already exists to design against (CSP, 2025).
- **A research gap:** existing work either tests whether digital exercise helps (Nagel *et al.*, 2024) or shows LLMs can be unsafe (He *et al.*, 2026). I want to measure whether a rule-based safety layer reduces that risk.
- **Supervisor guidance:** my supervisor advised that physiotherapy is the right setting, and that it fits the rehabilitation data model in the database pack shared with me (Bouhekoua, 2026).
- **Professional values and fit:** the BCS Code of Conduct asks members to protect public health and promote equal access to IT (BCS, 2022). The project builds on my SQL, systems analysis and security modules and my web development and digital transformation work.

- **My own reason:**

5. Background Research

Evidence so far shows a need and a safety problem. Digital exercise tools can raise adherence in the short term, but the review authors rate the evidence low to moderate quality, measured mostly by self-report, with longer-term benefit uncertain (Lang *et al.*, 2022). For a rule-based exercise app, 20 physiotherapists rated 944 recommended exercises and three agreed that 14 were risky, although the authors say expert consensus needs further validation (Griefahn *et al.*, 2024). For LLMs, a scoping review found evaluation methods highly varied (Lai *et al.*, 2025), and a systematic review found LLM plans worse than human experts in 5 of 6 head-to-head trials and unsafe as stand-alone replacements for expert supervision (He *et al.*, 2026). Co-design with disabled people is recommended (Haley *et al.*, 2026). Studies of general chatbots give mixed results: 55.8% of low back pain recommendations were accurate (Scaff *et al.*, 2025), and expert-rated ChatGPT-4.1 rehabilitation programmes scored lowest after ACL reconstruction (Cali *et al.*, 2026). In a conference study, local models scored high on safety yet mislabelled exercise intensity in 90.6% of plans, and the authors call for rule-based checks (de Mendonça, Sousa and Coelho, 2026). These studies use small samples and general-purpose chatbots, so they suggest where to be careful rather than prove how an app would behave. So prompting alone is not enough, and users and physiotherapists must be involved.

Similar products. Accessercise offers impairment-filtered exercise videos (Inside Indoor, no date) and Budy, launched in April 2026, offers AI voice coaching and seated programmes (Budy Labs, 2026), but I found no published safety evidence for either, and I have not tested them. I have not yet reviewed physiotherapy exercise-programme apps and will include them in my competitor analysis. **My prototype audit** (about 6,400 lines of TypeScript) found API keys in the client bundle (OWASP Gen AI Security Project, 2025), safety enforced only through prompt wording, no tests and no consent or delete controls for health data (ICO, no date). **Limits:** Physiq needs a smartphone, does not replace supervised care, and my rehabilitation content comes from a teaching database, so it must be reviewed by a physiotherapist before I call it safe. Synthetic personas test my rules, not real people.

6. Areas for Investigation

- **Home exercise adherence and digital rehabilitation**, and what the evidence does and does not show.
- **Condition-specific exercise safety and red flags**, so every rule can be traced to a source, such as NICE guidance for the conditions covered.
- **LLM guardrails and evaluation:** output validation, prompt injection and benchmark design (Huyen, 2024).

- **Accessible and user-centred design:** WCAG 2.2 and React Native accessibility (W3C, 2024).
- **Competitor apps in depth**, including physiotherapy programme apps, Accessercise and Budy.
- **Technologies and their limits:** Expo, on-device SQLite (still alpha on the web in SDK 54), a serverless proxy and a small LLM.
- **Evaluation methods:** SUS, think-aloud testing and thematic analysis (Rogers, Sharp and Preece, 2023).

Legal, social, ethical and professional issues (LSEPIs). *Legal:* UK data protection, because health data is special category data (ICO, no date), and the medical device boundary, which depends on an app's intended purpose (MHRA, 2023). *Social:* the digital divide, because users without a smartphone are excluded. *Professional:* the CSP principles on human oversight, consent and accountability (CSP, 2025) and the BCS, ACM, IEEE and AAAI codes. *Security:* API keys, prompt injection and protecting stored health data. *Net Zero Computing:* a small model, caching and fewer requests. *Deployment and maintenance:* how the app would be released and updated. *Project risks:* Section 10.

