Athleet.Ai

PERFORMANCE INTELLIGENCE
WHITEPAPER SERIES

PAPER TWO OF SEVEN · ELITE TRACK AND FIELD

The Intelligent Track

Artificial intelligence, event-specific expertise and the future of elite athletics coaching

Paul Forrest
Founder and Chief AI Officer, Athleet.AI
June 2025

About this series

This is the second of seven papers examining how artificial intelligence can be used responsibly in performance sport. Each paper stands alone. Together they set out a single argument, which is that the durable advantage in sport comes from better decisions rather than from more data, and that better decisions require systems designed around the person who is accountable for them.

The series shares one intellectual framework, described in Chapter 1, and one management thesis, described in Chapter 7. The flagship paper on governance, *The Responsible Performance System*, supplies the assurance model that the sport-specific papers draw on.

Editorial balance. Roughly seven parts in ten of this document is evidence and sector analysis, two parts in ten is original framework, and one part in ten concerns the Athleet.AI platform itself. Readers who want the argument without the product can stop at Chapter 14.

HOW TO READ THE EVIDENCE TAGS

Every substantive claim in this paper carries a tag stating what kind of evidence supports it. The tags are used consistently across the series.

PEER-REVIEWED A primary study in a peer-reviewed journal.

REVIEW A systematic review or meta-analysis.

CONSENSUS A consensus statement or governing-body position.

COMMENTARY Methodological criticism, which carries weight without being primary data.

PRACTICE Practitioner opinion, including the author's own.

ATHLEET.AI The author's own framework or analysis, proposed rather than proven.

Where the evidence is thin, the paper says so. Where it points the other way, the paper says that too.

Legal and regulatory notice. Parts of this paper describe legislation, regulatory guidance and governing-body requirements, in particular in Chapters 12 and 13. Nothing in this document constitutes legal advice. The provision of legal advice sits outside the terms of any engagement with the author or with Athleet.AI, and the material is presented to support discussion and further review by qualified advisers.

Statements about what a source says are attributed to that source. Where the author draws a practical implication the source does not itself state, the text says so. Legislative timetables in this area have moved more than once, and the position described is the position the author was able to verify in June 2025.

ILLUSTRATIVE MATERIAL

Every athlete, squad, programme and scenario in this paper is invented, and none describes a real athlete, coach or organisation. Cases are marked **illustrative scenario** wherever they appear. Published research is cited and numbered, and invented material is never presented as evidence.

CITATION

Forrest, P. (2025). *The Intelligent Track: Artificial Intelligence, Event-Specific Expertise and the Future of Elite Athletics Coaching.* Athleet.AI Performance Intelligence Whitepaper Series, Paper Two. Version 1.0, June 2025.

Comments and corrections are welcome and will be acknowledged in later versions.

ALSO IN THIS SERIES

- **Paper one.** *The Responsible Performance System.* Governance and trust in sports AI.
- **Paper three.** *Closing the Coaching Gap.* Responsible AI in grassroots athletics.
- **Paper four.** *Casting the Net Wider.* AI as a force multiplier in football pathways.
- **Paper five.** *The Connected Endurance Athlete.* Integrated preparation in triathlon.
- **Paper six.** *The Augmented Touchline.* Responsible AI in elite football.
- **Paper seven.** *The Augmented Golfer.* AI across technique, practice and strategy.

Contents

A NOTE FROM THE AUTHOR

Two problems that turned out to be one

I have spent my working life on two questions that took me an embarrassingly long time to recognise as the same question. The first is how a large organisation turns what it knows into what it does, which occupied me through business process work in the 1990s and through natural language and machine learning work since. The second is how a coach standing on a cold track on a Tuesday evening in February decides whether the athlete in front of them should do the session as written.

The literature on the first question is enormous. The literature on the second is thinner than most people outside the sport imagine, and it thins further the moment you step away from the 100m and the marathon. I coach in a track and field club as a volunteer, I hold coaching qualifications in sprints, hurdles and endurance, and I have watched enough good coaches make good decisions on incomplete evidence to be suspicious of any system that offers certainty in place of it.

What prompted this paper was a pattern I kept meeting. A programme would show me the data it had gathered, which was more than it had ever held before, and then describe decisions it was making with less confidence than it remembered having a decade earlier. Nobody in those rooms was short of numbers. What they lacked was a way of holding an athlete's history, their event, their staff's disagreements and their own methodology in one place, so that a decision could be made once, explained, and found again later.

That is the gap this paper addresses. It argues that athletics needs event-aware, explainable decision support built around the coach, and it argues equally firmly that several things being sold to the sport at present should be declined. Both halves matter. A paper that only enthuses is marketing, and a paper that only warns is of no use to anyone with a season to plan.

I have graded the evidence throughout because I would want that if I were reading it. Where I am giving an opinion, the text says so. Where the research contradicts a claim the industry likes to make, I have quoted the research rather than the claim.

TWO PAGES THAT CIRCULATE ON THEIR OWN

Executive summary

Elite athletics has acquired an impressive quantity of measurement and comparatively little shared understanding of what to do with it. Force plates, timing systems, markerless video, wearables and athlete management platforms have all improved. The decision they are meant to inform, which is whether this athlete should do this work today and why, is still made largely from memory, relationship and the coach's own record-keeping, and it is still very rarely written down in a form that survives a change of staff.

This paper argues that the opportunity in track and field is an explainable decision layer that is aware of the event, honest about uncertainty and designed to leave a named human being accountable. It argues against three things the market currently offers the sport, which are single-score readiness indices, individual injury prediction, and general-purpose training generators that have learned the shape of coaching advice without the substance of any particular event.

1

Athletics is a family of sports, and the evidence is unevenly distributed across it. Sprint and endurance research is comparatively rich. Throws, combined events and race walking are served far more thinly, and a model trained on the whole of published sports science will be confidently wrong most often exactly where coaches most need help.

2

Measurement has outrun interpretation. Markerless video now reports sagittal joint angles to within a few degrees of laboratory systems in gait, while transverse-plane rotations remain poor. Sprint force-velocity outputs are reproducible; the slope most often used to individualise training is not. The sport can measure a great deal that it cannot yet responsibly act on.

3

The dominant load metric does not survive its own statistics. A sustained methodological literature shows the acute to chronic workload ratio is mathematically coupled, that its apparent optimal zone can be an artefact of grouping, and that it produces the familiar association even when fed meaningless inputs. Load records remain useful. Load ratios as individual injury predictors do not.

4

Prediction is the wrong ambition; continuity is the right one. Prospective work in track and field reports that athletes who lose fewer weeks to injury and illness are more likely to meet their performance objectives. That association is a better foundation for a product than any prediction claim the current evidence supports.

5

Junior ranking tells a programme very little. Across Olympic sports, junior performance explains around two per cent of the reliable variance in senior performance. In athletics specifically, between six and 24 per cent of world-ranked juniors reach the equivalent senior ranking, depending on event and sex. Selection systems built on early results are discarding people for no good reason.

6

Explainability is a performance requirement, not a compliance chore. A recommendation a coach cannot interrogate will be ignored by the good ones and followed uncritically by the inexperienced, and both outcomes are bad. Provenance, stated confidence, visible assumptions, real alternatives and a recorded human approval are the minimum.

What follows for a high-performance programme

The practical recommendation is narrower than the technology permits and more useful than it promises. Programmes should connect their records before buying any intelligence, since a system reasoning across four incompatible spreadsheets will reason confidently about the wrong athlete. Every recommendation should name its sources, state its confidence and offer alternatives. And a named person should sit between any recommendation and any athlete, with that person's decision recorded alongside it, because a decision record is the only asset in this field that appreciates.

Pilots should also measure the right thing. Availability, completed against planned work, time-loss episodes, progression against individual objectives and season continuity are all measurable within a season. Career length is not, and any supplier claiming to have shortened that feedback loop is describing a hope rather than a result.

Chapter 14 sets out a maturity model and a 90-day proof of value. The distance from a filing cabinet to a connected record is analytical, and most programmes can cross it in a season. The distance from a connected record to an organisation that learns faster than its staff turn over is a management problem, and it is the harder of the two by some margin.

CHAPTER ONE

Why track and field needs an architecture of its own

A general-purpose model speaks about athletics with great fluency and no event knowledge. The gap between those two things is the whole opportunity, and it is also the whole risk.

Ask a general-purpose language model to write a training week for a hammer thrower and it will produce something plausible, orderly and subtly wrong. The volumes will look reasonable. The sequence will resemble a week of distance running with heavier implements attached, because the model has learned the statistical shape of training advice rather than the demands of the throw. A coach reads it and knows within 20 seconds that the author has never stood in a circle. An athlete reads it and cannot tell.

That asymmetry is the reason this paper exists. The useful question in elite athletics is narrower than the one the technology sector keeps asking. What deserves examination is which parts of the coaching task can be made faster, more consistent and better documented, and which parts must stay with the person whose licence, reputation and relationship with the athlete are attached to the outcome.

THE CORE CLAIM

Athletics needs event-aware, explainable decision support built around the coach. The market is currently offering it event-blind, unexplained automation instead.

A current-opinion paper in sports medicine states the commercial problem plainly. Developers of proprietary prediction models are often unwilling to report, or to release, how those models were developed and validated, and that reticence prevents any independent evaluation of performance, interpretability or generalisability before a system is used on athletes.[60] COMMENTARY No pub-

lished validation of an AI or language-model coaching tool in elite track and field was found during the preparation of this paper, and readers should treat that absence as the state of the field rather than as an oversight.

Three constraints shape everything that follows. Evidence in athletics is unevenly distributed, and thin in exactly the events where coaches most need help. Measurement has improved faster than interpretation, so the sport can now record a great many quantities it has no validated basis for acting on. And the accountable decision-maker is a named human being, which means any system that obscures how a suggestion was reached has made that person's job harder rather than easier.

1.1 The cycle this series uses

Every paper in this series describes the same five-stage cycle, because it is the sequence a coaching decision actually follows whether or not anybody has written it down. Information is gathered, converted into insight, weighed as a decision, delivered as work, and then compared against what happened. Most programmes do the first and fourth stages well. The second, third and fifth are where institutional knowledge either accumulates or evaporates.

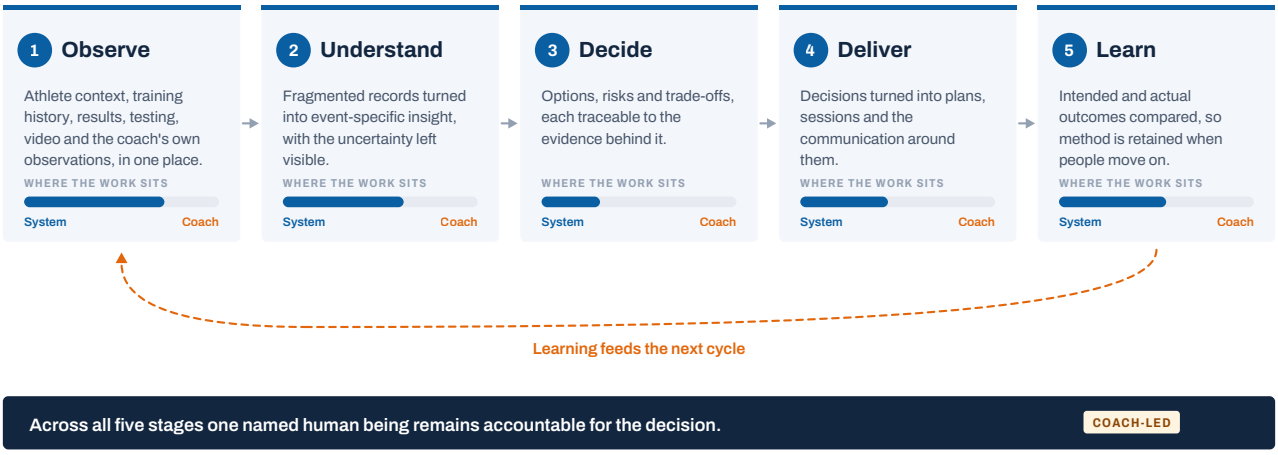


FIGURE 1 The performance intelligence cycle. The bar beneath each stage shows the balance of work between system and coach, which shifts deliberately towards the human at the point of decision.

Athleet.AI framework, proposed rather than validated.

The proportions in that diagram carry the argument. A system earns its place in observation, where it can hold more than a person can, and in learning, where it can compare intention against outcome without becoming bored. At the moment of decision its contribution falls sharply, and it should. Anyone selling a product whose contribution rises at that point is selling something the evidence does not support and the sport should not buy.

1.2 What the sport is actually short of

Programmes rarely lack data. They lack a place where an athlete's history, their event, the disagreements between their support staff and the squad's own methodology can sit together, so that a decision made in March can be found and understood in September. When a coach leaves, the reasoning leaves with them. When a physiotherapist changes, the rationale for a modified session becomes a memory held by two people who now recall it differently.

THE ARGUMENT IN ONE PARAGRAPH

Athletics does not need a machine that decides. It needs a machine that remembers precisely, reasons within the event, shows its working, and leaves the decision where the accountability already sits. Everything else in this paper is an elaboration of that sentence, including the parts that explain why several popular products fail it.

