

# Bridging Sensor Data and Human Comprehension: Designing a Large Language Model-Based Coaching Interface for Runners

Wearable sensors have significantly changed how runners track their performance, capturing a wealth of biomechanical data that was once confined to specialised laboratories. However, this has created a "data deluge," where recreational runners are often overwhelmed by complex metrics like ground contact time, hip flexion and joint angles, struggling to translate this information into actionable improvements for their training [1, 2]. This gap between data collection and human comprehension limits the potential of wearable technology to enhance performance and prevent injury. This research addresses this challenge by exploring the question: *How can a coaching interface be effectively designed to translate complex running sensor data into personalised, comprehensible, and actionable feedback for runners by leveraging Large Language Models (LLMs)?*

To answer this question, the study employed a user-centred Double Diamond design process [3], which integrated an extensive literature review, market analysis, and semi-structured interviews with runners of varying experience levels. These initial discovery phases defined the core user needs and technical requirements that informed the subsequent design and development. The process culminated in the creation and evaluation of a high-fidelity, fully functional prototype system.

Based on these insights, a socio-technical system was designed and developed, which consists of a robust backend architecture and an intuitive frontend interface. The back-end architecture is a robust, Multi-Stage Processing Pipeline designed for control, reliability, and verifiability, and was a deliberate architectural choice that prioritises reliability and control over more unpredictable agent-based models. This pipeline ingests raw sensor data and uses a combination of specialised LLMs, and uses Retrieval-Augmented Generation (RAG) to ground the LLM's responses in a curated knowledge base of sports science, preventing the generation of inaccurate or "hallucinated" advice [4]. The system also incorporates deterministic validation checks and external fact-checking APIs to ensure that all recommendations are based on verifiable evidence, prioritising user safety and trust. The frontend prototype was designed to make these complex backend processes transparent and accessible. It employs a progressive disclosure strategy, presenting users with high-level summaries first, such as a daily "Coach's Briefing" that synthesises key metrics into a natural-language narrative, while allowing them to drill down into more granular data as desired. This approach avoids overwhelming the user, catering to both novices seeking simple guidance and data-savvy runners who desire deeper analysis.

The prototype underwent a rigorous mixed-method evaluation, including an expert heuristic review and a task-based usability study with eight representative users. The results were highly positive,

with the system achieving an average System Usability Scale (SUS) score of 86.8, which is considered 'excellent' and well above the industry benchmark for good usability [5]. The qualitative findings provided deeper insight, revealing three pillars that are fundamental to creating a successful and trustworthy LLM-powered coaching experience:

1. Radical transparency that fosters trust, as the system was designed to explicitly reveal its reasoning process, allowing users to see the LLM's "chain-of-thought" and the specific data points or literature informing its advice. Participants consistently cited this transparency as the primary reason for their confidence in the LLM's recommendations, confirming that for high-stakes applications like health and fitness, a "black box" approach is insufficient [6, 7].
2. Synergy between visuals and conversation creates actionable insight, as the prototype's most effective feature was its ability to combine interactive data visualisations (such as the body map and dynamic charts) with conversational, LLM-generated explanations. This synergy was critical in bridging the comprehension gap, enabling users to not only see a metric but to understand its context, its implications, and what to do about it.
3. Granular control provides a sense of partnership, as the interface allows users to tailor the coaching experience to their preferences, from selecting a coach's personality and tone to choosing between a "Flash" mode for quick, exploratory queries and a "Deepthink" mode for slower, evidence-rich analysis. This level of control transformed the user's relationship with the LLM from one of passive consumption to an active, collaborative partnership.

In conclusion, this research contributes a validated framework for designing LLM-driven coaching technologies that effectively bridge the gap between complex sensor data and human understanding. It demonstrates that the solution lies not in the LLM alone, but in a carefully orchestrated system that pairs a controllable, fact-grounded back-end with a transparent, interactive, and customisable front-end. By doing so, such systems can move beyond simply presenting data and begin to empower individuals with genuine, actionable insight for their athletic journey.

## References

[1] L. Xiang, A. Wang, Y. Gu, L. Zhao, V. Shim, and J. Fernandez, "Recent Machine learning progress in lower limb running biomechanics with wearable technology: A Systematic review," *Frontiers in Neurorobotics*, vol. 16, Jun. 2022, doi: 10.3389/fnbot.2022.913052.

[2] C. A. Clermont, L. Duffett-Leger, B. A. Hettinga, and R. Ferber, "Runners' perspectives on 'Smart' wearable technology and its use for preventing injury," *International Journal of Human-Computer Interaction*, vol. 36, no. 1, pp. 31–40, Mar. 2019, doi: 10.1080/10447318.2019.1597575.

[3] "Implementation of double Diamond design model for user interface of Bandung City Dashboard," in *IEEE Conference Publication | IEEE Xplore*. [Online]. Available: <https://ieeexplore.ieee.org/document/10427040>

[4] R. Browne et al., "Reflective Dialogues with a Humanoid Robot Integrated with an LLM and a Curated NLU System for Positive Behavioural Change in Older Adults," *Electronics*, vol. 13, no. 22, p. 4364, Nov. 2024, doi: 10.3390/electronics13224364.

[5] P. Laubheimer, "Beyond the NPS: Measuring Perceived Usability with the SUS, NASA-TLX, and the Single Ease Question After Tasks and Usability Tests," *Nielsen Norman Group*, Mar. 03, 2025. [Online]. Available: <https://www.nngroup.com/articles/measuring-perceived-usability/>

[6] H. Muijlwijk, M. C. Willemsen, B. Smyth, and W. A. IJsselsteijn, "Benefits of Human-AI Interaction for Expert Users Interacting with Prediction Models: a Study on Marathon Running," in *Proc. 29th Int. Conf. Intelligent User Interfaces (IUI '24)*, Greenville, SC, USA, 2024, pp. 245–258, doi: 10.1145/3640543.3645205.

[7] I. Ali, "AI transparency and explainability," *ResearchGate*, Dec. 2024, [Online]. Available: [https://www.researchgate.net/publication/386416207\\_AI\\_Transparency\\_and\\_Explainability](https://www.researchgate.net/publication/386416207_AI_Transparency_and_Explainability)