DOCTORAL CONSORTIUM

Closing the Loop in AI-Driven Athletic Training Prescription

LUKA LAH, MIRJANA KLJAJIĆ BORŠTNAR
University of Maribor, Faculty of Organizational Sciences, Kranj, Slovenia
luka.lah5@um.si, mirjana.kljajic@um.si

DOI
https://doi.org/10.18690/um.fov.4.2026.82

ISBN
978-961-299-152-4

Training planning and load management in competitive athletics remain largely unsystematic, with most coaches and athletes relying on static tools that generate plans without any mechanism for ongoing adaptation. This open-loop limitation fails to account for the athlete's continuously evolving physiological state, increasing the risk of overtraining and injury, particularly among athletes without access to expensive professional infrastructure. This paper proposes a closed-loop decision-support architecture that integrates annual training plan generation with daily adaptation using consumer-grade wearable data and subjective athlete input. The proposed system could combine knowledge graph plan generation, Bayesian adaptation, longitudinal individual learning, and generative adversarial refinement into unified architecture. Using Design Science Research as the overarching methodological strategy, complemented by a PRISMA-ScR scoping review as its empirical foundation, the research aims to evaluate whether such a system could produce prescription quality comparable to that of experienced human coaches.

Keywords:
closed-loop training system,
artificial intelligence in sports,
dynamic training plan adaptation,
wearable device data,
athlete load management

University of Maribor Press

1 Introduction

The digitalization of athletic training and the widespread availability of wearable monitoring devices have fundamentally transformed the conditions for planning and executing training processes over the past decade. Continuously measuring physiological parameters such as heart rate, heart rate variability, sleep quality, and physical activity, these devices generate extensive data streams that potentially enable highly individualized and evidence-based load management for competitive athletes. In parallel, the rapid commercial growth of training platforms based on artificial intelligence has produced broadly accessible solutions for automated training plan generation (Mateus et al., 2024; Zhou et al., 2025).

Despite this progress, training planning and load management in competitive athletics remain largely unsystematic in practice. Coaches and athletes most commonly rely on paper-based records or general-purpose productivity tools, where digital solutions are adopted, these typically consist of static template-based applications that generate a training plan from a predefined dataset without any mechanism for ongoing adaptation. This approach is insufficient because it fails to account for the athlete's evolving functional state, including fatigue, sleep quality, and acute load responses, thereby increasing the risk of overtraining, suboptimal training stimuli, and delayed responses to changes in the athlete's condition (Puce et al., 2025). The consequences of this gap are not negligible: injury rates in competitive and professional sport remain persistently high, with a substantial proportion of injuries attributable to inadequate load monitoring and insufficient recovery management (Mateus et al., 2024; Zhou et al., 2025).

While real-time feedback-driven systems do exist, they are predominantly built on Internet of Things (IoT) architectures that require expensive sensor infrastructure, dedicated communication networks, and continuous technical maintenance. Such systems are feasible within elite professional environments but remain inaccessible to most athletes who train under budget constraints, including amateur competitors, youth athletes, and recreational sport participants. This creates a structural inequality: the athletes who could benefit most from systematic load management, precisely because they lack dedicated coaching staff, are also those for whom current technology-driven solutions are practically out of reach.

The root cause of this insufficiency lies in the fragmented data landscape surrounding the athlete. Wearable device data (heart rate, sleep metrics, training load), subjective athlete reports (fatigue, pain, mood), and long-term training goals exist in isolated silos, trapped within proprietary ecosystems without standardization or integration mechanisms. As a result, coaches and athletes who interact with AI-based tools must still manually enter textual descriptions rather than connecting real-time data feeds, effectively preserving the open-loop limitation of current systems (Puce et al., 2025; Qi et al., 2024).

This practical problem is compounded by a scientific gap: the empirical literature remains methodologically heterogeneous, comparative studies are scarce, and no validated framework yet exists for dynamic, feedback-driven training plan adaptation (Qi et al., 2024; Zhao et al., 2025). Rigorous evaluation of AI-based training platforms has not kept pace with their commercial expansion, leaving their effectiveness relative to classical periodization largely untested (Mateus et al., 2024; Zhou et al., 2025). Issues of data standardization, model explainability, and governance of autonomous systems in competitive sport also remain inadequately addressed in the literature (Qi et al., 2024; Zhao et al., 2025).