CHAPTER TWO

One sport, many performance models

The distance between a 400m hurdler and a shot putter is greater than the distance between many separate sports. Any system that treats athletics as a single domain will fail hardest where the coaching problem is most difficult.

Athletics shares a stadium, a governing body and a set of rules. Beyond that the events diverge sharply in the physiological system they tax, the technical model they demand, the competition structure they face and, most consequentially for anyone building software, the quality of the research available to the coach. A sprinter's training rests on a substantial body of work on force production and speed development. A decathlete's rests mainly on the accumulated judgement of the few people who have done it well.

That asymmetry has a practical consequence the technology sector consistently underestimates. A model trained across published sports science inherits the distribution of that literature, and the literature is heavily weighted towards running. Asked about throwing, it interpolates. The output remains fluent, the citations remain plausible, and the coach least able to detect the error is the one most likely to act on it.

WHERE MODELS FAIL

Generic systems fail most where evidence is thinnest, because that is where they must invent. The events with the least research therefore carry the greatest risk of confident error.

2.1 Reading the matrix

Figure 2 sets out my own assessment of where each event family stands on three dimensions, which are the strength of the published evidence, the quality of routine measurement available, and how well a general model is likely to perform. The grading is a judgement rather than a measurement, and I have published it in this form precisely so that coaches who disagree can say where and why.

EVENT-SPECIFIC INTELLIGENCE MATRIX

EVENT FAMILY	DOMINANT DEMAND	EVIDENCE	MEASUREMENT	MODEL FIT	THE PRACTICAL CONSTRAINT
Sprints and relays	Maximal power, horizontal force, acceleration mechanics	●●●●	●●●●	●●●○	Rich literature, dense measurement, short competition exposure.
Hurdles	Rhythm, stride pattern, unilateral tolerance	●●○○	●●●○	●●●○	Technical model interacts with fatigue in ways few models capture.
Middle distance	Aerobic and anaerobic balance, race tactics	●●●○	●●●●	●●●●	Physiological testing well established; tactical modelling is not.
Endurance and marathon	Aerobic capacity, durability, fuelling, environment	●●●●	●●●●	●●●●	The best-served group for wearable data and the worst for context.
Jumps	Approach control, take-off mechanics, elastic quality	●●○○	●●●○	●●○○	Small technical margins, high measurement sensitivity required.
Throws	Rate of force development, sequencing, implement specificity	●●○○	●●○○	●●○○	Sparse literature; heavy reliance on coaching tradition.
Combined events	Breadth across ten disciplines, recovery between them	●○○○	●●○○	●●○○	The thinnest evidence base and the hardest planning problem.
Race walking	Technique under judged constraint, heat tolerance	●○○○	●●○○	●●●○	Rule compliance adds a constraint no general model knows about.

Four dots indicate a comparatively strong position, one dot a comparatively weak one. The grading is the author's own assessment of the published literature, offered for challenge.

FIGURE 2 Event-specific intelligence matrix. Evidence, measurement quality and general-model fit vary sharply across the event families that share a track.

Author's assessment of the literature cited in the references. ATHLEET.AI

Two rows deserve comment. Endurance events sit at the top for measurement, since wearables have been designed around them for 15 years, and at the bottom for context, because those devices know an athlete's heart rate and nothing about their week. Combined events sit at the bottom of almost everything, which is unfortunate given that they present the hardest planning problem in the sport.

2.2 The evidence is thinner than the confidence

An integrative review of elite sprint performance makes the point plainly, observing that a considerable gap separates sprint science from sprint practice, largely because most published sprint research uses young team-sport athletes performing brief maximal efforts rather than sprinters doing the varied work that sprinting actually involves.[57] COMMENTARY If that is true of the best-served event group in the sport, it is a generous description of the others.

None of this argues for ignoring the research. It argues for a system that knows which event it is discussing, knows how strong the evidence behind a suggestion is, and says so, so that a coach can weigh a well-supported recommendation about sprint mechanics differently from a thinly supported one about throwing volume. That distinction is easy to state and rarely implemented.

ILLUSTRATIVE SCENARIO

The combined-events problem

An invented programme has one heptathlete and four specialists. On a Wednesday the heptathlete has a hurdles session and a shot session on the same afternoon, followed by a long jump session on Friday. The hurdles coach wants full recovery before Friday. The throws coach wants the shot work moved earlier in the week. Both are reasonable, both are arguing from their own event's evidence base, and neither has visibility of the other's plan.

A useful system does not resolve that disagreement. It makes the disagreement visible in one place, shows what the athlete's record says about how she has responded to similar weeks, records which option the lead coach chose, and makes that reasoning available in October when everybody has forgotten it. That is a modest claim and a genuinely valuable one.

CHAPTER THREE

What the technology can actually measure

A validated measurement and an actionable one are different things, and conflating them is the most expensive mistake a performance programme currently makes.

Measurement in athletics has improved substantially in a decade. Markerless capture has removed the marker set from biomechanics, force plates have become affordable enough for a well-funded club, and split timing that once needed a laboratory now runs from a phone. What has not kept pace is the evidence connecting those measurements to decisions, and the distance between the two is where a great deal of money is currently being wasted.

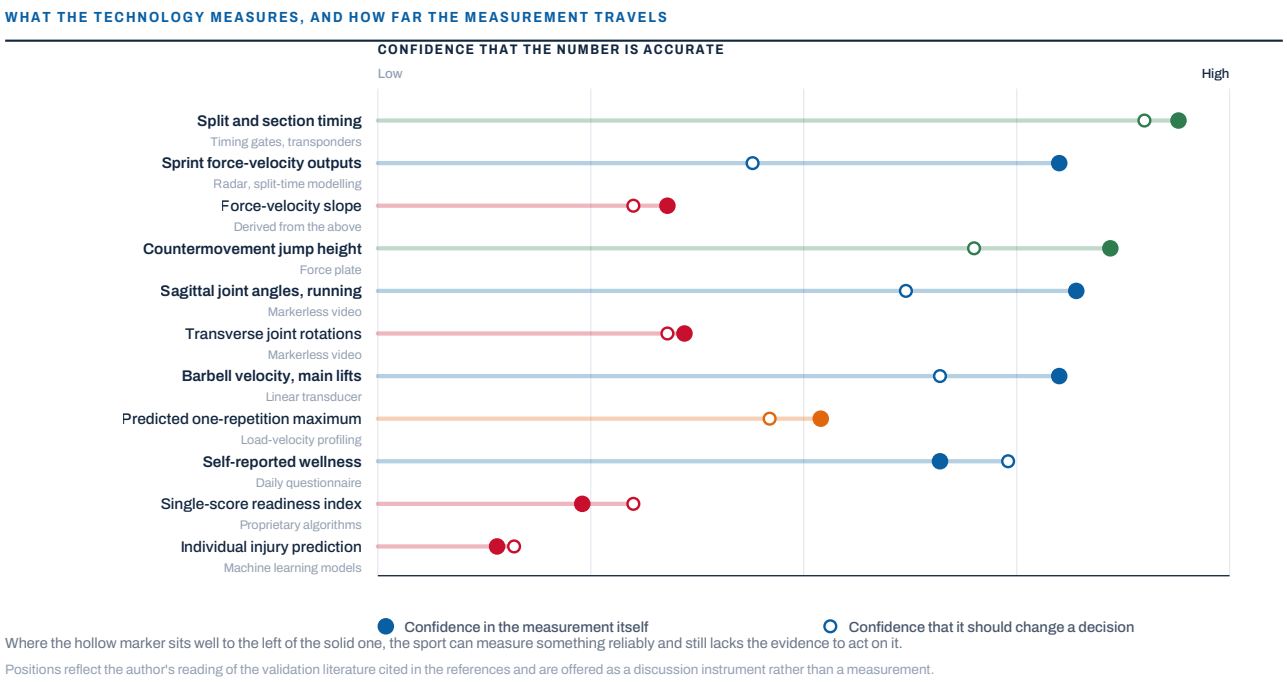


FIGURE 3 Two different confidences. The solid marker shows how far a measurement can be trusted as a number; the hollow marker shows how far it can be trusted to change a decision.

Author's reading of the validation literature at references 34 to 44. ATHLEET.AI

3.1 Markerless video

Validation work comparing a commercial markerless system against marker-based capture during treadmill gait reported joint-centre differences below 2.5 cm at every joint except the hip, where the difference was 3.6 cm, and segment angles within around 5.5 degrees except for rotations about the long axis of a segment, which performed considerably worse.[38] PEER-REVIEWED A systematic review and meta-analysis of 22 studies reached a compatible position, finding good to excellent accuracy for spatiotemporal measures such as walking speed, step time and step length, moderate to excellent agreement at the hip and knee in the sagittal plane, and poor validity at the ankle and in the transverse and frontal planes, with hip rotation carrying the largest error of all.[39] REVIEW

Both findings should be read with their populations in mind. The validation was performed during gait, and the sport wants to examine sprinting, throwing and jumping. Extending a walking validation to a javelin release is an inference the studies do not support, and a supplier quoting only the encouraging half of that literature is selling on a partial reading.

THE EXTRAPOLATION PROBLEM

Nearly all markerless validation has been conducted during walking or steady running. Very little addresses maximal sprinting, throwing or jumping, which is what athletics coaches want to look at.

Open-source pipelines using ordinary smartphone video have also been validated against laboratory methods, with the authors reporting that a clinician using the tool estimated movement dynamics roughly 25 times faster and at under one per cent of the cost.[40] PEER-REVIEWED That change in access also puts biomechanical output into the hands of people with no training in reading its error bounds, which is a governance problem rather than a technical one.

3.2 Force-velocity profiling

The field method for estimating sprint mechanical outputs from split times has been validated against force-plate measurement, with reported bias below five per cent for maximal horizontal force, velocity and power.[34] PEER-REVIEWED That is a good result for the outputs themselves. The variable most often used to indi-

vidualise a sprinter's training is the slope of the profile, and a reliability study in 30 young sprinters found that slope and the associated measure of mechanical effectiveness performed poorly between sessions, with intraclass correlations of roughly 0.31 to 0.33.[35] PEER-REVIEWED

Satellite-derived versions of the same profile have been validated against radar in elite under-18 rugby players, with moderate to nearly perfect correlations and good inter-unit reliability.[37] PEER-REVIEWED That population is a long way from a track sprinter at maximal velocity, and the absence of an equivalent validation in dedicated sprinters is a genuine gap in the literature rather than a detail.

A stronger criticism has also been published, arguing that sprint force-velocity profiling is a mathematical re-expression of velocity-time data that adds no physiological information and rests on an implausible assumption about training force and velocity capacities independently.[36] COMMENTARY My own reading, which goes beyond what any single paper states, is that the outputs are worth recording as descriptive quantities and the slope should not drive an individual's programme until the between-session reliability improves.

3.3 Jumps, bars and questionnaires

Countermovement jump monitoring has a substantial evidence base. A meta-analysis covering 151 articles and 531 effect sizes found that average jump height across a set was more sensitive to neuromuscular fatigue and supercompensation than the single best jump, which is the value most systems report by default.[41] REVIEW Earlier work showed that jump height alone can miss fatigue-related changes that other force-time variables detect.[42] PEER-REVIEWED A programme reporting best jump height is therefore using the least sensitive summary of a well-validated test.

Velocity-based training presents a similar picture. A systematic review of linear position and velocity transducers found validity and reliability that varied by device and by exercise, performing less well for plyometric and multi-planar movements.[43] REVIEW A review of load-velocity methods for predicting one-repetition maximum across 25 studies and 842 participants found accuracy varying with the number of loads used, the exercise, the velocity metric and the device, with no single approach uniformly accurate.[44] REVIEW

Against all of that sophistication sits the daily wellness questionnaire, which a systematic review found more sensitive to acute and chronic changes in training load than the objective markers commonly used alongside it.[18] REVIEW Sensitivity to load change is a different claim from predicting injury, and the review should not be stretched into the second. Even read narrowly it suggests that the cheapest instrument in the building is among the more informative, which is not the conclusion most procurement processes reach.

CHAPTER FOUR

From the annual plan to Tuesday morning

Periodisation is a planning language rather than a physiological law, and software that treats it as law produces plans that are internally consistent and externally useless.

Most planning software in sport encodes a version of periodisation theory and then assumes the athlete will comply with it. Coaches know better. The plan is a statement of intent that degrades on contact with illness, weather, selection, examinations and travel, and the skill of coaching lies substantially in deciding which departures matter.

The theoretical foundations have been questioned seriously and from within the discipline. Kiely has argued across two influential papers that periodisation's assumptions rest on outdated stress physiology and that the psychological and emotional state of the athlete modulates the training response enough to undermine any universal template.[29,30] COMMENTARY A systematic review of the meta-analyses supporting periodisation found that none of the 21 underlying studies compared periodised training against varied non-periodised training, since the comparator was almost always unvarying training.[32] REVIEW The claim that periodisation beats variation has, on that reading, not been tested.

READ CAREFULLY

Periodisation's critics do not argue for training without structure. They argue that the specific templates the sport inherited were never tested against the alternatives people assume they beat.