7. Research Ethics

I need **full School Ethics Committee approval**, because I will involve people, some of them disabled or living with long-term conditions, and possibly physiotherapists, and build a system that handles health data. **Status: not yet submitted.** No participant data will be collected before approval. Main safeguards:

- accessible information sheets and consent, with the right to withdraw and the right to decline the AI features (CSP, 2025);
- adults with capacity only, who judge the app and are not asked to exercise or change their own programme;
- fictional personas instead of real medical details, with only synthetic data sent to the AI, and the web demo uses only synthetic data;
- physiotherapists rate anonymised outputs only, and if approval or recruitment is late I run the benchmark first.

Physiq will state clearly that it is educational support and not medical advice. This follows the BCS duty to protect public health, privacy and wellbeing (BCS, 2022) and ACM principles 1.2 and 1.6 (ACM, 2018).

8. Review of Reference Materials

Source	Why I am using it
Books. Huyen (2024) <i>AI engineering</i>	Current practitioner text on evaluating and guarding LLM applications.
Rogers, Sharp and Preece (2023) <i>Interaction design</i> , 6th edn	The standard HCI textbook. Just over three years old, but it is the current edition and the seminal text on user evaluation.

Source	Why I am using it
Lu <i>et al.</i> (2023) <i>Responsible AI</i>	Patterns for trustworthy AI that I will map to the safety layer. Just over three years old and still the current edition.
Papers. Lang <i>et al.</i> (2022), <i>Archives of Physiotherapy</i>	Systematic review of digital tools and home exercise adherence. Over three years old, but it is the most directly relevant review, and Nagel <i>et al.</i> (2024) supports it.
Nagel <i>et al.</i> (2024), <i>JMIR mHealth and uHealth</i> ; Griefahn <i>et al.</i> (2024), <i>Scientific Reports</i>	A review of digital exercise for musculoskeletal conditions, and a study where physiotherapists rated app exercises for risk, which is the model for my expert review.
He <i>et al.</i> (2026), <i>Biology of Sport</i> ; Lai <i>et al.</i> (2025), <i>Journal of Medical Internet Research</i>	The LLM exercise safety review that justifies the safety layer, and a scoping review that shapes my evaluation. Haley <i>et al.</i> (2026) is opinion, not trial evidence.
Scaff <i>et al.</i> (2025), <i>Annals of the Rheumatic Diseases</i> ; Cali <i>et al.</i> (2026), <i>Frontiers in Medicine</i> ; de Mendonça, Sousa and Coelho (2026), <i>IEEE SeGAH</i>	Recent studies of chatbot and LLM accuracy and safety for back pain, rehabilitation programmes and exercise plans. All have small samples and I have so far read the abstracts of the first two, so I use only what they report.
Lewis and Sauro (2018), <i>Journal of Usability Studies</i>	Explains where the SUS average of 68 comes from and why it is a relative benchmark. I checked it against the source.
Websites. CSP (2025); WHO (2024); Arthritis UK (2026)	The physiotherapy profession's AI principles, and the scale of need for rehabilitation and arthritis. The Arthritis UK figure covers arthritis only.
W3C (2024); ICO (no date); MHRA (2023); Activity Alliance (2026)	Accessibility standard, health data rules, the medical device boundary (March 2023, over three years old but still the regulator's published guidance on intended purpose), and the disability inactivity gap.

Professional bodies. I will apply the ethics codes of BCS (2022), ACM (2018), IEEE (2020) and AAI (2019), including a human check on AI output.

9. Methodology

Research methods. I will use mixed methods: a short online questionnaire for adults who have followed a home exercise programme (aiming for 15-20 responses, minimum 10) to see how common needs are, and 3 to 5 interviews analysed thematically to understand why. Numbers alone would miss the reasons, and interviews alone would not show how common a need is. The drawback is small samples, so findings will be indicative only.

Development. I will use Scrum adapted for one person (Schwaber and Sutherland, 2020), with two-week sprints and a prioritised backlog. I rejected Waterfall because requirements are uncertain and testing would come too late, Kanban because it has no time-boxing for fixed deadlines, and Extreme Programming because pair programming is not possible alone, although I will borrow its test-first habit for the safety layer.

Evaluation. I will run the benchmark with and without the safety layer and report the unsafe-output rate. Following Griefahn *et al.* (2024), I will ask as many physiotherapists as I can recruit (aiming for three or more) to rate a sample of

outputs as risk or not risk and to review the red-flag list. If none is available, I will report the rule-based checks only and say so. I will run think-aloud tests with 5 to 8 people, report the mean SUS and compare it with the average of 68 from industrial studies, treating that as a relative benchmark and not a pass mark (Lewis and Sauro, 2018), and check core journeys with TalkBack and VoiceOver. I set reporting goals, not guaranteed results. Finally I will critically evaluate whether my objectives were met and reflect on my methodology and learning.

