If Artificial Intelligence is Writing Your Report, Why Are You Required? Defending the CRAFT of Performance Support in the Ever-Growing Age of Automated Analysis

Shane Malone, [1 2] **Martin Buchheit** [3 4 5 6 7]

[1]Technological University Dublin, Tallaght Campus
[2]LJMU, Liverpool, UK
[3]Aspetar Orthopaedic and Sports Medicine Hospital, Doha, Qatar
[4]Type 3.2 Performance
[5]HIIT Science, Revelstoke, Canada
[6] INSEP, Paris, France
[7]Optimo Performance Centre

Artificial Intelligence | **Large Language Models** | **Knowledge Translation** | **Automation Bias** | **Interpersonal Skills** | **Team Sports**

Headline

Within this expert opinion piece, we argue that practitioners should not compete with artificial intelligence (AI) on the technical execution of our daily roles, because this is a battle we will lose. We need to realise that the value of performance support was never fully located in the report we present. It sits in the appraisal of whether the report describes what actually happened within the training process, in the framing of the report against a proper structure such as a monitoring quadrant (4), and in the negotiation with a head coach who might want something different. AI automation does not diminish those functions. We propose the CRAFT model as a practical structure for drafting performance reports in this new age of AI within performance support departments. This is a five-stage workflow specifying what should be delegated to an AI model, what can't be, and what fails when the two sides of this equation become confused.

Aim

We aim to extend our earlier work on the POCKET model (19) into a performance landscape in which the analytical layer of support is becoming increasingly driven by automated AI support systems. We describe which parts of the reporting workflow are being absorbed by generative AI models, explain why simply inserting AI into a decision-making workflow doesn't always improve it, and offer the CRAFT model to ensure the practitioner retains the functions that generate trust, influence, and accountability within the training process.

AI Moves the Value, but not the Practitioner

The direction of travel is not speculative. Reviews of AI and machine learning in sport describe rapid growth across analysis, injury risk assessment, training planning and talent identification, with methods moving steadily into commercial platforms (10, 11, 23). Large language models (LLM) already translate technical findings into plain language, build dashboards and run analyses that would traditionally take practitioners considerable time (21, 27). The clearest empirical sign comes from clinical practice, where randomised trials of documentation tools report reductions in documentation time, cognitive task load and burnout, though more modest than the marketing implied (18, 22). The technology has not eliminated the clinician, but it has decisively changed which part of the clinician's day carries the most value. The documentation stopped being the deliverable. The consultation became the deliverable. Sport science is on the same trajectory, and arguably further along, because the majority of the data we produce daily is already structured and AI-readable in a way that a patient conversation is not. This is not a threat to the profession, but a redistribution of where its value sits. The value now sits in the translation of the report to coaches and not the technical expertise to develop said report.

AI predictions must be translated into Training Process Adjustments

There is a persistent assumption in our field that better models will produce better decisions, but there is little evidence to support it. Systematic appraisal of musculoskeletal injury prediction models in sport found that the great majority carried high or unclear risk of bias (7, 28), and Bullock and colleagues (6, 8) argue that poorly validated methods should not be deployed in athlete care, because a model untested beyond its developmental sample generates confident but misleading outputs. An AI pipeline that computes a flawed construct faster simply distributes the flaw more efficiently (15); ultimately, working faster does not mean working better. Even a well-calibrated, externally validated AI model does not make the final performance decision. It produces an output that must be interpreted with context and argued in plain coach-based language in a room where a head coach may hold an opposing view and greater authority. Explainable AI has been proposed as the bridge, but Kranzinger and colleagues (17) report that such explanations are rarely validated before dissemination to the practitioners and coaches supposed to act on them. An explanation that satisfies a data scientist does not persuade a head coach on an MD-1; this is where the practitioner is still required to translate AI predictions into training process adjustments that move the dial from a performance angle.