Structured alternatives have their defenders. Block periodisation has been argued for on the grounds that concentrated, sequenced loading suits the multi-peaking demands of high-performance sport better than simultaneous mixed training,[31] COMMENTARY although that case is made by the model's own author rather than tested against it.

Evidence in the other direction exists and should be stated. A meta-analysis of 35 volume-equated resistance-training studies found greater one-repetition-maximum gains from periodised programmes, with undulating models outperforming linear ones in trained participants, and no difference for hypertrophy.[33] REVIEW That finding concerns strength training rather than sprinting or throwing, and extending it to the track would be exactly the kind of move this paper argues against.

4.1 Four horizons, one record

The practical resolution is to treat the plan as a set of nested horizons that revise one another, and to insist that evidence travels upward as readily as instructions travel downward. Most software moves instructions down the hierarchy and loses the return path entirely, which is why so many athlete management systems become expensive filing cabinets within a season.

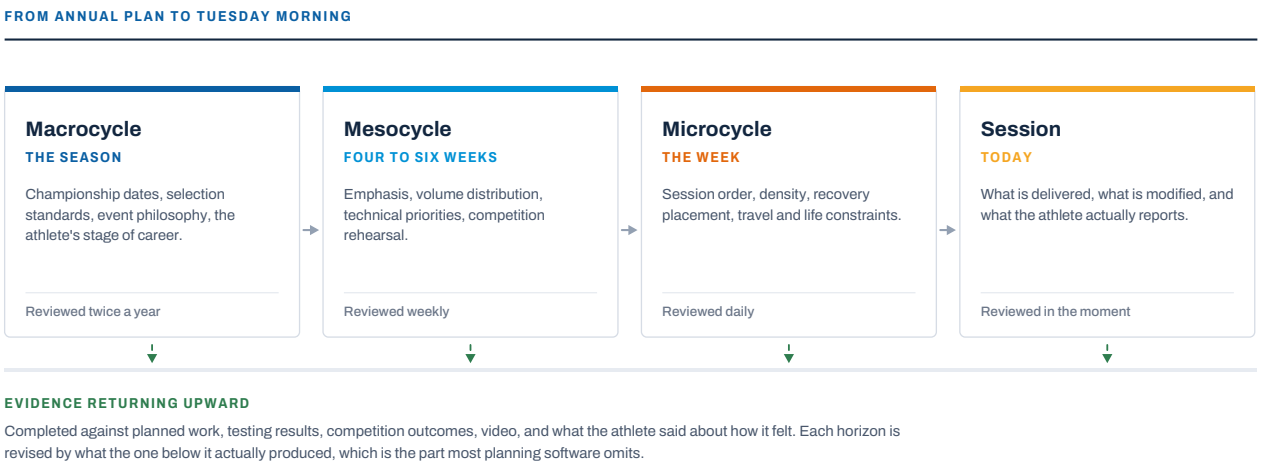


FIGURE 4 Four planning horizons and the evidence that revises them. The upward arrows are the part most planning tools omit. Athleet.AI framework. ATHLEET.AI

Two consequences follow for anybody specifying a system. Completed work must be recorded against planned work at session level, because the difference between the two is the most informative quantity a programme holds and is almost never analysed. And every modification must record its reason, since a session shortened for a hamstring complaint and one shortened for an examination look identical in a database and mean entirely different things.

4.2 What a planning assistant should and should not do

A planning assistant earns its place by drafting, by checking a proposed week against the squad's own policies, by flagging where this week departs from what has worked before, and by preparing the evidence a coach needs before changing their mind. It loses that place the moment it changes a plan without a person seeing the change, since the coach can then no longer answer the only question that matters after a poor outcome.

SPECIFICATION

Four requirements for planning software in athletics

- The event model is explicit and editable, so a coach can state their own philosophy and have the system hold it rather than override it.
- Planned and completed work are stored separately and compared automatically, with the reason for every modification captured at the moment it happens.
- Every generated suggestion is traceable to the athlete's own record and to the policies the squad has written down.
- No plan reaches an athlete without a named person approving it, and that approval is stored with the plan.

CHAPTER FIVE

The technical athlete

Video analysis becomes dangerous at the point where a measured joint angle is mistaken for a coaching instruction. The measurement is the easy part.

Every technical event carries a model of how the movement should look, and every good coach holds that model loosely. Athletes solve the same mechanical problem differently because they have different levers, histories and tolerances, and the pursuit of a single correct position has wasted more seasons in this sport than almost any other idea. Automated video analysis reintroduces that pursuit with new authority, because a number attached to a joint angle looks more objective than an opinion.

The measurement problem was set out in Chapter 3 and does not need repeating, beyond the observation that the axes on which markerless systems perform worst, meaning rotations about the long axis of a segment, are exactly the axes coaches most often want to discuss in throwing and hurdling.[38,39] PEER-REVIEWED A supplier presenting hip rotation from smartphone video to three decimal places is presenting precision the underlying method does not have.

THE REAL QUESTION

The useful question about a technical measurement is whether it distinguishes this athlete from their own previous self, rather than from an idealised model.

5.1 What video analysis is genuinely good at

Set against those limits, three uses stand up well. Retrieval comes first, meaning the ability to find every one of an athlete's attempts at a technical element across four seasons in seconds, work that used to cost an assistant coach an evening. Longitudinal comparison of an athlete against themselves comes second, since systematic error partly cancels and the change over time carries more information than any absolute value. Communication comes third, because an athlete who sees two of her own attempts side by side learns faster than one being told about the difference.

None of those uses requires the system to say what is correct. They require it to organise, to retrieve and to present, which is work software does extremely well and which currently consumes an enormous amount of skilled coaching time.

ILLUSTRATIVE SCENARIO

The invented jumper and the tidy conclusion

An invented long jumper has lost distance across four competitions. Automated analysis reports that her penultimate step has shortened and her take-off knee angle has changed by four degrees. The system, if poorly designed, presents this as a technical fault to correct.

Her coach knows three further things. She changed spikes in April, she has been carrying a heavy academic term, and the approach at two of those venues ran into a headwind. The measurement is real. The conclusion drawn from it in isolation is worthless, and a system that offered it as an instruction would have made the coach's job harder.

5.2 Holding the coach's own model

The most valuable thing a system can hold in a technical event is the coach's own philosophy, stated explicitly and applied consistently. Two respected sprint coaches will describe acceleration mechanics in ways that partly contradict each

other, and both produce international athletes. A system that imposes a third view helps nobody, while one that records what this coach believes and shows where an athlete departs from it is doing something no coach can do alone across 15 athletes.

CHAPTER SIX

Testing that changes training

A test that never alters a session is an expensive ritual. Most testing batteries in the sport contain several.

Ask a performance programme why it runs a particular test and the answers divide into three groups. Some tests inform a decision that would otherwise be made blind. Some produce a number that is recorded, filed and never consulted. The rest survive because a previous member of staff introduced them and nobody has felt able to stop, and that third group is larger than most programmes admit. An honest audit is usually the cheapest performance intervention available to them.

Two properties determine whether a test can change a decision. The first is reliability, meaning how much the number moves when nothing has changed, since a change smaller than the measurement noise carries no information whatever the software displays. The second is a threshold established in a comparable population. Many tests in athletics satisfy the first and fail the second, so they tell a coach that something has moved without telling them whether it matters.

A PRACTICAL AUDIT

For each test in the battery, name the decision it changes and the threshold at which it changes it. Tests that survive that question are worth keeping. The rest are ritual.

6.1 The evidence-to-intervention ladder

The framework I use to keep those distinctions visible is a ladder. Each rung deserves a different degree of trust, and the most common failure in performance software is a recommendation that mixes rungs without saying which is which, so that a measured fact, a model estimate and an opinion arrive in the same sentence wearing the same clothes.

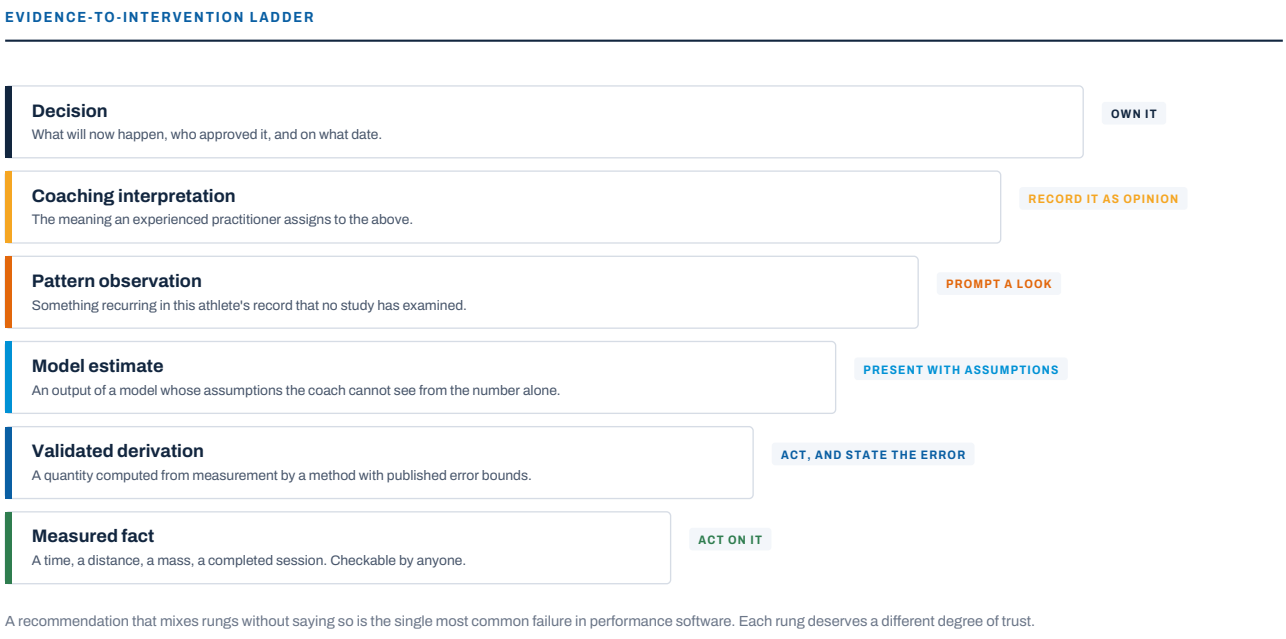


FIGURE 5 The evidence-to-intervention ladder. The width of each rung indicates how much interpretation has been added between the measurement and the action.

Athleet.AI framework. ATHLEET.AI

Applying the ladder to the countermovement jump makes the point concrete. The force-time trace is a measured fact, and jump height computed from flight time is a validated derivation with a known error. A fatigue index built from several force-time variables is a model estimate whose assumptions the coach cannot see. That this athlete's values fall in the fortnight after a competition block is a pattern in her record, and the judgement that she needs an easier Thursday is coaching interpretation. Only the last is a decision, and only a person should make it.

6.2 Turning a battery into a decision

A system earns its place by holding the reliability of each test alongside the result, so that a change is flagged only when it exceeds what the test can detect, and by comparing against the athlete's own history rather than a norm derived from undergraduate team-sport players. It should also be capable of saying nothing. A dashboard producing an insight every week while the athlete trains steadily is generating noise and calling it analysis.

TABLE 1 Four common tests, and the decision each can honestly support

TEST	WHAT IT MEASURES WELL	THE DECISION IT CAN SUPPORT	WHAT IT CANNOT DO
Countermovement jump	Neuromuscular state, using average rather than best jump height	Whether to proceed with a planned high-intensity session	Identify why the change occurred, or predict injury
Sprint split timing	Acceleration and maximal velocity phases, with high reliability	Whether the technical emphasis of a block has transferred	Separate mechanical change from weather and surface
Load-velocity profile	Bar velocity at a given load, device permitting	Daily adjustment of prescribed load in the main lifts	Deliver an accurate one-repetition maximum across all methods
Daily wellness entry	The athlete's own perception, tracked against their own range	Whether to open a conversation before the session starts	Serve as a medical assessment or a readiness score

Compiled by the author from references 18, 34, 35, 41, 42, 43 and 44. The final column matters more than the third.

CHAPTER SEVEN

Smarter training

Load is not a single number, and the sport's habit of treating it as one has produced a decade of confident analysis resting on arithmetic that does not hold.

Every paper in this series examines the same management problem through the demands of a different sport. What the event requires, what the coach plans, how the athlete responds, what the coach then decides, whether those decisions are building capacity, and what the resulting continuity is worth. Athletics presents that problem in an unusually pure form, because the training is highly specific, the competition exposure is short and the consequences of a poorly judged fortnight are visible within a month.