10. Project Plan

The module is 600 hours, about 24 a week over 25 weeks, so I front-load research, ethics and design into Semester 1 and keep Semester 2 for building, testing and writing. Dates to be checked against Blackboard.

Stage	Dates	Length	Deadlines and other work
Planning and proposal	28 Sep to 30 Oct 2026	5 weeks	Final proposal Fri 30 Oct, 16:00 (5%)
Research, analysis and design	2 Nov to 18 Dec 2026	7 weeks	Poster upload Mon 7 Dec; presentation weeks 11-12 (5%); <i>[other module hand-ins]</i>
Implementation and testing	7 Dec 2026 to 17 Mar 2027	about 12 weeks without the break	Artefact and walkthrough video Thu 18 Mar, 16:00 (40%); <i>[other module hand-ins]</i>
Critical evaluation and write-up	19 Mar to 26 Apr 2027	about 6 weeks	ExpoTees Wed 21 Apr (10%); final report Mon 26 Apr, 16:00 (40%)

Testing continues throughout, and the project journal is updated weekly. Figure 1 is the Gantt chart, also supplied as an Excel file. **Dependencies and risks.** Each stage depends on the one before: the safety layer cannot be evaluated until the proxy and data store work, and no participant data is collected before ethics approval.

Physiotherapist review of the red-flag list is planned for December to January so rules are reviewed before they are tested. Stretch work sits in the weeks before the March deadline as my buffer, so if the core runs late it is dropped first. **Other modules.**