This problem is theoretically grounded in cybernetic systems theory, which holds that the effectiveness of a regulatory system increases with the quality and speed of feedback (Wiener, 1948). Within athletic periodization, this aligns with the individualized training model (Bompa & Haff, 2009), which requires adjusting training loads according to the athlete's current functional status. Puce et al. (2025) explicitly identify the development of hybrid LLM and wearable approaches for dynamic adaptation as a priority for future research, directly framing the gap this paper addresses.

The proposed research therefore responds to a concrete, practical problem: the inability to reliably and continuously adapt training prescriptions based on real-time athlete data in a way that is accessible beyond high-budget professional environments. It investigates whether and how a closed-loop architecture integrating Large Language Models (LLMs) with consumer-grade wearable device data and subjective athlete feedback can serve as a viable and accessible decision support concept. The absence of such validated architecture is not the problem itself, but the proposed direction for its resolution.

This paper presents a research proposition for the design, development, and evaluation of AI closed-loop training system. The following research questions guide the investigation:

- **RQ1:** Which artificial intelligence and machine learning systems have been developed for generating athletic training plans, and what are their documented capabilities within sports science contexts?
- **RQ2:** How does the quality of training plans generated by AI systems compare to those created by experienced human coaches, and to what extent do model type, level of input detail, and athletic context influence this quality?
- **RQ3:** How effective are machine learning systems that incorporate wearable device data and subjective athlete wellness measures in predicting fatigue and optimizing load management among competitive athletes?
- **RQ4:** Does an AI system that generates an annual training plan and adapts it daily based on wearable device data and athlete feedback produce higher quality training prescriptions and more effective load management than classical non-adaptive periodization methods?

AI systems in the sports domain rest on the assumption that artificial intelligence can generate high-quality personalized training plans and effectively manage athlete loads. Within the existing scientific literature, this assumption has not been sufficiently tested (Mateus et al., 2024; Puce et al., 2025).

Qi et al., (2024); Zhao et al. (2025) point out that questions of data standardization, model explainability, and the governance of autonomous systems in competitive sports contexts receive inadequate attention in literature. Puce et al. (2025) explicitly argues that future research should prioritize reinforcement learning models or hybrid approaches combining large language models with wearable device data for dynamic training plan adaptation. Without a synthesis of available evidence, developers of decision-support solutions are left without an empirical foundation on which to base architectural decisions.

1.1 Prior research

The existing body of prior research can be organized into three mutually complementary clusters: systems for generating and adapting training plans within a closed-loop architecture, the evaluation of Large Language Model (LLM) quality in training plan generation, and machine learning systems for athlete load management.

Three prior studies directly address closed-loop architectures that combine training plan generation with dynamic sensor-driven adaptation. Using deep learning and Internet of Things (IoT) sensors, Wang et al. (2026) achieved an 18 to 23 percent improvement in athletic performance with the CoachXNet system compared to systems without individualization. Razo-Ballon et al. (2025) proposed architecture integrating IoT infrastructure with the Llama-2-7b large language model to produce daily adapted training plans based on physiological and environmental data collected from smartwatches, representing the closest existing technological design to the problem addressed in this research. By combining reinforcement learning with an inertial measurement unit (IMU), electromyography, and heart rate monitoring. Li (2025) developed a system for competitive aerobics athletes that demonstrated approximately 16 percent average improvement in training performance and athlete satisfaction exceeding 90 percent.

Several comparative studies have evaluated training plan quality produced by large language models. Genç et al. (2025) found statistically significant quality differences across ChatGPT-3.5, ChatGPT-4o, and ChatGPT-4.1, with scores increasing progressively across versions. Havers et al. (2025) had domain experts rate muscle hypertrophy and strength plans, with Google Gemini receiving the highest ratings, followed by Copilot and GPT-3.5. J. Yang et al. (2026) evaluated six large language models on soccer training plan quality, with ChatGPT-o1 scoring highest among expert evaluators. Washif et al. (2024) benchmarked three ChatGPT versions against established resistance training guidelines, concluding that generated plans are realistic but frequently require adjustment and are best treated as coach templates. Düking et al. (2024) showed that input detail significantly influences output quality in runner-targeted plans, with the least detailed input receiving significantly lower expert ratings on 15 of 22 criteria compared to the most detailed. Z. Yang et al. (2026) corroborated this finding, demonstrating that detailed input reduces variability and raises overall quality across dimensions of personal adaptability, safety, and feasibility. Tan & Chen (2025) reported that a GAN-based approach achieved a 22 percent reduction in MSE, and 45 percent faster generation compared to classical machine learning, with higher subjective ratings from athletes and coaches. Masagca (2024) confirmed in a quasi-experimental study that a five-week AI-generated calisthenics program produced significant improvements in flexibility and muscular endurance, though cardiovascular outcomes remained inferior to those achieved through professionally designed programs.