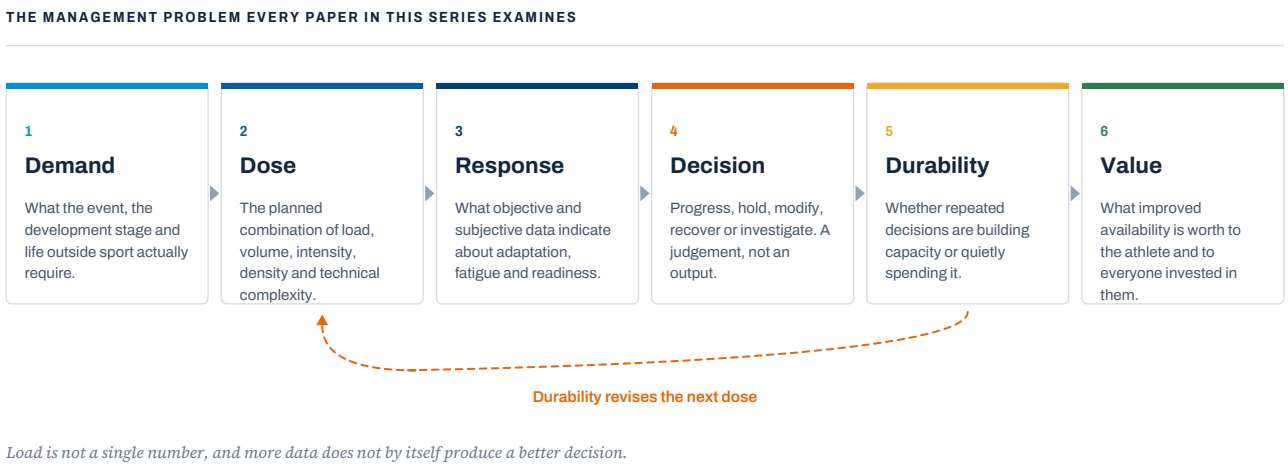


FIGURE 6 Demand, dose, response, decision, durability and value. The dotted return path is the part that distinguishes a training record from a training system.

Athleet.AI framework, used across the whole series. ATHLEET.AI

7.1 Internal load, external load and the distance between them

The conceptual separation between what an athlete was asked to do and what that work cost them has been in the literature for two decades and is still applied loosely.[13] COMMENTARY The external dose is the session as written, meaning the distance, the number of throws, the mass on the bar. The internal response is what that dose demanded of this athlete on this day, given sleep, illness, heat, examinations and everything else the plan competes with. Two athletes completing an identical session have completed two different sessions.

The simplest instrument for capturing the internal side is session rating of perceived exertion, validated against heart-rate-based measures in the original study and used almost universally since.[14] PEER-REVIEWED It costs one question after training. A programme that has spent heavily on external measurement while collecting no internal measure has bought half of a picture and will interpret it as a whole one.

SESSION-RPE

Duration multiplied by rating of perceived exertion remains among the most useful instruments in the sport, largely because it captures the internal response at almost no cost.

7.2 What the consensus statements actually say

The 2016 consensus statement from the International Olympic Committee concluded that both rapid increases in load and insufficient load are associated with

injury risk, and offered practical monitoring recommendations on that basis.[15] A companion statement addressed the relationship between load and illness.[16] A separate expert consensus reviewed the available internal and external monitoring methods and their practical use.[17] CONSENSUS

Those documents are expert synthesis rather than primary evidence, and they inherit the limitations of the observational literature beneath them. Read carefully, they support a modest and defensible position, which is that load should be recorded, that large sudden changes deserve attention, and that the relationship between load and injury is real, complicated and not reducible to a threshold. That is a good deal less than the sport has been told.

7.3 Density, complexity and the things nobody counts

Athletics carries dimensions of load that general monitoring frameworks handle poorly. Technical complexity is one, since 30 competition-standard hurdle clearances tax an athlete differently from 30 rhythm runs at the same speed. Landing volume in the jumps is a quantity many programmes still do not count. Competition density is a third, and for a hurdler facing three rounds in two days it may be the dominant variable of the week.

A system that records only distance and intensity will describe those weeks as identical. My own view, which the published evidence supports in direction rather than in detail, is that event-specific load dimensions deserve to be first-class quantities in any athletics system, and that their absence from most commercial platforms is a design decision inherited from team sport rather than a considered one.

CHAPTER EIGHT

Readiness, availability and the limits of prediction

The sport has been sold individual injury prediction for a decade. The methodological literature has spent that decade explaining, with some patience, why it does not work.

This is the chapter where a paper of this kind is expected to become enthusiastic about machine learning, and it is the chapter where the evidence most firmly declines to cooperate. The claim that a model can predict which athlete will be injured, and when, has been examined repeatedly and has not survived the examination. Stating that plainly is more useful to a performance director than any amount of qualified optimism.

8.1 The arithmetic problem

The acute to chronic workload ratio entered the sport through a widely cited narrative review proposing that high chronic loads can be protective when they are reached gradually.[1] PRACTICE Empirical work in cricket fast bowlers had already linked spikes in acute workload to injury.[2] PEER-REVIEWED The idea spread quickly, in part because it was intuitive and in part because it produced a single number that fitted comfortably on a dashboard.

READ THIS CHAPTER TWICE

The evidence against ratio-based individual prediction is stronger and more consistent than the evidence that established the idea. Very few commercial products reflect that.

A conceptual model published shortly afterwards proposed a multi-stage causal chain running from external load through internal load to response and adaptation, in an attempt to explain why workload and injury findings had been so inconsistent.[3] COMMENTARY That model was never itself tested against data, and the ratio spread faster than the caution attached to it.

The statistical objections arrived afterwards and have accumulated. Because the acute period sits inside the chronic period, the numerator forms part of its own denominator, which generates correlation independent of any physiological relationship.[4] A biostatistical critique argued that the widely reproduced optimal zone can be an artefact of how continuous values are grouped, and that reported effects may be substantially inflated.[5,6] Most damaging of all, a later paper demonstrated that the ratio continues to appear predictive when the chronic component is replaced with contrived values, and its authors recommended abandoning the measure and the theory beneath it.[7] COMMENTARY

8.2 The inference problem

Separate objections concern causation rather than arithmetic. Aggregated load says nothing about the tissue that failed, and a call for structure-specific load and explicit causal reasoning has been made from within the field.[8] COMMENTARY Two companion papers argued that the discipline's belief in manipulating load to prevent injury has outrun the associational evidence supporting it.[9,10] COMMENTARY

Underneath all of it sits a simpler mathematical point. Bahr showed that because injuries are low-incidence and multifactorial events, a screening test with respectable sensitivity and specificity will still have poor positive predictive value at the individual level.[11] COMMENTARY A complementary argument holds that injury emerges from a complex system of interacting factors and resists single-variable prediction altogether.[12] COMMENTARY

8.3 Machine learning has not rescued it

Two systematic reviews settle the question for now. The first found widespread methodological weakness across machine learning injury-prediction studies, including small and imbalanced datasets and an absence of external validation.[20] REVIEW The second appraised 204 models across 30 studies and reported that 98 per cent were at high or unclear risk of bias, with external validation inadequate or absent in most, and concluded that no model was ready for practical use.[21] REVIEW

WHAT THIS RULES OUT

Three claims a supplier should be asked to withdraw

- That a single readiness score, however derived, indicates whether an individual athlete should train today.
- That a model can identify which athletes will be injured in the coming weeks at a useful individual accuracy.
- That a workload ratio has an optimal zone applicable to a specific athlete in a specific event.

None of these is a question of tuning. Each fails on grounds the published methodological literature sets out at length, and a supplier unable to engage with that literature is not ready to sell into a high-performance programme.

8.4 What survives, and why it is enough

Removing the prediction claim leaves a great deal standing. A load record describes exposure accurately, which supports an honest conversation between coach, athlete and physiotherapist. Departures from an athlete's own established pattern are worth a coach's attention, presented as a prompt to look rather than an instruction to act. Completed against planned work is a fact. Self-reported wellness tracks the training response with useful sensitivity.[18] REVIEW Heart-rate variability has been argued to index autonomic status in elite endurance athletes when read as a rolling trend rather than a single day, on the basis of small and largely endurance-specific studies.[19] PRACTICE And availability itself, meaning weeks of uninterrupted training, is measurable, meaningful and directly connected to performance, which is the subject of Chapter 11.

My own position, stated as opinion, is that the sport spent a decade chasing prediction because prediction is exciting, and neglected continuity because continuity is administrative. The administrative one is where the value turned out to be.

CHAPTER NINE

Peaking and championship preparation

A championship compresses a season's decisions into ten days, and it is the worst possible moment to be reconstructing a rationale from memory and a group chat.

Championship preparation is where the quality of a programme's record-keeping stops being an administrative virtue and becomes a competitive one. Selection has to be argued, travel and heat planned, a taper designed for an athlete whose spring has already departed from the plan twice, and rounds recovered from. Every one of those decisions is made by people who are tired, in an unfamiliar hotel, working from partial information.

FIGURE 7 The championship decision timeline. Each marker is a decision that will be examined afterwards, which is a reason to record the reasoning at the time.

Athleet.AI framework. ATHLEET.AI

9.1 What the taper literature supports

Tapering is among the better-evidenced areas of the sport. A meta-analysis of 27 studies found the largest performance improvement from a taper of around two weeks with an exponential reduction in training volume of 41 to 60 per cent, with intensity and frequency largely maintained, reporting an overall effect of 0.72 for the volume-reduction comparison.[45] REVIEW An earlier narrative synthesis set out the physiological mechanisms proposed to underlie the effect.[46] COMMENTARY

Those numbers describe a central tendency across several sports, and the studies were mostly pre-post designs rather than controlled comparisons with a non-tapering arm. A system should hold that range as a starting position rather than a prescription, alongside what this athlete's record says about her last three tapers, which is information no meta-analysis contains.

41 to 60%

VOLUME REDUCTION

The range associated with the largest tapering effect in a meta-analysis of 27 studies, across roughly two weeks, with intensity maintained.

9.2 Heat, travel and the things that are not training

Environmental preparation has moved from folklore to guidance. A multidisciplinary consensus recommends heat acclimatisation over one to two weeks of repeated exercise-heat exposure, alongside hydration and cooling strategies and adjustments to scheduling and recovery.[47] CONSENSUS Field data from a world championship held in extreme conditions found that ice-vest pre-cooling significantly lowered pre-race core temperature, at 37.5 degrees against 37.8 degrees, while core temperature itself was unrelated to finishing position.[48] PEER-REVIEWED

A survey of 66 elite race walkers at a subsequent championship is more revealing about practice than about physiology. Every medallist had used heat acclimation, top-ten finishers were more likely to have used it, and 43 per cent of the surveyed athletes had undertaken no specific heat preparation at all.[49] PEER-REVIEWED The design is cross-sectional and cannot establish causation. It does establish that a well-evidenced and inexpensive preparation was skipped by a large minority of elite competitors, which is a coordination failure rather than a knowledge failure, and coordination failures are precisely what a system can help with.

ILLUSTRATIVE SCENARIO

The invented 800m runner and the third round

An invented middle-distance runner faces heats, a semi-final and a final across four days in heat. The performance team has agreed a cooling protocol, a fuelling plan and a contingency should the semi-final be unusually fast. On the second evening the coach is asked whether to change the warm-up before the semi-final because the athlete reported heavy legs.

What helps at that moment is a single view holding the agreed plan, what has already been done, what the athlete reported after her heat, and what the team decided the last time this happened. What does not help is a readiness score. The decision is a coaching judgement made under time pressure, and the system's contribution is to ensure it is made with the full picture rather than a partial one, and to record what was chosen so that the reasoning survives into the debrief.

9.3 Selection, and the discipline of writing it down

Selection is the most contested decision in any federation and the one most often documented after the fact. Where a system helps is by assembling the evidence against published criteria before the meeting, showing which athletes meet which standards and where the panel is exercising discretion, and recording the reasoning at the time. That record protects the panel, and it protects the athlete who was not selected and deserves a straight answer.

My view, which goes beyond anything the literature addresses, is that selection support is the single most commercially attractive and reputationally dangerous application in this whole field. Any system touching it must present evidence and stop there. The moment a model produces a ranking, the panel's discretion becomes a deviation from an algorithm, and the federation has acquired a legal and cultural problem it did not have before.

CHAPTER TEN

Mixed squads and the multidisciplinary team

Individualisation is a resource problem before it is a scientific one. One coach, 15 athletes and six events is the constraint that actually shapes the sport.

The literature on individualised training assumes a coach with the time to individualise. In practice a national event lead may carry 15 athletes across six events, a job elsewhere, and a support team who meet properly once a month. Individualisation then collapses towards the group, because attention is finite and Tuesday arrives on schedule regardless of how much the coach knows.

FIGURE 8 The athlete digital context model. The fourth and fifth elements, meaning life context and the decision record, are the two most commonly missing and the two that most determine whether advice is any use.

Athleet.AI framework. ATHLEET.AI

This is where a system multiplies a coach rather than replacing one. Preparing 15 individual weekly reviews takes an evening. Preparing 15 drafts for a coach to correct takes minutes, and the corrections are where the coaching happens. The distinction sounds like a technicality and is the whole design.

10.1 The gap between science and practice

Research on knowledge translation in elite sport reports that coaches acquire knowledge mainly through personal interaction rather than through journals or formal courses, and that barriers to translating sport science into practice include time, funding and the willingness of practitioners to adopt it.[58] COMMENTARY A related framework proposes structured approaches for sport scientists to improve that translation, though its own effectiveness has not been evaluated.[59] PRACTICE

That has a design consequence most sports software ignores. If coaches learn through conversation rather than reading, evidence should arrive inside the decision a coach is already making and in the language of that decision. A well-designed system is a colleague who has read the literature and mentions the relevant part at the relevant moment, which is a more modest ambition than most product marketing in this sector.