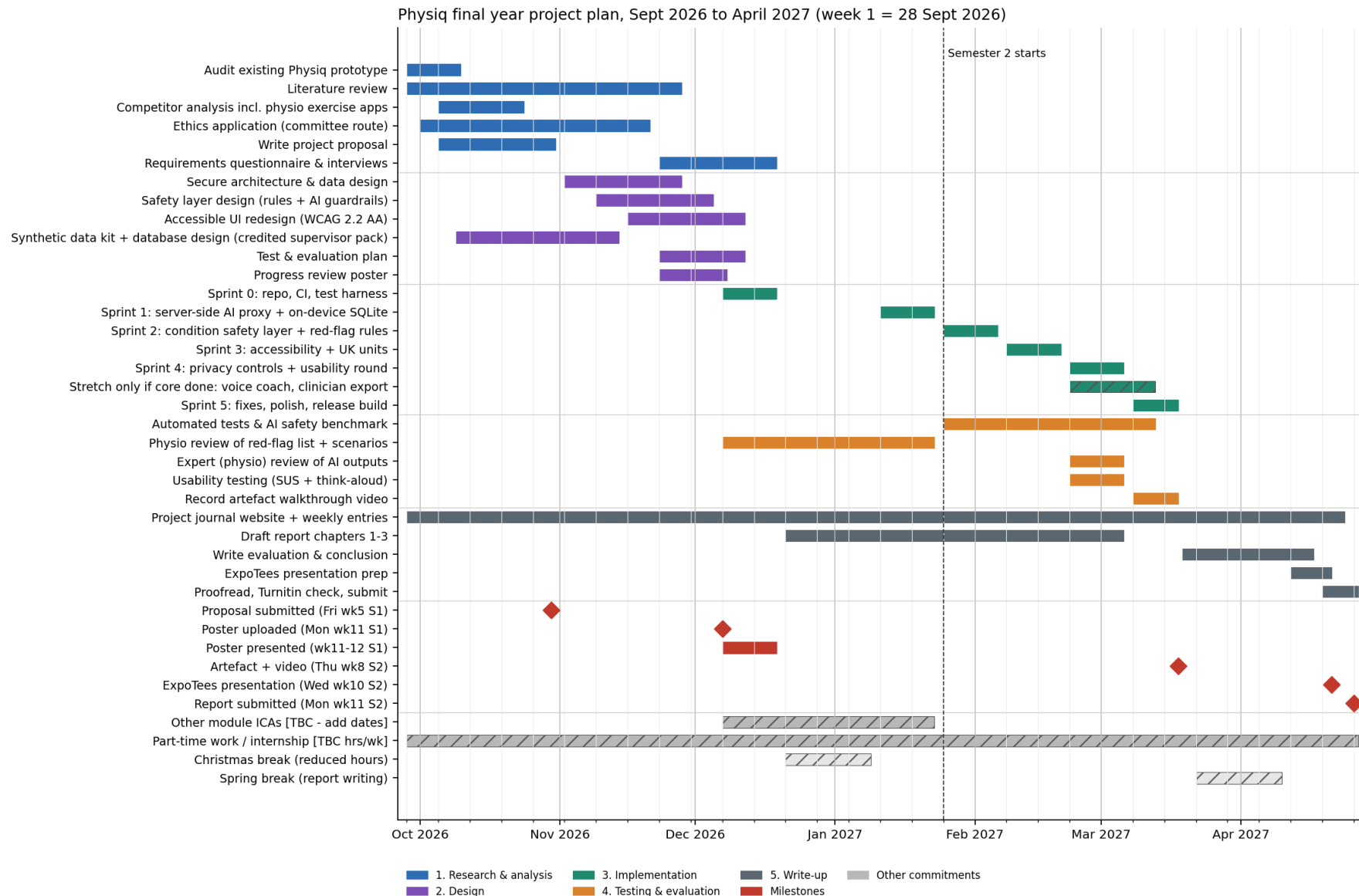


Figure 1: Project Gantt chart, September 2026 to April 2027. Hatched bars are other commitments (dates and hours to be added) or optional stretch work; diamonds are assessment deadlines.

11. References

- AAAI (2019) *AAAI Code of Professional Ethics and Conduct*. Available at: <https://aaai.org/about-aaai/ethics-and-plurality/> (Accessed: 1 October 2026).
- ACM (2018) *ACM Code of Ethics and Professional Conduct*. Available at: <https://www.acm.org/code-of-ethics> (Accessed: 14 October 2026).
- Activity Alliance (2026) *Our response to Sport England's Active Lives 2024-25 report*. Available at: <https://www.activityalliance.org.uk/news/10563-our-response-to-sport-englands-active-lives-202425-report> (Accessed: 14 October 2026).
- Anthropic (2026) *Pricing*. Available at: <https://platform.claude.com/docs/en/about-claude/pricing> (Accessed: 14 October 2026).
- Apple (2026) *Become a member: Apple Developer Program*. Available at: <https://developer.apple.com/programs/enroll/> (Accessed: 14 October 2026).
- Arthritis UK (2026) *The state of musculoskeletal health*. Available at: <https://www.arthritis-uk.org/policy-and-data/health-intelligence/the-state-of-musculoskeletal-health> (Accessed: 15 October 2026).
- BCS (2022) *BCS Code of Conduct*. Available at: <https://www.bcs.org/membership-and-registrations/become-a-member/bcs-code-of-conduct/> (Accessed: 14 October 2026).
- Boucekoua, M. (2026) *mHealth_Physiq_RawDB_V0_5 default SQL pack (Hotfix3)* [SQL Server database scripts]. Unpublished teaching resource, created 15 June 2026 and shared with me for use in my project. [Confirm with your supervisor how they want this cited.]
- Budy Labs (2026) *Budy launches AI fitness app with adaptive training, nutrition, and accessibility-first design* [Press release]. Fitt Insider, 23 April. Available at: <https://insider.fitt.co/press-release/budy-launches-ai-fitness-platform-with-adaptive-training-nutrition-and-accessibility-first-design/> (Accessed: 14 October 2026).
- Cali, A., Yorukoglu, M.E., Acar, G. and Safran, E. (2026) 'ChatGPT-generated rehabilitation programs in sports physiotherapy: an expert evaluation and a mixed-methods study of clinical applicability', *Frontiers in Medicine*, 13, 1935702. Available at: <https://doi.org/10.3389/fmed.2026.1935702>.
- CSP (2025) *Artificial intelligence (AI) principles in action*. Chartered Society of Physiotherapy. Available at: <https://www.csp.org.uk/professional-clinical/professional-guidance/ai-use-physiotherapy> (Accessed: 15 October 2026).
- de Mendonça, A.M.P., Sousa, F.P. and Coelho, I.M. (2026) 'Evaluating the safety of local LLM-generated exercise prescriptions against clinical guidelines', *2026 IEEE Conference on Serious Games and Applications for Health (SeGAH)*, Niteroi, Brazil, 29-31 July. IEEE. Available at: <https://doi.org/10.1109/SEGAH69272.2026.11653312>.
- Griefahn, A., Zalpour, C. and Luedtke, K. (2024) 'Identifying the risk of exercises, recommended by an artificial intelligence for patients with musculoskeletal disorders', *Scientific Reports*, 14, 14472. Available at: <https://doi.org/10.1038/s41598-024-65016-1>.
- Haley, J.A., Bodde, A., Helsel, B.C. and Lee, D. (2026) 'Beyond accessibility: co-designing mHealth to bridge the physical activity gap for people with disabilities', *mHealth*, 12, 15. Available at: <https://doi.org/10.21037/mhealth-2026-0003>.
- He, T., Lu, D., Ma, Y., He, J., Li, D., Li, G. and Sun, J. (2026) 'The AI recommendation paradox: a systematic review evaluating the promise, peril, and path forward for large language models in exercise recommendation', *Biology of Sport*, 43, pp. 949–970. Available at: <https://doi.org/10.5114/biolsport.2026.158676>.
- Huyen, C. (2024) *AI engineering: building applications with foundation models*. Sebastopol, CA: O'Reilly Media.