The Synergy Problem: Adding a Human Element to AI can Sometimes Make Things Worse

Most people tend to treat human judgement with the addition of AI as superior to either alone (25). There is one place where the evidence supports this, as Vaccaro and colleagues (29) found consistent gains for human–AI combinations on content creation tasks. This is the basis for the position of the current opinion piece. These generative AI tools should be pointed at the drafting stage and back-end development of the report and kept away from the final decision-making process of the report, which should be human-led. Indeed, across the same analysis (29), human–AI combinations performed significantly worse than humans alone or AI alone, an effect driven entirely by the end-stage decision-making process, with competence a key determining factor of this: when the human outperformed the AI system, the combination of the two then improved both; where the system outperformed the human, it degraded the system. Synergy is therefore conditional on the human knowing something the model does not and being able to prompt the model correctly. If an MDT member cannot say from memory which player is managing a calf that week, inserting them between the model and the coach adds little value and may even degrade the final output.

AI Deskilling and the Erosion of Contextual Knowledge

Automation bias has been well documented in clinical decision support, producing both omission errors, where problems go unnoticed, and commission errors, where incorrect recommendations are accepted as fact (14). Recent work in medicine has extended this into a concern about deskilling, with early-career practitioners most exposed because they are still consolidating the judgement that the tool is replacing (24). AI deskilling in some areas may be seen as an advantageous development of AI, such as vibe coding, backend development and the speed of questioning and reasoning; however, deskilling within the applied role is a concern that needs to be addressed for example the junior practitioner in team sports such as rugby union who has never manually cleaned a GPS file may never develop the pattern recognition to question the reliability of data from units and, for example, when a velocity spike is present within training or match data, where a winger has been pushed over the try-line. Contextual knowledge is accumulated by being the practitioner in the room, but also by being present on the field of training or match and learning from these moments. This experience cannot be extracted from any data set or any AI-prompted conversation, because it was never in the model to begin with. A model can tell you that a player's high-speed running dropped 22%, the direction of the smallest worthwhile change, and the flag associated with this change is that the player is underloaded; only someone who has the contextual knowledge can tell you the why. If AI reduces the practitioner's need to be present, it erodes the very thing that makes them worth consulting in the training process.

You Cannot Delegate Confidentiality and Accountability

Some processes sit outside AI models' reach, however capable it may seem. The first is data governance. Athlete physiological and medical data attracts heightened protection as special category data under the General Data Protection Regulation (GDPR), and consent for continuous wearable-derived data is often formal rather than substantive, with contracts treating athlete data as an institutional asset (16). Pasting a squad's medical or wellness data into a general-purpose model, a common informal practice, is a governance issue whether acknowledged or not within the sports performance field (9). The second is accountability. A recommendation to withdraw a player from a fixture, or to escalate a fatigue concern, is a professional judgement attached to a named person who can be questioned, who can be wrong, and who carries the consequence. A report generated end-to-end by a model has no author in this sense. It is just a producer of an output; the AI producer cannot stand behind a recommendation, so it gives the coach nothing to trust. The practitioner who can author the judgement of the AI-drafted solution, deliver it and shape it based on the context of the environment can then own the decision and resolve any coach's scepticism in a way that a sophisticated AI model never will.

The CRAFT Model

What is missing from the models in our field is a structure that specifies the division of workload between the practitioner and AI at each stage of producing a report. Therefore, we have developed the CRAFT model (Figure 1): Context, Retrieval and Reduction, Appraisal, Framing, Transmission. The model's organising principle is simple. Delegate the stages where output needs to be confirmed, and error is visible. Retain the stages where judgement, context and accountability are required in the workflow.

C – Context

Nothing is placed into an AI model until every stakeholder has confirmed who the report is for. The context drives the report, not the other way around: which stakeholder receives the report, what they are trying to achieve this week, what they are biased toward, and what a realistic zone of possible agreement looks like are all questions that precede analysis and what can't be derived from the data (19). This is also where the practitioner decides what would change if the answer from the AI model comes back differently, and if nothing would change, the report should not be produced. Skipping this stage is the most common failure in AI reporting structures, which provides a flawless, cool-looking report with all the applied statistical analysis that answers a question no one is looking to answer and is ultimately useless (4).