Empirical studies on machine learning-supported load management and physiological monitoring form a closely related body of evidence. Simonelli et al. (2026) achieved subjective fatigue prediction accuracy of 70 to 82 percent in professional soccer players using four machine learning models combined with wearable sensor data. Carvalho et al. (2025) demonstrated through explainable artificial intelligence that heart rate variability effectively tracks daily load changes in high-performance swimmers, with two predictive models performing well at a 27 percent increase in training load and a 20 percent increase in perceived effort. Teixeira et al. (2024) found that maturation offset was a statistically significant but weak predictor of high-intensity actions in youth soccer, with all tested algorithms showing poor performance when used as the sole predictor. Mandorino et al. (2024), in a two-season longitudinal study with 106 professional soccer players, showed that large weekly load decreases improved locomotor efficiency in less fit athletes, while both large decreases and large increases were detrimental in well-conditioned players. Hao & Qian (2024) proposed a support vector machine-based performance evaluation model designed to provide multi-dimensional decision support for training planning.

2 Proposed methodology and results

To achieve the objectives of the proposed research, we will adopt Design Science Research (DSR) as the overarching methodological strategy (figure 2). Hevner et al. (2004) and Peffers et al. (2007) recommend DSR as the primary strategy for developing and evaluating novel information artifacts within the context of information systems research. The framework is methodologically appropriate here because the proposed research does not merely describe or explain an existing phenomenon. Rather, it seeks to design, construct, and evaluate a new technological artifact that addresses a clearly identified gap in scientific literature.

Several conceptual directions could plausibly address the identified problem of personalized sport-specific training prescription. One direction could involve rule-based expert systems, in which domain knowledge from sports scientists and coaches is encoded as explicit logical rules. Another direction could rely on collaborative filtering, recommending training content based on patterns observed across large populations of athletes with similar profiles. A third direction could center on retrieval-augmented generation, pairing data mining approaches with structured databases of validated training plans. Each of these approaches carries distinct trade-offs in terms of explainability, data requirements, and adaptability to individual athletes. The model proposed in this research draws from the strengths of each direction while addressing their individual limitations through a multi-phase architecture that combines semantic embedding, knowledge graph construction, Bayesian adaptation, and generative refinement.

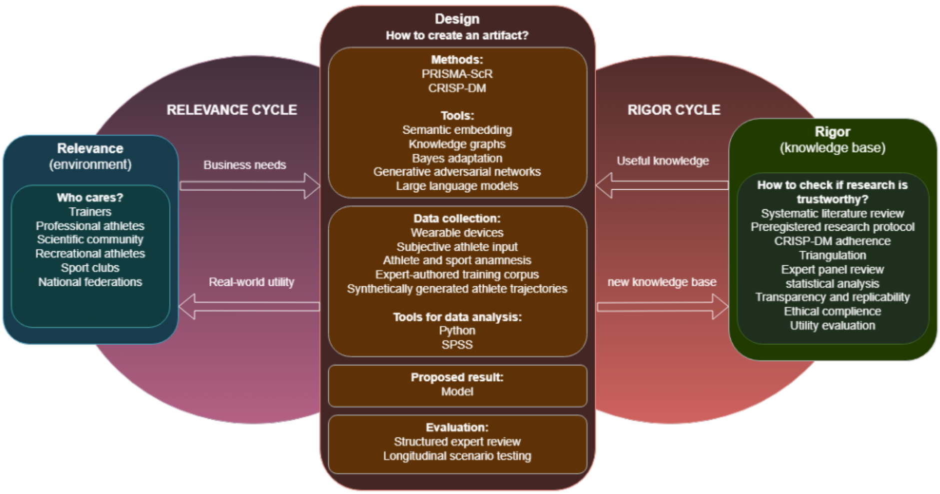

Figure 2: Design Science Research Framework applied to the closed-loop AI training system (adapted from Hevner et al., 2004)

Source: Own

The DSR framework is complemented by a systematic scoping review conducted in accordance with the PRISMA-ScR protocol (Tricco et al., 2018). This review serves as the foundational step of the research, ensuring that the artifact is grounded in the best available evidence regarding existing methodologies, data sources, and algorithmic approaches in the field. Within the Design cycle, the development of all machine learning components will follow the CRISP-DM process model. This means the development proceeds through the stages of business understanding, data understanding, data preparation, modeling, evaluation, and deployment in an iterative manner. DSR therefore functions as overarching research logic, while CRISP-DM provides the structured procedural scaffold governing each machine learning development iteration within that broader framework.