DESIGN CONSEQUENCE

If coaches learn through conversation, a system that delivers findings as documents will be ignored. Delivering them inside the decision they are already making is the only route that works.

10.2 Making disagreement productive

Multidisciplinary teams disagree, and the disagreements are frequently the most valuable information in the room. A physiotherapist's caution and a coach's ambition are both legitimate and they conflict, and the outcome usually depends on who spoke last and how confident they sounded. Recording the positions, the evidence behind each and the decision that followed turns a recurring argument into an accumulating body of institutional knowledge.

The practical instrument is unglamorous. Every performance meeting should produce a written record of what was decided, who decided it, what evidence was

in front of them and what would cause the decision to be revisited. A system that prepares that record and stores it is doing something genuinely valuable, and it is doing it in an area where the technology carries almost no risk of harm.

CHAPTER ELEVEN

Career durability and what it is worth

Training continuity is measurable within a season, connected to performance, and almost entirely ignored as a performance objective in its own right.

Chapter 8 argued that individual injury prediction is beyond the current evidence. This chapter argues that the sport should pursue something less exciting and considerably better supported, which is training continuity, meaning the number of weeks in which an athlete trains as intended without interruption.

A five-year prospective study of elite Australian track and field athletes reported that fewer weeks lost to injury and illness was associated with a greater likelihood of achieving performance objectives.[22] PEER-REVIEWED A systematic review across sports reached the same direction of travel.[23] REVIEW A cross-sectional analysis across eight international championships and 13,059 athletes found that teams with lower injury and illness rates achieved relatively better medal outcomes, though as an ecological analysis it cannot establish individual-level causation.[27] PEER-REVIEWED

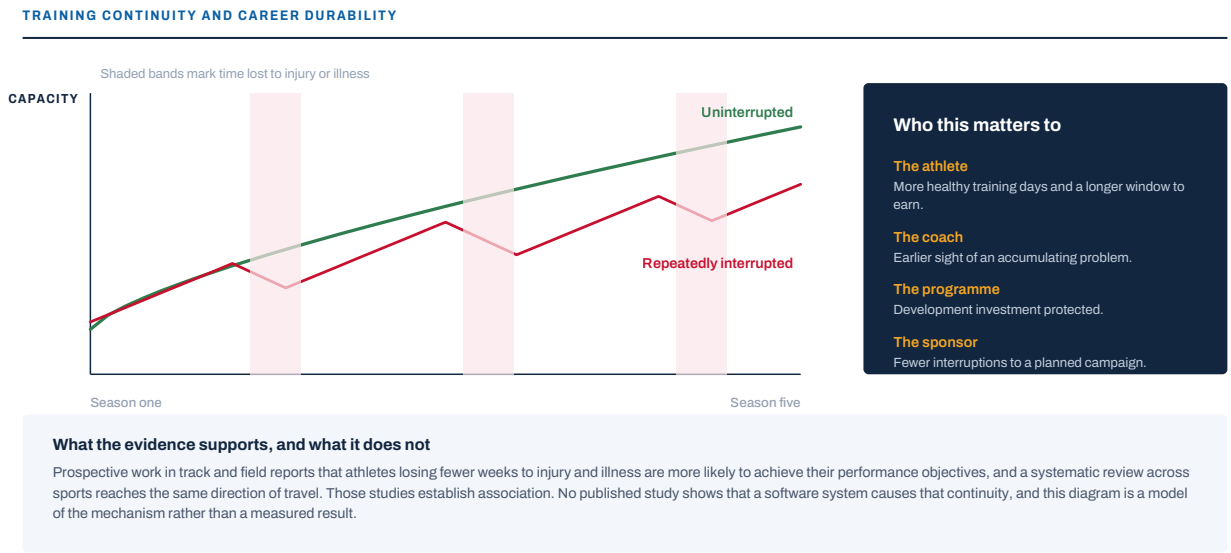


FIGURE 9 Two illustrative capacity trajectories across five seasons. The curves are a model of the mechanism rather than measured data, and the panel beneath states precisely what the evidence does and does not support.

Model by the author; evidence at references 22, 23 and 27. ATHLEET.AI

11.1 Where the interruptions come from

Championship surveillance gives a clear picture of what interrupts athletics careers. Muscle injuries are the dominant type across international championships, with hamstring injury the leading single diagnosis in pooled analysis covering 2007 to 2015.[28] Surveillance at a single championship covering 1,851 athletes reported 134.5 injuries and 68.1 illnesses per 1,000 athletes, with hamstring

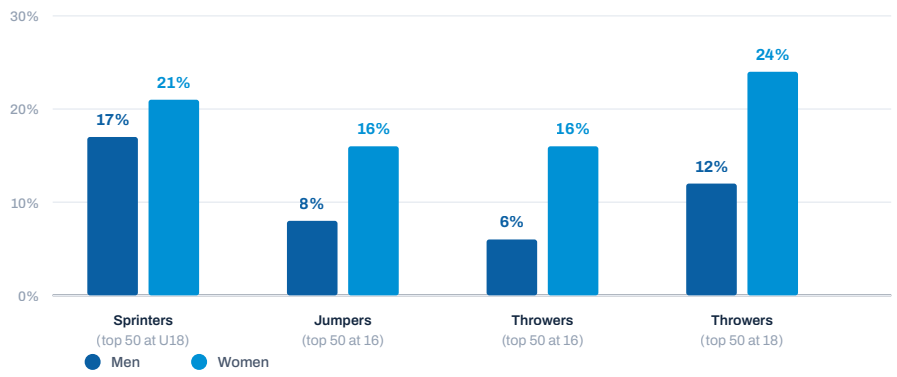
strain the most frequent diagnosis and a substantial share attributable to overuse.[26] Earlier championship surveillance established the method and the baseline.[24,25] PEER-REVIEWED

Those figures describe competition periods rather than whole seasons and should not be read as an annual rate. They do establish that interruptions concentrate in a few predictable categories, which makes them a reasonable target for attention even without a prediction model.

11.2 How little early success tells you

Any argument about career length runs into a widespread assumption that the sport can identify who will have one. The evidence is unkind to that assumption. A meta-analysis of 129 effect sizes across 13,392 athletes in Olympic sports found a pooled junior-to-senior performance correlation of 0.148, meaning junior performance explained around 2.2 per cent of the reliable variance in senior performance.[56] REVIEW

HOW LITTLE JUNIOR RANKING TELLS YOU



Share of world-ranked junior athletes who reached the equivalent senior ranking. Data from Boccia and colleagues, World Athletics ranking records.

2.2%

of the reliable variance in senior performance is explained by junior performance, pooled across 129 effect sizes and 13,392 athletes in Olympic sports.

BARTH AND COLLEAGUES, 2024

FIGURE 10 Junior-to-senior conversion among world-ranked athletes, by event group and sex.

Boccia and colleagues, references 52, 53 and 54, from World Athletics ranking records; variance figure from reference 56 (Barth and colleagues, 2024). PEER-REVIEWED

Athletics-specific work sharpens the picture considerably. Among world-ranked sprinters, 17 per cent of men and 21 per cent of women in the top 50 at under-18 level reached the senior top 50.[52] Among throwers the transition rate was six per cent for men and 16 per cent for women at age 16, rising to 12 and 24 per cent at age 18.[53] Among jumpers, eight per cent of men and 16 per cent of women in the top 50 at 16 reached the senior top 50.[54] An Italian dataset of 5,929 athletes approached the same point from the other direction, finding that only 17 to 26 per cent of top-level adult athletes had themselves been top-level at 14 to 17.[51] PEER-REVIEWED

A further finding deserves attention from anyone designing a pathway. Among world-class sprinters and jumpers, relatively younger athletes within a selection year were less likely to be elite as juniors, and those who did reach that level had higher odds of remaining elite as seniors.[55] PEER-REVIEWED Age at peak performance meanwhile ranges from roughly 24 years in men's 10,000m to roughly 28 in the discus.[50] PEER-REVIEWED A system that preserves the record of an athlete who was overlooked at 17, and makes it retrievable at 21, is doing something the evidence says the sport badly needs.

11.3 The value, stated honestly

Continuity is worth something to several parties at once. To the athlete it means more healthy training days and a longer window in which to earn. To the coach it means earlier sight of an accumulating problem. To a programme it protects development investment. To a sponsor it means fewer interruptions to a planned campaign and a more predictable relationship.

HOW TO MEASURE THIS IN A PILOT

Five measures available within one season

- Availability rate, meaning the proportion of planned training weeks completed without interruption.
- Completed against planned work at session level, recorded with the reason for every modification.
- Time-loss episodes, their duration and their category.
- Progression against individual development objectives agreed at the start of the season.
- Competition starts and season continuity, meaning the length of the unbroken competitive block.

Career duration is not on that list, because a season cannot measure it. Any supplier presenting career extension as a pilot outcome is describing a hypothesis, and it should be stated as one.

CHAPTER TWELVE

Explainable recommendations

A recommendation a coach cannot interrogate will be ignored by the experienced and followed uncritically by the inexperienced. Both outcomes damage the athlete.

Explainability is often treated as a compliance requirement bolted on after the model works. In coaching it is a performance requirement, because a good coach's first response to any suggestion is to ask why, and a system that cannot answer has failed at the first exchange. The inexperienced coach who accepts without asking is the more worrying case, since the system has then substituted its confidence for their judgement without either party noticing.

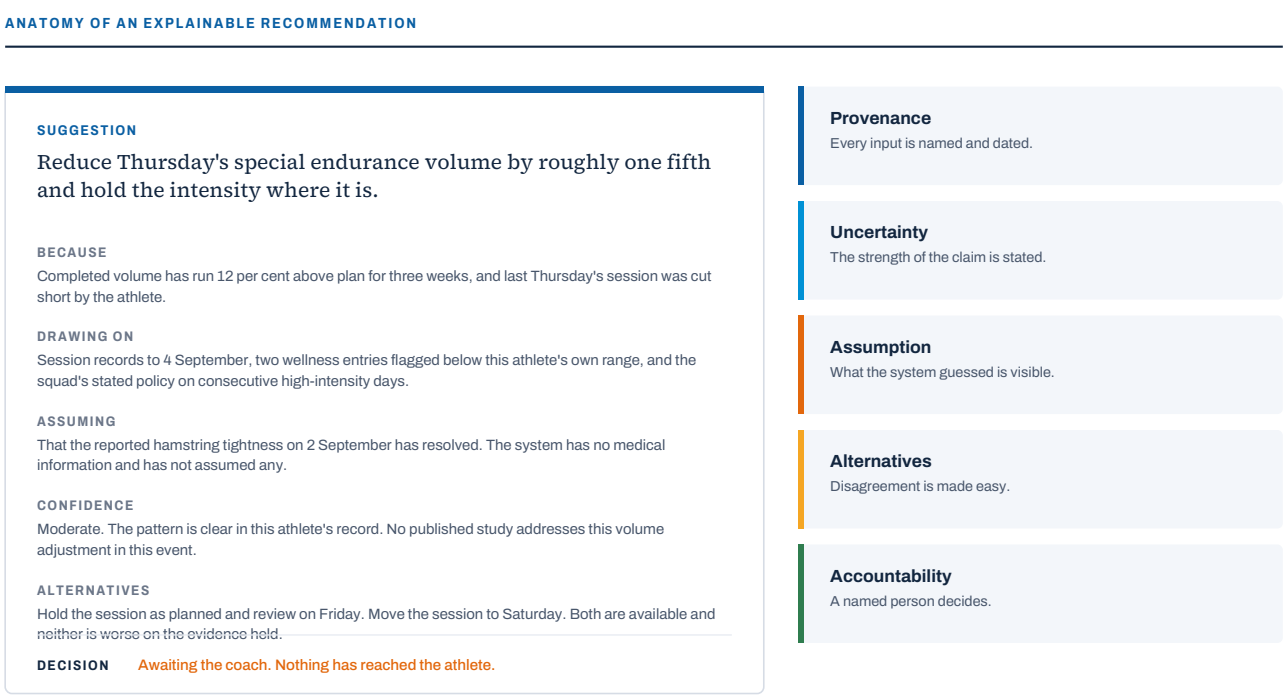


FIGURE 11 The anatomy of an explainable recommendation, with the five tests it must pass. The example is invented and no session was modified. Athleet.AI framework. ATHLEET.AI

12.1 Interpretable models and post-hoc excuses

An influential argument in machine learning holds that post-hoc explanation of black-box models can be unreliable or misleading in high-stakes settings, and that inherently interpretable models should be preferred wherever comparable accuracy is achievable.[65] PEER-REVIEWED Earlier work argued that interpretability depends on the task and on who is doing the interpreting, rather than being one property a model either has or lacks.[76] PEER-REVIEWED

Applied to athletics, that argument points somewhere unfashionable. For most decisions a coach faces, a transparent rule, a clear comparison against the athlete's own history and a stated confidence will be more useful than a sophisticated model with a plausible explanation attached afterwards. The sport's decisions are not mostly limited by the sophistication of the available models. They are limited by the quality and completeness of the record, which is a much less glamorous problem.