R – Retrieval and Reduction

This is the appropriate home for automation; this is where AI is most effective. Cleaning, merging, aggregating, computing typical error, smallest worthwhile change, z-scores and magnitude-based inferences, and drafting first-pass visualisations are all high-volume, rule-governed, verifiable tasks. Delegating them to AI is not a compromise; it is the correct allocation of a practitioner's time. The practitioner's role at this stage is upstream specification, defining which metrics, thresholds and comparison windows the AI model should use and report on, rather than execution.

A – Appraisal

This is a high workload stage for the practitioner and the one most often skipped under the time constraints of daily tasks. This means checking the AI output against what was observed on the day: was that session what the plan said it was, was that player carrying an injury, was the drill a match simulation or an open game. It also means interrogating rather than accepting the AI model's output, given the documented tendency toward automation bias when practitioners are under time pressure (14, 24). Practically, we recommend a standing appraisal question that is asked daily of AI-generated reports. Before any report is released: which number in this document would I not be keen to have an error with, in front of the head coach, and have I verified it, before I release this report?

F – Framing

Framing is the step where the evidence supports genuine human–machine synergy (29), and, chronologically, it is the point in the reporting structure where AI can aid our process. Generative tools are effective at producing alternative phrasings, tightening prose, removing jargon, and proposing structure;

they are poor at knowing which two points matter this week. The practitioner selects the message and owns the words; the AI accelerates their construction, not their selection. The principles established for coach-facing communication are unchanged: filtered rather than exhaustive data, plain language rather than technical vocabulary, a clear dose–response narrative, and a small number of specific points rather than a comprehensive data dump (2, 5, 13).

T – Transmission

The conversation between the performance staff and the coach is still the most important step in performance management. Transmission covers the delivery, the negotiation, the concessions, the agreement and the review afterwards, and it remains entirely a human process. It is here that the POCKET sequence operates: a planned opening position, a communication style matched to the individual coach, genuine exploration of options, and a terminated conversation with a unified message to players irrespective of outcome (19). AI models can help prepare for this stage by drafting briefing notes or anticipating counter-arguments. It can not have the conversation for the practitioner. Table 1 sets out this division of labour, including a failure mode when each stage is assigned to the wrong party.