Expected contributions and evaluation

Personalized conditioning training presents a multi-dimensional optimization problem that resists straightforward algorithmic solutions. The core difficulty lies in simultaneously representing the complexity of individual athlete profiles and encoding the structured demands of diverse sports in a form that is both computationally tractable and scientifically grounded. Existing approaches each address a part of this challenge: rule-based expert systems struggle to scale and capture individual variability, collaborative filtering is limited by cold-start problems and cross sport generalization failure, and retrieval-augmented generation raises concerns about prescriptive consistency and traceability (Tan & Chen, 2025; Washif et al., 2024). No existing system has addressed these limitations as an integrated architecture.

The proposed model could respond to this gap through a closed-loop architecture combining four computational mechanisms: knowledge-graph-based plan generation, Bayesian closed-loop adaptation, longitudinal individual learning, and generative adversarial refinement. The system would address exclusively the conditioning dimension of athletic preparation, deliberately excluding technique-oriented content so that technical coaching authority remains with human practitioners.

A foundational challenge concerns the computational representation of athlete identity and sport structure. An athlete profile could encode medical history, injury records, prior performance assessments, training age, competition level, and contextual constraints such as equipment access. A sport profile could encode dominant movement patterns, energy system demands, and competitive calendar structure. Both representations could be independently embedded into high-dimensional vectors and blended into a unified athlete-sport fingerprint, whose weighting between sport-general and athlete-specific dimensions could function as an adjustable parameter reflecting athlete experience level. The model could then construct a knowledge graph in which nodes represent exercises and athlete and sport characteristics, with edges encoding biomechanical similarity and prescriptive relationships derived from a corpus of expert-authored plans. This formulation would enable relevance-scored exercise selection and structured plan assembly according to established periodization principles (Bompa & Haff, 2009).

A second challenge concerns temporal validity. Any plan generated at a single point in time will progressively diverge from an athlete's actual physiological state, as fatigue, recovery, and adaptation continuously alter the appropriate training stimulus (Carvalho et al., 2025; Simonelli et al., 2026). The proposed adaptation mechanism could model this as a Bayesian inference problem, in which the system maintains and continuously updates a probabilistic belief over prescription parameters. At each observation interval, this belief could be updated using objective wearable signals such as heart rate variability, sleep quality, and activity load, together with a subjective wellness rating from the athlete. Adjustable parameters could include training volume, intensity, exercise selection, and recovery duration, all constrained within the periodization structure established during plan generation.

A third challenge is that population-level models cannot account for pronounced inter-individual variability in physiological response. Two athletes with identical intake profiles may respond very differently to the same prescription, and this divergence compounds over time (Mandorino et al., 2024). The proposed longitudinal learning component could address this by progressively shifting the model's generative assumptions away from corpus-derived population priors and toward the specific response history of the individual. In early use, the model would rely primarily on population patterns; as personal response data accumulates, it could assign progressively greater weight to that individual's own trajectory. The rate and conditions under which individual data should override population priors represent an open methodological challenge that would constitute an original contribution of this research. Initial development and stress-testing of this component could rely on synthetically generated athlete trajectories produced by a generative adversarial network conditioned on corpus-level patterns, before iterative retraining on real-world observations.

A fourth challenge concerns the structural coherence of generated output. Even when individual exercise parameters are appropriate, a computationally assembled plan may lack the contextual flow and compositional logic characteristic of expert-authored programs, a limitation that Bayesian parameter optimization alone cannot resolve (Tan & Chen, 2025). A generative adversarial component introduced as the final architectural layer could address this: a generator would combine the Bayesian adaptation output with the athlete-sport embedding to produce candidate plans, while a discriminator would evaluate whether each candidate is structurally indistinguishable from expert-authored programs. Iterative adversarial training could progressively refine the generator's outputs toward expert quality, with the Wasserstein formulation with gradient penalty applied to improve training stability and output diversity.