A USEFUL TEST

Ask a supplier to show the same recommendation as it appears to a coach, to an athlete and to a lawyer. If the three differ in substance rather than in wording, the explanation is decoration.

12.2 What the regulators expect

Guidance produced jointly by the Information Commissioner's Office and the Alan Turing Institute sets out the types of explanation an organisation should be able to give about AI-assisted decisions, covering rationale, responsibility, data, fairness, safety and impact.[62] CONSENSUS The guidance is not binding, and my own view is that it remains the most practical checklist available to anyone specifying a system in this area.

Two legal instruments bear on athletics programmes operating in the United Kingdom and Europe. Under the UK General Data Protection Regulation, Article 9 treats data concerning health, and biometric data used to identify a person, as special category data requiring a further condition beyond a lawful basis.[63] CONSENSUS Routine performance metrics such as distance, speed and session rating of perceived exertion do not fall within that definition automatically, and

this paper does not claim they do. They fall within it once a programme uses them to infer or record something about health, which is exactly what an availability or injury-risk model does. The regulator's own test turns on whether you intend to make a health-linked inference, or to treat someone differently because of one, and not on how confident you are that the inference is right. Separately, the Data (Use and Access) Act 2025 received Royal Assent on 19 June 2025. Its section 80 will replace Article 22 of the UK GDPR with new Articles 22A to 22D on automated decision-making, but that section awaits commencement regulations, so the existing Article 22 is the operative law as this paper is written.[64] CONSENSUS

The European Union's Artificial Intelligence Act, Regulation (EU) 2024/1689, entered into force on 1 August 2024 and applies in stages.[61] CONSENSUS Sports performance systems are not named in the Annex III high-risk categories, and on my own reading they would ordinarily sit outside that tier, though a system used to make employment-related decisions about a contracted athlete would need separate consideration. That reading is mine and it is an inference rather than a regulator's position, so a federation should take its own advice. Transparency obligations for certain systems, including disclosure that a person is interacting with AI, apply from 2 August 2026.

TO CONFIRM WITH YOUR OWN ADVISERS

The dates above are those set out in Article 113 of the Regulation as enacted, which brought in prohibitions and the AI literacy duty from 2 February 2025, general-purpose model obligations from 2 August 2025, general application from 2 August 2026, and product-embedded high-risk systems from 2 August 2027. Timetables of this kind attract amendment. Any programme relying on a particular deadline should verify it against the current official position rather than against this paper.

12.3 Whose data it is

The question of athlete agency over performance data is being settled by practice faster than by law. A charter of player data rights developed in professional football sets out eight rights over performance and personal data, including access, portability, rectification and erasure, citing survey evidence that 80 per cent of professional footballers surveyed wanted access to their own performance data.[74] CONSENSUS Athletics has no equivalent instrument. The scholarly literature has argued at length that athletes hold limited practical control over biometric data commercialised by others.[75] PEER-REVIEWED

My own view is that any programme building a longitudinal athlete record should decide now, in writing, what happens to it when the athlete leaves. Doing so is cheap while the record is small and becomes contentious once it is valuable, and an athlete who cannot take their own history with them will reasonably conclude the system was never built for them.

CHAPTER THIRTEEN

Federation intelligence and institutional memory

Programmes lose more knowledge through staff turnover than through any other cause, and almost none of them treat that as a solvable problem.

A national programme typically holds its methodology in the heads of between five and 20 people. When one leaves, some of the reasoning leaves too, and the successor rebuilds it over two years from conversations and inherited spreadsheets. Other safety-critical fields solved the equivalent problem by writing decisions down and making them retrievable, and nothing about athletics prevents the same approach.

Three assets are worth building deliberately. The first is approved knowledge, meaning the federation's own guidance, event philosophies and coach education material held in one place and used as the basis for any generated advice. The second is the decision record described in Chapter 10. The third is a case library, meaning real anonymised situations with what was decided and what followed, which is the material coach education has always been short of.

THE CHEAPEST ASSET

A decision record costs almost nothing to create at the moment of decision and cannot be reconstructed afterwards at any price.

13.1 Evidence gaps that models will inherit

Any system trained on sports science will inherit that literature's blind spots. An analysis of 5,261 publications covering roughly 12.5 million participants across six sport and exercise science journals found that female participants made up around 34 per cent of the total, with female-only studies a small minority.[67] PEER-REVIEWED Earlier work had documented the same under-representation.[68] PEER-REVIEWED

A model that has learned from that literature will be more confident about male athletes than female ones, and it will not say so unless designed to. A well-documented healthcare case shows how such bias arrives through a proxy rather than intent, where an algorithm used cost as a stand-in for health need and underestimated the needs of Black patients because less had historically been spent on their care.[66] PEER-REVIEWED Selection history, funding history and squad membership are all cost-like proxies in sport.

13.2 Documentation that makes a system reviewable

Two established practices from machine learning transfer directly. Model cards document a model's intended use, its performance across subgroups and its ethical considerations.[69] PEER-REVIEWED Datasheets document a dataset's motivation, composition, collection and maintenance.[70] PEER-REVIEWED Both are inexpensive, both are demanded by almost nobody in sport, and either would improve most procurement conversations immediately.

At organisational level, the NIST AI Risk Management Framework organises the work into governing, mapping, measuring and managing risk across the lifecycle,[71] and ISO/IEC 42001, published in 2023, provides a certifiable management-system standard for AI.[72] The International Olympic Committee has also published principles for AI across the Olympic Movement, covering privacy, integrity, fairness, robustness and clear human accountability.[73] CONSENSUS

TABLE 2 Six questions to put to any supplier before signing

QUESTION	WHAT A GOOD ANSWER LOOKS LIKE
What was this trained on?	A named dataset with its population, sexes, events and limitations stated. A refusal to answer is itself an answer.
Where does it perform worst?	A specific answer naming events, populations or conditions. A supplier who claims uniform performance has not measured it.
What does it do when uncertain?	It says so, and declines to recommend. A system that always produces an answer is not measuring its own confidence.
Who owns the athlete record?	The athlete and the organisation, with export in a usable format and a written position on departure.
Does our data train your models?	A clear yes or no, in the contract rather than the sales meeting.
What has been independently validated?	A published or auditable evaluation. Internal benchmarks are marketing until somebody outside the company has checked them.

Compiled by the author, drawing on references 69 to 73. ATHLEET.AI

CHAPTER FOURTEEN

A blueprint for an AI-enabled athletics programme

The analytical distance is short and most programmes can cross it in a season. The organisational distance is longer, and it is the one that decides whether any of this holds.

Bringing the argument together produces a sequence rather than a purchase. A programme that connects its records, insists on explainable output, places a named person between recommendation and athlete, and measures continuity rather than prediction will have done almost everything this paper recommends, and it can begin without buying anything at all.

FIGURE 12 The elite athletics AI maturity model. The test beneath each stage is the question a programme at that stage can answer about its own decisions.

Athleet.AI framework. ATHLEET.AI

Most programmes I have seen sit between the first and second stages, holding good data in incompatible places. The move to the third stage is analytical and can be bought. The move to the fourth and fifth is a change in how the organisation makes and records decisions, which cannot be bought at any price and is the reason so many capable platforms end their first year unused.

14.1 Where a recommendation may act

The governing question for any deployment is which decisions the system may make, which it may propose, and which it may only inform. Figure 13 sets out the framework used across this series, plotting the consequence of an error against the strength of the evidence behind the recommendation.

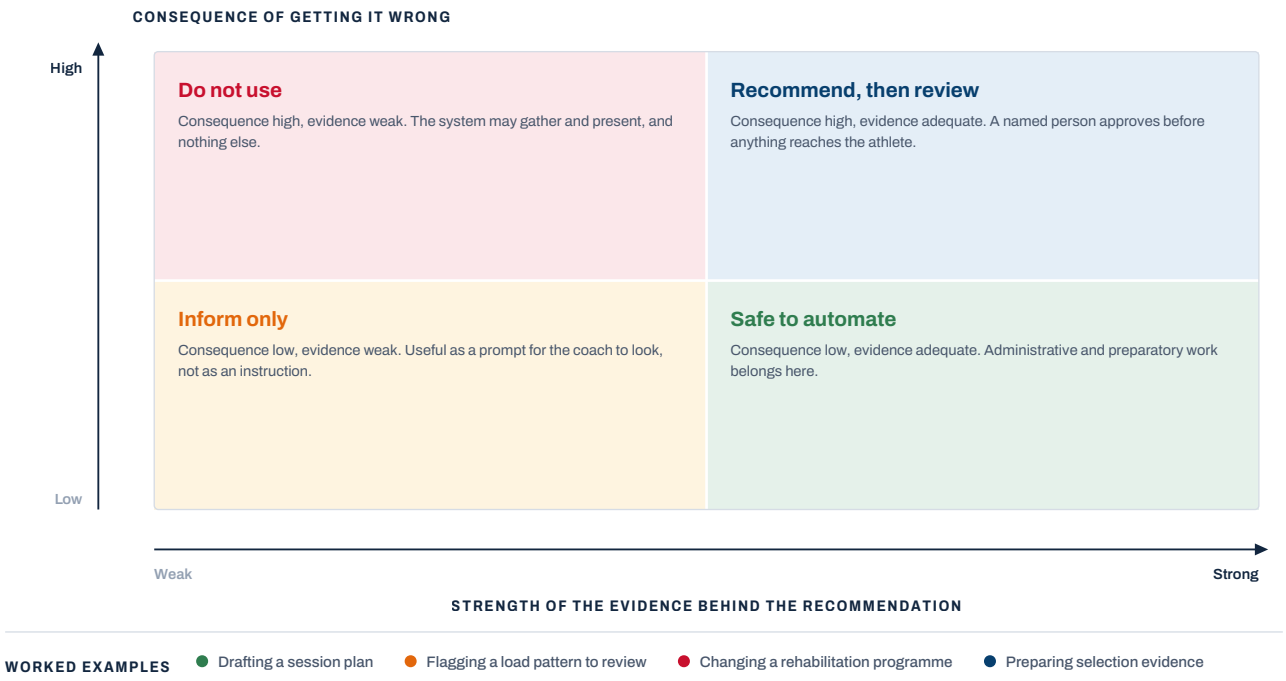


FIGURE 13 Where a recommendation may act, and where it may only inform. Most of what is currently sold to sport sits in the upper-left quadrant and is marketed as though it sat in the upper-right.

Athleet.AI framework, used across the whole series. ATHLEET.AI

14.2 A 90-day proof of value

A pilot should be small enough to complete and specific enough to fail honestly. I would run it on one event group, with one lead coach, one agreed set of measures and a written statement of what would count as a disappointing result.

PROPOSED STRUCTURE

- **Days 1 to 20.** Connect existing records for one event group under a single athlete identifier. Record planned and completed work separately. Agree the measures and the disappointing result in writing before anything else begins.
- **Days 21 to 55.** Run drafting and retrieval only. The system prepares weekly reviews, session drafts and meeting evidence. Nothing reaches an athlete without the coach approving it, and every approval is recorded.
- **Days 56 to 80.** Add explainable prompts based on departures from the athlete's own pattern. Each prompt names its sources and states its confidence. Coaches log which prompts were useful and which were noise, and the noise count is reported.
- **Days 81 to 90.** Review against the agreed measures with someone outside the programme in the room. Decide to stop, continue or extend, and write down the reasoning.

Two features of that design matter more than the technology. Declaring in advance what a disappointing result looks like prevents the near-universal habit of redefining success at the end of a pilot. And counting the useless prompts alongside the useful ones is the only honest measure of whether a system is helping, since a tool that produces 40 alerts a week will be switched off by March whatever its accuracy.

14.3 What good looks like in three years

A programme that has done this well will not look dramatically different from the outside. The coaches will still be coaching, the arguments will still happen, and the athletes will still do the work. What will have changed is that a decision made in March can be found in September with its reasoning intact, that a coach joining in year two inherits method rather than folklore, and that nobody is being asked to trust a number whose provenance they cannot establish.

That is a modest description of success and a demanding one to achieve. It is also, on the evidence assembled in this paper, considerably more valuable than the prediction the sport keeps being offered instead.

END MATTER

Implementation checklist

Twelve things to settle before a single recommendation reaches a coach. Most cost nothing and none of them requires a supplier.