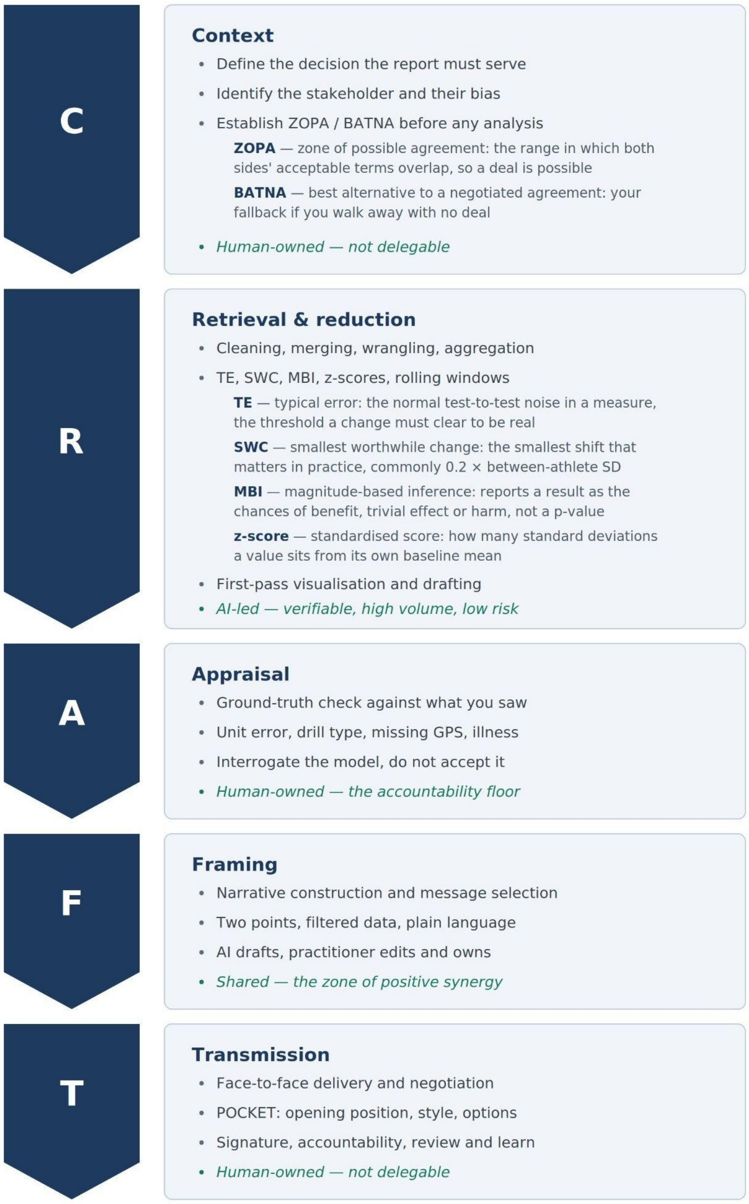

Fig. 1. The CRAFT model. A five-stage workflow for AI-assisted performance reporting. Ownership is stated at each stage: three stages are practitioner-owned, one is machine-led, and one is shared, and represents the accountability floor of the process.

Table 1. The CRAFT division of labour across the reporting workflow.

Stage	Owner	What the machine does well	What only the practitioner can do	Failure mode if mis-assigned
C — Context	Practitioner	Retrieve prior reports, precedents and season trends on request	Define the decision the report must serve, the stakeholder, their bias, and the Zone of possible agreement (ZOPA): the range in which both sides' acceptable terms overlap, so a deal is possible/Best alternative to a negotiated agreement (BATNA): this is your fallback if you walk away with no deal or your best alternative outcome	A technically flawless report answering a question nobody asked
R — Retrieval & Reduction	Machine	Clean, merge, aggregate; draft first-pass visuals, and compute the following: **Typical Error:** the normal test-to-test noise in a measure, the threshold a change must clear to be a real change, **Smallest Worthwhile Change:** the smallest shift in a measure that matters in practical terms, commonly 0.2 x between athlete standard deviation **Z-Scores:** How many standard deviations a measure shifts away from its own baseline mean **Magnitude-Based Inference:** Reports a measure as the chances of benefit, trivial effect or harm and not a p-value	Specify the metrics, thresholds and comparison windows the computation should use	Practitioner time consumed by wrangling that adds no decision value
A — Appraisal	Practitioner	Flag anomalies, outliers and missing data for human attention	Ground-truth the output against what was actually seen on the pitch that day	Automation bias: a plausible, fluent, confidently wrong report reaches the coach
F — Framing	Shared	Draft alternative narratives, tighten language, strip jargon, propose structure	Select the two points that matter, own the wording, carry the clinical meaning	Generic prose that reads as though it could describe any squad in any week
T — Transmission	Practitioner	Prepare briefing notes and anticipated counter-arguments before the meeting	Deliver, negotiate, read the room, concede, commit and sign the outcome	No accountable human in the room; trust and influence erode permanently

Some Questions Before Delegating to AI models

For practitioners looking for a usable heuristic that want to lean into AI modelling, we suggest four questions be applied to any task before it is handed to an AI generative model. The task should only be delegated if the answer to all four is satisfactory.

1. Does my judgement change the output? If the practitioner would accept the AI tool's result unchanged across each use case, they are not supervising the model; they are relying on it. In this case, the human is just a rubber stamp on the process, and a rubber stamp is the first thing a streamlined pipeline removes.
2. Would an error be visible? Some errors announce themselves; others are silent. A miscalculated total distance is visible. A player quietly dropped from a merge because of an identifier mismatch is not. Silent-failure tasks require

human checkpoints regardless of how reliable the tool appears.

3. Am I still accountable for it? If the answer is yes, and for anything reaching a coach or athlete it always is, then the practitioner must be able to explain and defend every element of the output, in their own words, without reference to the AI tool that produced it.
4. Would I be comfortable if this input were disclosed? If athlete-identifiable physiological or medical data is being passed to a third-party system, that comfort should be grounded in a documented data processing arrangement rather than in the absence of anyone having asked (16).