Evaluating a decision-support model of this nature presents challenges not fully addressed by conventional software testing, since prescription quality is not reducible to a single metric and expert judgment remains an indispensable reference point. A two-stage evaluation strategy is proposed within the DSR evaluation cycle (Hevner et al., 2004). In the first stage, domain experts could assess model outputs against structured criteria derived from sports science literature, covering progressive overload adherence, periodization integrity, sport-specific appropriateness, and practical feasibility. In the second stage, the dynamic behavior of the adaptation mechanism could be examined across representative longitudinal scenarios defined by distinct combinations of wearable biometric trajectories and subjective wellness ratings, testing whether the Bayesian update mechanism produces responses that are both directionally correct and proportionate over time. Semi-structured interviews with coaches and sports scientists could provide qualitative triangulation alongside quantitative performance metrics.

The primary anticipated contribution of this research is a functioning decision-support artifact integrating the described mechanisms into a unified closed-loop architecture (Hevner et al., 2004; Peffers et al., 2007). The scientific contribution lies not only in the artifact itself but in the knowledge design generated through its construction and evaluation, as the proposed combination of knowledge graph generation, Bayesian adaptation, longitudinal learning, and generative refinement has not been examined as an integrated system in the existing literature. The research could yield a meaningful contribution regardless of empirical outcome. If expert evaluators judge generated prescriptions as comparable to coach-authored plans, this would constitute evidence that architecture is a viable computational approach to a problem previously considered to require exclusively human expertise. If generated plans prove inferior, the evaluation process could identify the specific architectural or data-related limitations responsible, providing a theoretically grounded foundation for future development.

References

Bompa, T. O., & Haff, G. (2009). *Periodization: Theory and Methodology of Training*. Human Kinetics.

Carvalho, D. D., Goethel, M. F., Erblang, M., Vilas-Boas, J. P., Pyne, D. B., Fernandes, R. J., & Lopes, P. (2025). Impact of an Overload Period on Heart Rate Variability, Sleep Quality, Motivation, and Performance in High-level Swimmers: Use of Explainable Artificial Intelligence (XAI) to Assess Training Load Variations. *Sports Medicine*. https://doi.org/10.1007/s40279-025-02369-1

Düking, P., Sperlich, B., Voigt, L., Hooren, B. V., Zanini, M., & Zinner, C. (2024). ChatGPT Generated Training Plans for Runners are not Rated Optimal by Coaching Experts, but Increase in Quality with Additional Input Information. *Journal of Sports Science and Medicine*, 56–72. https://doi.org/10.52082/jssm.2024.56

Genç, A., Aydın, G. R., Kasap, M., Özkan, A., Rakhymzhanov, A., Gürcan, H. H., Güllü, S., & Demirhan, B. (2025). Comparative evaluation of ChatGPT versions in training program design: Scientific approach, accuracy, and practical applicability. *BMC Sports Science, Medicine and Rehabilitation*, *18*(1), 19. https://doi.org/10.1186/s13102-025-01409-7

Hao, P., & Qian, K. (2024). The Integration of Personalized Training Program Design and Information Technology for Athletes. *Scalable Computing: Practice and Experience*, *25*(5), 4351–4359. https://doi.org/10.12694/scpe.v25i5.3083

Havers, T., Jelonnek, C., Masur, L., Isenmann, E., Sperlich, B., Geisler, S., & Düking, P. (2025). A professional assessment of training plans for muscle hypertrophy and maximal strength developed by generative artificial intelligence. *Biology of Sport*, *42*(4), 353–380. https://doi.org/10.5114/biolsport.2026.152350

Hevner, A., R, A., March, S., T, S., Park, Park, J., Ram, & Sudha. (2004). Design Science in Information Systems Research. *Management Information Systems Quarterly*, *28*, 75.

Li, H. (2025). Intelligent IoT-based training decision and evaluation system for competitive aerobics athletes. *Discover Computing*, *28*(1), 182. https://doi.org/10.1007/s10791-025-09709-9

Mandorino, M., Tessitore, A., & Lacome, M. (2024). Loading or Unloading? This Is the Question! A Multi-Season Study in Professional Football Players. *Sports*, *12*(6), 148. https://doi.org/10.3390/sports12060148

Masagca, R. C. (2024). The AI coach: A 5-week AI-generated calisthenics training program on health-related physical fitness components of untrained collegiate students. *Journal of Human Sport and Exercise*, *20*, 39–56. https://doi.org/10.55860/13v7e679

Mateus, N., Abade, E., Coutinho, D., Gómez, M.-Á., Peñas, C. L., & Sampaio, J. (2024). Empowering the Sports Scientist with Artificial Intelligence in Training, Performance, and Health Management. *Sensors*, *25*(1), 139. https://doi.org/10.3390/s25010139

Peffers, K., Tuunanen, T., Rothenberger, M., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*, 45–77.