- [ ] **One athlete identifier across every system**
  A record that cannot be joined cannot be reasoned across, and this is the most common single blocker in the sport.
- [ ] **Planned and completed work stored separately**
  The difference between the two is the most informative quantity a programme holds, and it disappears the moment the two are merged.
- [ ] **A reason captured with every modification**
  A session shortened for a hamstring complaint and one shortened for an examination look identical in a database and mean different things.
- [ ] **An internal load measure alongside the external one**
  Session rating of perceived exertion costs one question and captures what the work actually demanded of this athlete.
- [ ] **Event-specific load dimensions defined**
  Landing volume, technical complexity, competition density and eccentric exposure, named and counted rather than assumed.
- [ ] **Test reliability recorded with every test**
  So that a change smaller than the measurement noise is never presented as a finding.
- [ ] **A written statement of the squad's own method**
  Event philosophy, session policies and the coach's technical model, held by the system and applied consistently rather than overridden.
- [ ] **Provenance on every recommendation**
  Sources named, assumptions visible, confidence stated, alternatives offered.
- [ ] **A named approver between recommendation and athlete**
  Recorded with the recommendation, at the time, without exception.
- [ ] **A decision record from every performance meeting**
  What was decided, by whom, on what evidence, and what would cause it to be revisited.
- [ ] **A written position on athlete data on departure**
  Agreed while the record is small and uncontroversial rather than when it has become valuable.
- [ ] **An agreed definition of a disappointing pilot result**
  Written down before the pilot starts, by the people who will be tempted to redefine success at the end of it.

END MATTER

Glossary

Terms used in a specific sense in this paper, defined once and held to throughout.

Acute to chronic workload ratio
Recent training load expressed against a longer-term average of the same load. Widely adopted and, on the methodological evidence set out in Chapter 8, unsuited to individual injury prediction.

Availability
The proportion of planned training weeks an athlete completes without interruption from injury or illness. Measurable within a season and used in this paper in preference to injury prediction.

Decision support
Assembling and presenting evidence, options and trade-offs so that a person can decide. Distinct from automation, in which the system acts.

Explainable
A recommendation whose sources, assumptions, confidence and alternatives are visible to the person deciding, in language they can interrogate.

Internal and external load
External load is the work as prescribed, in distance, repetitions or mass. Internal load is what that work demanded of a particular athlete on a particular day, captured through perceived exertion or self-report.

Markerless motion capture
Estimation of body-segment position from video without physical markers. Well validated for lower-limb sagittal kinematics during gait, considerably weaker for transverse-plane rotations.

Performance intelligence
The organised connection of athlete data, coaching knowledge and decision records so that a programme can say what happened, what it meant, and why it chose what it chose.

Provenance
The record of which inputs produced a given output, dated and named, so that a recommendation can be traced back to its evidence.

Responsible AI
Systems designed so that their limits are visible, their outputs are contestable, and accountability for consequences rests with an identified person.

Smarter training
Managing demand, dose, response, decision and durability as one connected problem rather than as separate monitoring exercises.

END MATTER

Method and evidence grading

How the sources in this paper were selected, checked and graded, and what the reader should distrust.

Sources were gathered by searching the peer-reviewed sport science literature, governing-body publications, regulator guidance and primary legislation, with priority given to systematic reviews, consensus statements and methodological criticism over single primary studies. Every reference in the list that follows was checked against the publishing journal, publisher or issuing body, and the bibliographic details were verified independently a second time before publication. Where a detail could not be confirmed, the source was removed rather than approximated.

Several sources that would have strengthened particular arguments were excluded on those grounds, including material on athlete data rights in athletics

specifically, for which no equivalent to football's player data charter appears to exist.

What the grades mean

The tags used through the paper distinguish primary peer-reviewed studies, systematic reviews and meta-analyses, consensus and governing-body positions, methodological commentary, practitioner opinion, and the author's own frameworks. Methodological commentary is graded separately because several of the strongest arguments in this paper are criticisms rather than experiments, and a reader deserves to know which is which.

What to distrust in this paper

Three things. The event-specific intelligence matrix in Chapter 2 and the measurement chart in Chapter 3 are the author's assessments of a literature that others read differently, and they are published in a form designed to invite disagreement. The durability model in Chapter 11 is a model of a mechanism rather than measured data, and the underlying studies establish association rather than causation. And the regulatory position in Chapter 12 describes a timetable that has moved more than once already, so it should be verified against the current official position rather than relied on here.

Conflict of interest

The author is the founder of Athleet.AI, which builds software in the area this paper examines, and the paper nonetheless recommends against several categories of product, including some that would be commercially straightforward to build and sell into high-performance sport. Readers should weigh the argument accordingly. Test the citations rather than the author.

END MATTER

About the author

Paul Forrest is the founder and Chief AI Officer of Athleet.AI. He works with organisations on the deployment of artificial intelligence, having begun in business process reengineering in the 1990s and moved into natural language and machine learning work from 2014, building small, domain-specific models in preference to general-purpose ones.

He is a qualified athletics coach, holding level three qualifications in sprints, hurdles and endurance, and he coaches as a volunteer at club level. That combination of enterprise systems work and time on a track is the reason this series exists, and it is also the reason the papers decline to recommend several things that would be straightforward to sell.

He is the author of *Tales from AI Development Hell* and *AI Made the Decision: Who Answers When Nobody Owns the Outcome?*, and holds an adjunct professorship teaching digital acumen and digital entrepreneurship.

END MATTER

References

Seventy-six sources, each verified twice against the publishing journal, publisher or issuing body.

Training load, monitoring and injury

1. **Gabbett, T. J.** (2016). The training-injury prevention paradox: should athletes be training smarter and harder? *British Journal of Sports Medicine*, 50(5), 273–280. doi:10.1136/bjsports-2015-095788
2. **Hulin, B. T., Gabbett, T. J., Blanch, P., Chapman, P., Bailey, D., & Orchard, J. W.** (2014). Spikes in acute workload are associated with increased injury risk in elite cricket fast bowlers. *British Journal of Sports Medicine*, 48(8), 708–712. doi:10.1136/bjsports-2013-092524
3. **Windt, J., & Gabbett, T. J.** (2017). How do training and competition workloads relate to injury? The workload–injury aetiology model. *British Journal of Sports Medicine*, 51(5), 428–435. doi:10.1136/bjsports-2016-096040
4. **Lolli, L., Batterham, A. M., Hawkins, R., Kelly, D. M., Strudwick, A. J., Thorpe, R., Gregson, W., & Atkinson, G.** (2019). Mathematical coupling causes spurious correlation within the conventional acute-to-chronic workload ratio calculations. *British Journal of Sports Medicine*, 53(15), 921–922. doi:10.1136/bjsports-2017-098110
5. **Wang, C., Vargas, J. T., Stokes, T., Steele, R., & Shrier, I.** (2020). Analyzing activity and injury: lessons learned from the acute:chronic workload ratio. *Sports Medicine*, 50(7), 1243–1254. doi:10.1007/s40279-020-01280-1
6. **Impellizzeri, F. M., Tenan, M. S., Kempton, T., Novak, A., & Coutts, A. J.** (2020). Acute:chronic workload ratio: conceptual issues and fundamental pitfalls. *International Journal of Sports Physiology and Performance*, 15(6), 907–913. doi:10.1123/ijspp.2019-0864
7. **Impellizzeri, F. M., Woodcock, S., Coutts, A. J., Fanchini, M., McCall, A., & Vigotsky, A. D.** (2021). What role do chronic workloads play in the acute to chronic workload ratio? Time to dismiss ACWR and its underlying theory. *Sports Medicine*, 51(3), 581–592. doi:10.1007/s40279-020-01378-6
8. **Nielsen, R. O., Bertelsen, M. L., Møller, M., Hulme, A., Windt, J., Verhagen, E., Mansournia, M. A., Casals, M., & Parner, E. T.** (2018). Training load and structure-specific load: applications for sport injury causality and data analyses. *British Journal of Sports Medicine*, 52(16), 1016–1017. doi:10.1136/bjsports-2017-097838
9. **Impellizzeri, F. M., Menaspà, P., Coutts, A. J., Kalkhoven, J., & Menaspà, M. J.** (2020). Training load and its role in injury prevention, part I: back to the future. *Journal of Athletic Training*, 55(9), 885–892. doi:10.4085/1062-6050-500-19
10. **Impellizzeri, F. M., McCall, A., Ward, P., Bornn, L., & Coutts, A. J.** (2020). Training load and its role in injury prevention, part 2: conceptual and methodologic pitfalls. *Journal of Athletic Training*, 55(9), 893–901. doi:10.4085/1062-6050-501-19
11. **Bahr, R.** (2016). Why screening tests to predict injury do not work, and probably never will: a critical review. *British Journal of Sports Medicine*, 50(13), 776–780. doi:10.1136/bjsports-2016-096256
12. **Bittencourt, N. F. N., Meeuwisse, W. H., Mendonça, L. D., Nettel-Aguirre, A., Ocarino, J. M., & Fonseca, S. T.** (2016). Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition, narrative review and new concept. *British Journal of Sports Medicine*, 50(21), 1309–1314. doi:10.1136/bjsports-2015-095850
13. **Impellizzeri, F. M., Marcora, S. M., & Coutts, A. J.** (2019). Internal and external training load: 15 years on. *International Journal of Sports Physiology and Performance*, 14(2), 270–273. doi:10.1123/ijspp.2018-0935
14. **Foster, C., Florhaug, J. A., Franklin, J., Gottschall, L., Hrovatin, L. A., Parker, S., Doleshal, P., & Dodge, C.** (2001). A new approach to monitoring exercise training. *Journal of Strength and Conditioning Research*, 15(1), 109–115.
15. **Soligard, T., Schwellnus, M., Alonso, J. M., Bahr, R., Clarsen, B., Dijkstra, H. P., et al.** (2016). How much is too much? (Part 1) International Olympic Committee consensus statement on load in sport and risk of injury. *British Journal of Sports Medicine*, 50(17), 1030–1041. doi:10.1136/bjsports-2016-096581
16. **Schwellnus, M., Soligard, T., Alonso, J. M., Bahr, R., Clarsen, B., Dijkstra, H. P., et al.** (2016). How much is too much? (Part 2) International Olympic Committee consensus statement on load in sport and risk of illness. *British Journal of Sports Medicine*, 50(17), 1043–1052. doi:10.1136/bjsports-2016-096572
17. **Bourdon, P. C., Cardinale, M., Murray, A., Gastin, P., Kellmann, M., Varley, M. C., et al.** (2017). Monitoring athlete training loads: consensus statement. *International Journal of Sports Physiology and Performance*, 12(s2), S2-161–S2-170. doi:10.1123/ijspp.2017-0208
18. **Saw, A. E., Main, L. C., & Gastin, P. B.** (2016). Monitoring the athlete training response: subjective self-reported measures trump commonly used objective measures. A systematic review. *British Journal of Sports Medicine*, 50(5), 281–291. doi:10.1136/bjsports-2015-094758
19. **Plews, D. J., Laursen, P. B., Stanley, J., Kilding, A. E., & Buchheit, M.** (2013). Training adaptation and heart rate variability in elite endurance athletes: opening the door to effective monitoring. *Sports Medicine*, 43(9), 773–781. doi:10.1007/s40279-013-0071-8
20. **Van Eetvelde, H., Mendonça, L. D., Ley, C., Seil, R., & Tischer, T.** (2021). Machine learning methods in sport injury prediction and prevention: a systematic review. *Journal of Experimental Orthopaedics*, 8, 27. doi:10.1186/s40634-021-00346-x
21. **Bullock, G. S., Mylott, J., Hughes, T., Nicholson, K. F., Riley, R. D., & Collins, G. S.** (2022). 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 Medicine*, 52(10), 2469–2482. doi:10.1007/s40279-022-01698-9
22. **Raysmith, B. P., & Drew, M. K.** (2016). Performance success or failure is influenced by weeks lost to injury and illness in elite Australian track and field athletes: a 5-year prospective study. *Journal of Science and Medicine in Sport*, 19(10), 778–783. doi:10.1016/j.jsams.2015.12.515
23. **Drew, M. K., Raysmith, B. P., & Charlton, P. C.** (2017). Injuries impair the chance of successful performance by sportspeople: a systematic review. *British Journal of Sports Medicine*, 51(16), 1209–1214. doi:10.1136/bjsports-2016-096731
24. **Alonso, J. M., Junge, A., Renström, P., Engebretsen, L., Mountjoy, M., & Dvorak, J.** (2009). Sports injuries surveillance during the 2007 IAAF World Athletics Championships. *Clinical Journal of Sport Medicine*, 19(1), 26–32. doi:10.1097/JSM.0b013e318191c8e7
25. **Alonso, J. M., Tscholl, P. M., Engebretsen, L., Mountjoy, M., Dvorak, J., & Junge, A.** (2010). Occurrence of injuries and illnesses during the 2009 IAAF World Athletics Championships. *British Journal of Sports Medicine*, 44(15), 1100–1105. doi:10.1136/bjsm.2010.078030
26. **Alonso, J. M., Edouard, P., Fischetto, G., Adams, B., Depiesse, F., & Mountjoy, M.** (2012). Determination of future prevention strategies in elite track and field: analysis of Daegu 2011 IAAF Championships injuries and illnesses surveillance. *British Journal of Sports Medicine*, 46(7), 505–514. doi:10.1136/bjsports-2012-091008
27. **Edouard, P., Richardson, A., Navarro, L., Gremeaux, V., Branco, P., & Junge, A.** (2019). Relation of team size and success with injuries and illnesses during eight international outdoor athletics championships. *Frontiers in Sports and Active Living*, 1, 8. doi:10.3389/fspor.2019.00008
28. **Edouard, P., Branco, P., & Alonso, J. M.** (2016). Muscle injury is the principal injury type and hamstring muscle injury is the first injury diagnosis during top-level international athletics championships between 2007 and 2015. *British Journal of Sports Medicine*, 50(10), 619–630. doi:10.1136/bjsports-2015-095559