The Crossover

Underlying the CRAFT model is the process of our role (Figure 2). As AI analysis becomes more capable, the marginal value of the technical aspect of our role begins to fall, not because the work stops mattering, but because it stops being scarce, while the marginal value of contextual and interpersonal judgement rises with the volume of output requiring appraisal, framing and negotiation (26, 30). We argue that in well-resourced team sport environments these curves may already have met. A hiring process that filters primarily on programming ability and statistical technique is therefore selecting for the portion of the role with the steepest depreciation, and technical competence becomes a floor rather than a differentiator.

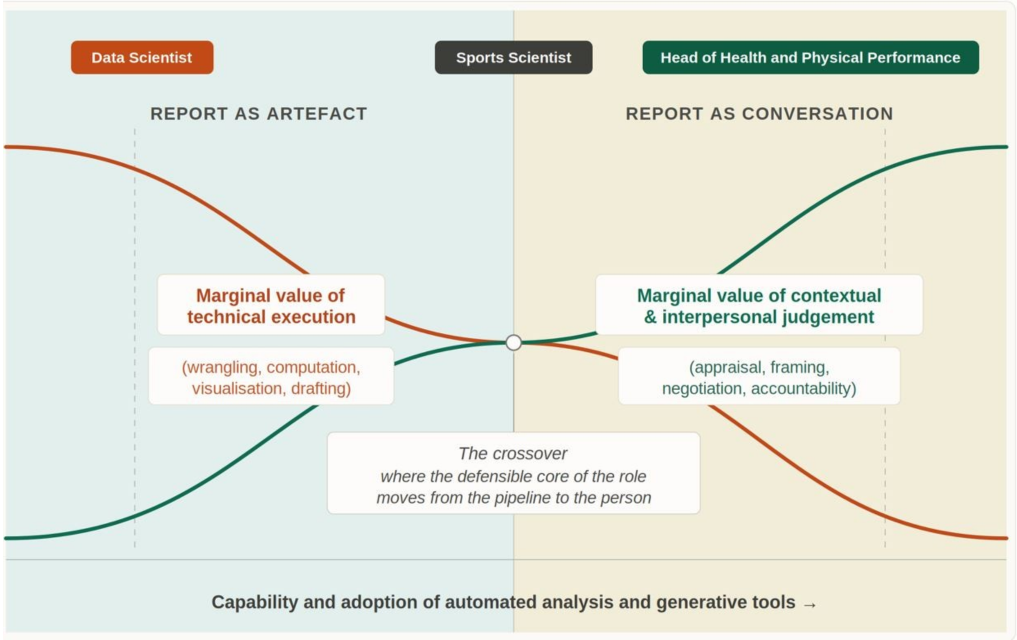

Fig. 2. The crossover curve. As automated analysis and generative tooling become more capable, the marginal value of technical execution declines while the marginal value of contextual appraisal, framing, negotiation and accountability rises. The defensible core of the practitioner's role migrates from the report as an artefact to the report as a conversation.

The Beautiful AI-Supported Report Still Fails at the Art of Performance Science

We have argued previously that effective performance support requires blending theoretical intelligence with emotional intelligence, and science with art, to arrive at appropriate performance outcomes (3,19). Automation does not change this argument. It sharpens it by removing the ambiguity about which half of the equation is required. The literature has consistently found that coaches acquire and act on knowledge through personal interaction rather than through documents, and that trust built over time is the primary determinant of whether evidence influences the training process (1, 12, 19, 20). No advance in generative AI capability alters this equation, because the constraint is never the availability of information. It is the willingness of a decision-maker to act on it, which is a function of the relationship between stakeholders in the performance process. A department that produces beautiful, automated reports and has no one able to hold a difficult conversation with the head coach has optimised the wrong variable. The most useful framing we can offer is the following. The machine has taken the part of the job that could be written down. What remains is the part that could not: knowing the squad, knowing the coach, knowing which number matters this week, and being willing to put a name to a recommendation and defend it. That was always the difficult part. It is now the main part of the performance process.