Puce, L., Bragazzi, N. L., Currà, A., & Trompetto, C. (2025). Harnessing Generative Artificial Intelligence for Exercise and Training Prescription: Applications and Implications in Sports and Physical Activity—A Systematic Literature Review. *Applied Sciences*, *15*(7), 3497. https://doi.org/10.3390/app15073497

Qi, Y., Sajadi, S. M., Baghaei, S., Rezaei, R., & Li, W. (2024). Digital technologies in sports: Opportunities, challenges, and strategies for safeguarding athlete wellbeing and competitive integrity in the digital era. *Technology in Society*, *77*, 102496. https://doi.org/10.1016/j.techsoc.2024.102496

Razo-Ballon, H., Ticona-Esquivel, R., & Montalvo, P. (2025). Design of IoT Architecture and LLM Model for Personalized Training Recommendations for Athletes. In T. Guarda, F. Portela, & G. Gatica (Eds), *Advanced Research in Technologies, Information, Innovation and Sustainability* (pp. 306–320). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-83210-9_23

Simonelli, C., Trecroci, A., Formenti, D., & Rossi, A. (2026). Prediction of Subjective Fatigue in Professional Soccer Players: A Data-Driven Method to Optimize Training Approach to the Match. *Research Quarterly for Exercise and Sport*, *97*(1), 137–145. https://doi.org/10.1080/02701367.2025.2560562

Tan, J., & Chen, J. (2025). Generating context-specific sports training plans by combining generative adversarial networks. *PLOS ONE*, *20*(1), e0318321. https://doi.org/10.1371/journal.pone.0318321

Teixeira, J. E., Encarnação, S., Branquinho, L., Morgans, R., Afonso, P., Rocha, J., Graça, F., Barbosa, T. M., Monteiro, A. M., Ferraz, R., & Forte, P. (2024). Data Mining Paths for Standard Weekly Training Load in Sub-Elite Young Football Players: A Machine Learning Approach. *Journal of Functional Morphology and Kinesiology*, *9*(3), 114. https://doi.org/10.3390/jfmk9030114

Tricco, A., Lillie, E., Zarin, W., O'Brien, K., Colquhoun, H., Levac, D., Moher, D., Peters, M., Weeks, L., Hempel, S., Akl, E., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M., Garritty, C., & Straus, S. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, *169*. https://doi.org/10.7326/M18-0850

Wang, Y., Feng, X., Wu, Y., Zhang, X., & Wu, J. (2026). CoachXNet: An Artificial Intelligence and Internet of Things Integrated Platform for Personalized Training and Feedback in Digital Sports. *International Journal of Computational Intelligence Systems*, *19*(1), 49. https://doi.org/10.1007/s44196-025-01146-2

Washif, J., Pagaduan, J., James, C., Dergaa, I., & Beaven, C. (2024). Artificial intelligence in sport: Exploring the potential of usingChatGPT in resistance training prescription. *Biology of Sport*, *41*(2), 209–220. https://doi.org/10.5114/biolsport.2024.132987

Wiener, N. (1948). Cybernetics. *Scientific American*, *179*(5), 14–19.

Yang, J., Qin, S., & Ren, D. (2026). Artificial intelligence coaches on the sidelines: Evaluating readability and quality of soccer training plans from six generative models. *International Journal of Sports Science & Coaching*, *21*(1), 75–81. https://doi.org/10.1177/17479541251369593

Yang, Z., Zhang, X., Li, H., & Ye, J. (2026). More details, less variability? A crossover design study on the impact of information granularity on ChatGPT's training program stability. *Biology of Sport*. https://doi.org/10.5114/biolsport.2026.154148

Zhao, X., Zheng, H., & Zhang, Q. (2025). Modern Advances in Artificial Intelligence Across the Athletic Domain. *Concurrency and Computation: Practice and Experience*, *37*. https://doi.org/10.1002/cpe.70068

Zhou, D., Keogh, J. W. L., Ma, Y., Tong, R. K. Y., Khan, A. R., & Jennings, N. R. (2025). Artificial intelligence in sport: A narrative review of applications, challenges and future trends. *Journal of Sports Sciences*, 1–16. https://doi.org/10.1080/02640414.2025.2518694