Planning, measurement and performance science in athletics

1. **Kiely, J.** (2012). Periodization paradigms in the 21st century: evidence-led or tradition-driven? *International Journal of Sports Physiology and Performance*, 7(3), 242–250. doi:10.1123/ijspp.7.3.242
2. **Kiely, J.** (2018). Periodization theory: confronting an inconvenient truth. *Sports Medicine*, 48(4), 753–764. doi:10.1007/s40279-017-0823-y
3. **Issurin, V. B.** (2010). New horizons for the methodology and physiology of training periodization. *Sports Medicine*, 40(3), 189–206. doi:10.2165/11319770-000000000-00000
4. **Afonso, J., Rocha, T., Nikolaidis, P. T., Clemente, F. M., Rosemann, T., & Knechtle, B.** (2019). A systematic review of meta-analyses comparing periodized and non-periodized exercise programs: why we should go back to original research. *Frontiers in Physiology*, 10, 1023. doi:10.3389/fphys.2019.01023
5. **Moesgaard, L., Beck, M. M., Christiansen, L., Aagaard, P., & Lundbye-Jensen, J.** (2022). Effects of periodization on strength and muscle hypertrophy in volume-equated resistance training programs: a systematic review and meta-analysis. *Sports Medicine*, 52(7), 1647–1666. doi:10.1007/s40279-021-01636-1

6. **Samozino, P., Rabita, G., Dorel, S., Slawinski, J., Peyrot, N., Saez de Villarreal, E., & Morin, J. B.** (2016). A simple method for measuring power, force, velocity properties, and mechanical effectiveness in sprint running. *Scandinavian Journal of Medicine & Science in Sports*, 26(6), 648–658. doi:10.1111/sms.12490
7. **Šarabon, N., Kozinc, Ž., Ramos, A. G., Knežević, O. M., Čoh, M., & Mirkov, D. M.** (2021). Reliability of sprint force-velocity-power profiles obtained with KiSprint system. *Journal of Sports Science and Medicine*, 20(2), 357–364. doi:10.52082/jssm.2021.357
8. **Ettema, G.** (2024). The force–velocity profiling concept for sprint running is a dead end. *International Journal of Sports Physiology and Performance*, 19(1), 88–91. doi:10.1123/ijspp.2023-0110
9. **Clavel, P., Leduc, C., Morin, J. B., Owen, C., Samozino, P., Peeters, A., Buchheit, M., & Lacome, M.** (2022). Concurrent validity and reliability of sprinting force-velocity profile assessed with GPS devices in elite athletes. *International Journal of Sports Physiology and Performance*, 17(10), 1527–1531. doi:10.1123/ijspp.2021-0339
10. **Kanko, R. M., Laende, E. K., Davis, E. M., Selbie, W. S., & Deluzio, K. J.** (2021). Concurrent assessment of gait kinematics using marker-based and markerless motion capture. *Journal of Biomechanics*, 127, 110665. doi:10.1016/j.jbiomech.2021.110665
11. **Scataglini, S., Abts, E., Van Bocxlaer, C., Van den Bussche, M., Meletani, S., & Truijen, S.** (2024). Accuracy, validity and reliability of markerless camera-based 3D motion capture systems versus marker-based 3D motion capture systems in gait analysis: a systematic review and meta-analysis. *Sensors*, 24(11), 3686. doi:10.3390/s24113686
12. **Uhlrich, S. D., Falisse, A., Kidziński, Ł., Muccini, J., Ko, M., Chaudhari, A. S., Hicks, J. L., & Delp, S. L.** (2023). OpenCap: human movement dynamics from smartphone videos. *PLOS Computational Biology*, 19(10), e1011462. doi:10.1371/journal.pcbi.1011462
13. **Claudino, J. G., Cronin, J., Mezêncio, B., McMaster, D. T., McGuigan, M., Tricoli, V., Amadio, A. C., & Serrão, J. C.** (2017). The countermovement jump to monitor neuromuscular status: a meta-analysis. *Journal of Science and Medicine in Sport*, 20(4), 397–402. doi:10.1016/j.jsams.2016.08.011
14. **Gathercole, R., Sporer, B., Stellingwerff, T., & Sleivert, G.** (2015). Alternative countermovement-jump analysis to quantify acute neuromuscular fatigue. *International Journal of Sports Physiology and Performance*, 10(1), 84–92. doi:10.1123/ijspp.2013-0413
15. **Moreno-Villanueva, A., Pino-Ortega, J., & Rico-González, M.** (2024). Validity and reliability of linear position transducers and linear velocity transducers: a systematic review. *Sports Biomechanics*, 23(10), 1340–1369. doi:10.1080/14763141.2021.1988136
16. **Marston, K. J., Forrest, M. R. L., Teo, S. Y. M., Mansfield, S. K., Peiffer, J. J., & Scott, B. R.** (2022). Load-velocity relationships and predicted maximal strength: a systematic review of the validity and reliability of current methods. *PLOS ONE*, 17(10), e0267937. doi:10.1371/journal.pone.0267937
17. **Bosquet, L., Montpetit, J., Arvisais, D., & Mujika, I.** (2007). Effects of tapering on performance: a meta-analysis. *Medicine and Science in Sports and Exercise*, 39(8), 1358–1365. doi:10.1249/mss.0b013e31806010e0
18. **Mujika, I., & Padilla, S.** (2003). Scientific bases for precompetition tapering strategies. *Medicine and Science in Sports and Exercise*, 35(7), 1182–1187. doi:10.1249/01.MSS.0000074448.73931.11
19. **Racinais, S., Alonso, J. M., Coutts, A. J., Flouris, A. D., Girard, O., González-Alonso, J., et al.** (2015). Consensus recommendations on training and competing in the heat. *British Journal of Sports Medicine*, 49(18), 1164–1173. doi:10.1136/bjsports-2015-094915
20. **Racinais, S., Ihsan, M., Taylor, L., Cardinale, M., Adami, P. E., Alonso, J. M., et al.** (2021). Hydration and cooling in elite athletes: relationship with performance, body mass loss and body temperatures during the Doha 2019 IAAF World Athletics Championships. *British Journal of Sports Medicine*, 55(23), 1335–1341. doi:10.1136/bjsports-2020-103613
21. **Galan-Lopez, N., Esh, C. J., Leal, D. V., Gandini, S., Lucas, R., Garrandes, F., et al.** (2023). Heat preparation and knowledge at the World Athletics Race Walking Team Championships Muscat 2022. *International Journal of Sports Physiology and Performance*, 18(8), 813–824. doi:10.1123/ijspp.2022-0446
22. **Hollings, S. C., Hopkins, W. G., & Hume, P. A.** (2014). Age at peak performance of successful track & field athletes. *International Journal of Sports Science & Coaching*, 9(4), 651–661. doi:10.1260/1747-9541.9.4.651
23. **Boccia, G., Brustio, P. R., Moisè, P., Franceschi, A., La Torre, A., Schena, F., Rainoldi, A., & Cardinale, M.** (2019). Elite national athletes reach their peak performance later than non-elite in sprints and throwing events. *Journal of Science and Medicine in Sport*, 22(3), 342–347. doi:10.1016/j.jsams.2018.08.011
24. **Boccia, G., Cardinale, M., & Brustio, P. R.** (2021). World-class sprinters' careers: early success does not guarantee success at adult age. *International Journal of Sports Physiology and Performance*, 16(3), 367–374. doi:10.1123/ijspp.2020-0090
25. **Boccia, G., Cardinale, M., & Brustio, P. R.** (2021). Elite junior throwers unlikely to remain at the top level in the senior category. *International Journal of Sports Physiology and Performance*, 16(9), 1281–1287. doi:10.1123/ijspp.2020-0699
26. **Boccia, G., Cardinale, M., & Brustio, P. R.** (2021). Performance progression of elite jumpers: early performances do not predict later success. *Scandinavian Journal of Medicine & Science in Sports*, 31(1), 132–139. doi:10.1111/sms.13819
27. **Brustio, P. R., Stival, M., & Boccia, G.** (2023). Relative age effect reversal on the junior-to-senior transition in world-class athletics. *Journal of Sports Sciences*, 41(9), 903–909. doi:10.1080/02640414.2023.2245647
28. **Barth, M., Güllich, A., Macnamara, B. N., & Hambrick, D. Z.** (2024). Quantifying the extent to which junior performance predicts senior performance in Olympic sports: a systematic review and meta-analysis. *Sports Medicine*, 54(1), 95–104. doi:10.1007/s40279-023-01906-0
29. **Haugen, T., Seiler, S., Sandbakk, Ø., & Tønnessen, E.** (2019). The training and development of elite sprint performance: an integration of scientific and best practice literature. *Sports Medicine – Open*, 5, 44. doi:10.1186/s40798-019-0221-0
30. **Fullagar, H. H. K., McCall, A., Impellizzeri, F. M., Favero, T., & Coutts, A. J.** (2019). The translation of sport science research to the field: a current opinion and overview on the perceptions of practitioners, researchers and coaches. *Sports Medicine*, 49(12), 1817–1824. doi:10.1007/s40279-019-01139-0
31. **Bartlett, J. D., & Drust, B.** (2021). A framework for effective knowledge translation and performance delivery of sport scientists in professional sport. *European Journal of Sport Science*, 21(11), 1579–1587. doi:10.1080/17461391.2020.1842511
32. **Bullock, G. S., Hughes, T., Arundale, A. H., Ward, P., Collins, G. S., & Kluzek, S.** (2022). 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 Medicine*, 52(8), 1729–1735. doi:10.1007/s40279-022-01655-6

Governance, explainability, regulation and bias

1. **European Parliament and Council of the European Union** (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*, OJ L, 2024/1689, 12 July 2024. In force 1 August 2024, applying in stages.
2. **Information Commissioner's Office & The Alan Turing Institute** (2020). *Explaining decisions made with AI.* Wilmslow, ICO. Guidance in three parts, covering the basics, practice and organisational responsibilities.
3. **United Kingdom General Data Protection Regulation**, Article 9. Processing of special categories of personal data, including data concerning health and biometric data used for unique identification.
4. **Data (Use and Access) Act 2025** (c. 18), section 80. Will replace Article 22 of the UK GDPR with new Articles 22A to 22D on automated decision-making. Royal Assent 19 June 2025; section 80 not yet commenced at the date of this paper.
5. **Rudin, C.** (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1, 206–215. doi:10.1038/s42256-019-0048-x
6. **Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S.** (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. doi:10.1126/science.aax2342
7. **Cowley, E. S., Olenick, A. A., McNulty, K. L., & Ross, E. J.** (2021). "Invisible sportswomen": the sex data gap in sport and exercise science research. *Women in Sport and Physical Activity Journal*, 29(2), 146–151. doi:10.1123/wspaj.2021-0028
8. **Costello, J. T., Bieuzen, F., & Bleakley, C. M.** (2014). Where are all the female participants in sports and exercise medicine research? *European Journal of Sport Science*, 14(8), 847–851. doi:10.1080/17461391.2014.911354
9. **Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T.** (2019). Model cards for model reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229. doi:10.1145/3287560.3287596
10. **Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K.** (2021). Datasheets for datasets. *Communications of the ACM*, 64(12), 86–92. doi:10.1145/3458723
11. **National Institute of Standards and Technology** (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, NIST AI 100-1. Gaithersburg, NIST. Published 26 January 2023.
12. **International Organization for Standardization** (2023). *ISO/IEC 42001:2023. Information technology, artificial intelligence, management system.* Geneva, ISO.
13. **International Olympic Committee** (2024). *Olympic AI Agenda.* Lausanne, IOC. Principles covering privacy, integrity, fairness, robustness and human accountability across the Olympic Movement.
14. **FIFPRO & FIFA** (2022). *Charter of Player Data Rights.* Hoofddorp, FIFPRO. Published 19 September 2022, setting out eight rights over performance and personal data.
15. **Gale, K.** (2016). The sports industry's new power play: athlete biometric data domination. Who owns it and what may be done with it? *Arizona State University Sports and Entertainment Law Journal*, 6(1), 7–46.
16. **Doshi-Velez, F., & Kim, B.** (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint* arXiv:1702.08608.

WHERE THIS GOES NEXT

A high-performance programme pilot

The argument in this paper is testable, and the fastest way to test it is on one event group for one season with measures agreed in advance.

Athleet.AI works with national federations and high-performance programmes to run a defined pilot on a single event group, using the structure set out in Chapter 14. The platform is an event-aware orchestration layer connecting plans, sessions, testing, video, athlete feedback and the coach's own recorded knowledge, and every recommendation it produces carries its sources, its confidence and its alternatives.

What a pilot involves

- One event group, one lead coach, one agreed set of measures and a written definition of a disappointing result.
- Existing records connected under a single athlete identifier, with planned and completed work stored separately.
- Drafting and retrieval first, explainable prompts second, and a named approver between every recommendation and every athlete.
- A review at 90 days against availability, completed against planned work, time-loss episodes and progression against individual objectives.
- An honest count of the prompts coaches found useless alongside those they found useful.