Conclusion

Automating the analytical layer is not a threat to applied sport science but an accelerator, and potentially one that puts more power in the practitioner's hands. This is conditional: the technology delivers only when applied appropriately, and only if we learn to appreciate its function rather than adopt it blindly. What it does remove is the option of defining our role through technical execution. Prediction modelling shows that better algorithms have not produced better decisions; human–machine collaboration shows that inserting a human into an automated workflow can reduce its usefulness unless that human holds context the system lacks; clinical practice shows that deskilling is a worrying consequence of AI models, not a hypothetical risk. All three point the same way. The practitioner's contribution lies in contextual appraisal, framing, negotiation and accountability, the domains which the proposed CRAFT model gives structure to. Departments that build their workflows around this division of labour can gain the AI model's productivity without diminishing the value of the human in the loop that the coaches require daily.

Take Home Messages

- As AI automation spreads, the value of technical execution falls, and the need for contextual judgement rises. Where a practitioner sits on that crossover, from data scientist to head of health and physical performance, increasingly predicts their exposure to automation.
- The technical layer of performance reporting, data wrangling, computation, visualisation and first-draft narrative is largely automatable and is being automated in most team sport settings. Practitioners should be delegating this process as it is a major strength of AI models.
- Better prediction does not produce better decision-making. Most injury prediction models in sport carry a high risk of bias, and even well-validated models do not perform the interpretation, framing and negotiation through which the training process becomes adjusted.
- Appraisal is the non-negotiable stage. Contextual knowledge, what happened at that session, cannot be recovered from data that never contained it.
- The CRAFT model (Context, Retrieval and Reduction, Appraisal, Framing, Transmission) assigns ownership at each stage of the reporting workflow and names the failure mode when a stage is mis-assigned.
- Recruitment weighted primarily toward technical ability is selecting practitioners who have the fastest-depreciating asset for a performance department. Technical competence is becoming a floor of practitioner requirements rather than a differentiator.

AI Use Statement

Aligning with the CRAFT structure presented here, the conceptual framework, literature screening, and primary drafting represent the independent work of SM. Generative AI models were employed strictly during the Framing phase to refine language fluency on the initial manuscript, with subsequent AI iterations utilized to condense and streamline specific sections following critical appraisal and feedback from MB. Figure 1 was developed using Microsoft PowerPoint, whereas Figure 2 was initially crafted via Microsoft PowerPoint and Canva prior to final visual editing with AI assistance. Generative AI support was facilitated through Claude Pro, specifically employing the Fable 5.1 and Opus 5 High architecture models.

References

1. Bartlett JD, Drust B. A framework for effective knowledge translation and performance delivery of sport scientists in professional sport. Eur J Sport Sci. 2021; 21(11):1579–1587.

2. Buchheit M. Want to see my report, coach? Sport science reporting in the real world. Aspetar Sports Med J. 2017; 6:36–43.

3. Buchheit M, Allen SV. To optimise, first empathize. Int J Sports Physiol Perform. 2022; 17(4):505–506.

4. Buchheit M, Hader K. Data everywhere, insight nowhere: a practical quadrant-based model for monitoring training load vs response in elite football. Sport Perform Sci Rep. 2025; #258, v1.

5. Buchheit M, Simpson BM. Player-tracking technology: half-full or half-empty glass? Int J Sports Physiol Perform. 2017; 12(Suppl 2):S235–S241.

6. Bullock GS, Hughes T, Arundale AH, Ward P, Collins GS, Kluzek S. Black box prediction methods in sports medicine deserve a red card for reckless practice: a change of tactics is needed to advance athlete care. Sports Med. 2022; 52(8):1729–1735.

7. Bullock GS, Mylott J, Hughes T, Nicholson KF, Riley RD, Collins GS. Just how confident can we be in predicting sports injuries? A systematic review of the methodological conduct and performance of existing musculoskeletal injury prediction models in sport. Sports Med. 2022; 52(10):2469–2482.

8. Bullock GS, Ward P, Collins GS, Hughes T, Impellizzeri FM. Comment on: machine learning for understanding and predicting injuries in football. Sports Med Open. 2024; 10:84.

9. Connor M, O'Neill M. Large language models in sport science and medicine: opportunities, risks and considerations. arXiv:2305.03851. 2023.