- ICO (no date) *What is special category data?* Available at: <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/lawful-basis/special-category-data/what-is-special-category-data/> (Accessed: 15 October 2026).
- IEEE (2020) *IEEE Code of Ethics* (IEEE Policies, Section 7.8). Available at: <https://www.ieee.org/about/corporate/governance/p7-8> (Accessed: 1 October 2026).
- Inside Indoor (no date) *Introducing Accessercise: the accessible exercise app*. Available at: <https://insideindoor.com/introducing-accessercise-the-accessible-exercise-app/> (Accessed: 14 October 2026).
- Lai, X., Lai, Y., Chen, J., Huang, S., Gao, Q. and Huang, C. (2025) 'Evaluation strategies for large language model-based models in exercise and health coaching: scoping review', *Journal of Medical Internet Research*, 27(1), e79217. Available at: <https://doi.org/10.2196/79217>.
- Lang, S., McLelland, C., MacDonald, D. and Hamilton, D.F. (2022) 'Do digital interventions increase adherence to home exercise rehabilitation? A systematic review of randomised controlled trials', *Archives of Physiotherapy*, 12, 24. Available at: <https://doi.org/10.1186/s40945-022-00148-z>.
- Lewis, J.R. and Sauro, J. (2018) 'Item benchmarks for the System Usability Scale', *Journal of Usability Studies*, 13(3), pp. 158-167.
- Lu, Q., Zhu, L., Whittle, J. and Xu, X. (2023) *Responsible AI: best practices for creating trustworthy AI systems*. Boston, MA: Addison-Wesley Professional.
- MHRA (2023) *Crafting an intended purpose in the context of Software as a Medical Device (SaMD)*. Available at: <https://www.gov.uk/government/publications/crafting-an-intended-purpose-in-the-context-of-software-as-a-medical-device-samd> (Accessed: 14 October 2026).
- Nagel, J., Wegener, F., Grim, C. and Hoppe, M.W. (2024) 'Effects of digital physical health exercises on musculoskeletal diseases: systematic review with best-evidence synthesis', *JMIR mHealth and uHealth*, 12(1), e50616. Available at: <https://doi.org/10.2196/50616>.
- OWASP Gen AI Security Project (2025) *LLM01:2025 Prompt injection*. Available at: <https://genai.owasp.org/llmrisk/llm01-prompt-injection/> (Accessed: 14 October 2026).
- Rogers, Y., Sharp, H. and Preece, J. (2023) *Interaction design: beyond human-computer interaction*. 6th edn. Hoboken, NJ: Wiley.
- Scaff, S.P.S., Reis, F.J.J., Ferreira, G.E., Jacob, M.F. and Saragiotto, B.T. (2025) 'Assessing the performance of AI chatbots in answering patients' common questions about low back pain', *Annals of the Rheumatic Diseases*, 84(1), pp. 143-149. Available at: <https://doi.org/10.1136/ard-2024-226202>.
- Schwaber, K. and Sutherland, J. (2020) *The Scrum guide*. Available at: <https://scrumguides.org/scrum-guide.html> (Accessed: 14 October 2026).
- W3C (2024) *Web Content Accessibility Guidelines (WCAG) 2.2*. Available at: <https://www.w3.org/TR/WCAG22/> (Accessed: 14 October 2026).
- WHO (2024) *Rehabilitation* [Fact sheet]. World Health Organization. Available at: <https://www.who.int/news-room/fact-sheets/detail/rehabilitation> (Accessed: 15 October 2026).