10. Cust EE, Sweeting AJ, Ball K, Robertson S. Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance. J Sports Sci. 2019; 37(5):568–600.

11. Gonçalves CE, Costa E. Editorial: training in sports — the role of artificial intelligence and machine learning. Front Sports Act Living. 2025; 7:1590162.

12. Fullagar HHK, McCall A, Impellizzeri FM, Favero T, Coutts AJ. The translation of sport science research to the field: a current opinion and overview on the perceptions of practitioners, researchers and coaches. Sports Med. 2019; 49:1817–1824.

13. Gregson W, Carling C, Gualtieri A, O'Brien J, Reilly P, Tavares F, Bonanno D, Lopez E, Marques J, Lolli L. A survey of organizational structure and operational practices of elite youth football academies and national federations from around the world. Front Sports Act Living. 2022; 4:1031721.

14. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. J Am Med Inform Assoc. 2012; 19(1):121–127.

15. Impellizzeri FM, Tenan MS, Kempton T, Novak A, Coutts AJ. Acute:chronic workload ratio: conceptual issues and fundamental pitfalls. Int J Sports Physiol Perform. 2020; 15(6):907–913.

16. Kwon JW. Athlete data sovereignty: addressing the legal and policy gaps in sports technology. Front Sports Act Living. 2025; 7:1742484.

17. Kranzinger S, Halmich C, Hofer D, Kranzinger C. A scoping review of explainable artificial intelligence in sports science. Discov Artif Intell. 2026; 6:5.

18. Lukac PJ, Turner W, Vangala S, Chin AT, Khalili J, Shih YT, Sarkisian C, Cheng EM, Mafi JN. Ambient AI scribes in clinical practice: a randomized trial. NEJM AI. 2025; 2(12).

19. Malone S, Buchheit M. Dropping into the POCKET: blending science and art to negotiate the appropriate performance outcomes in team sports environments. Sport Perform

Sci Rep. 2024; #229, v1.

20. Malone S, Hughes B, Collins K. Effect of training load distribution on aerobic fitness measures in hurling players. J Strength Cond Res. 2019; 33(3):825–830.

21. Musat CL, Mereuta C, Nechita A, Tutunaru D, Voipan AE, Voipan D, et al. Diagnostic applications of AI in sports: a comprehensive review of injury risk prediction methods. Diagnostics. 2024; 14(22):2516.

22. Olson KD, Meeker D, Troup M, Barker TD, Nguyen VH, Manders JB, Stults CD, Jones VG, Shah SD, Shah T, Schwamm LH. Use of ambient AI scribes to reduce administrative burden and professional burnout. JAMA Netw Open. 2025; 8(10):e2534976.

23. Reis FJ, Alaiti RK, Vallio CS, Hespanhol L. Artificial intelligence and machine-learning approaches in sports: concepts, applications, challenges, and future perspectives. Braz J Phys Ther. 2024; 28(3):101083.

24. El Tarhouny S, Farghaly A. Deskilling dilemma: brain over automation. Front Med. 2026; 13:1765692.

25. Robertson S. Man and machine: adaptive tools for the contemporary performance analyst. J Sports Sci. 2020; 38(18):2118–2126.

26. Ward P, Windt J, Kempton T. Business intelligence: how sport scientists can support organization decision making in professional sport. Int J Sports Physiol Perform. 2019; 14(4):544–546.

27. Van Eetvelde H, Mendonça LD, Ley C, Seil R, Tischer T. Machine learning methods in sport injury prediction and prevention: a systematic review. J Exp Orthop. 2021; 8:27.

28. Leckey C, van Dyk N, Doherty C, et al. Machine learning approaches to injury risk prediction in sport: a scoping review with evidence synthesis. Br J Sports Med. 2025; 59:491–500.

29. Vaccaro M, Almaatouq A, Malone T. When combinations of humans and AI are useful: a systematic review and meta-analysis. Nat Hum Behav. 2024; 8(12):2293–2303. doi:10.1038/s41562-024-02024-1.

30. Windt J, Taylor D, Little D, Sporer BC. Making everyone's job easier: how do data scientists fit as a critical member of integrated support teams? Br J Sports Med. 2021; 55(2):73–75.