



Democratised or Stratified?

AI, Competitive Fairness and the Grassroots Coaching Gap

About this paper

This paper asks whether artificial intelligence in coaching will narrow the distance between elite and grassroots provision or widen it, and what governing bodies can do to influence the outcome. It is written for the boards, performance directors and coach development leads of national governing bodies, and for the club committees and volunteer coaches who will meet the question in practice before any policy reaches them. It assumes no technical background. It does assume some familiarity with how coach education and competition rules are made.

This is the sixth and final paper in the Athleet.AI strand of the working papers. The strand has argued throughout that AI in coaching and athlete development must respect the non-linear, human and developmental nature of athletic progress. Here that argument meets the question of distribution. A tool that respects development is of limited public value if only the best-resourced programmes can use it well, and a tool that reaches every coach becomes a hazard at scale if it encodes a narrow view of how young athletes grow.

Editorial balance. Roughly two fifths of the paper reports published evidence and official documents, about two fifths is my own analysis of how that evidence bears on sport, and the remaining fifth, concentrated in the final chapter, sets out a framework of options for governing bodies that I propose and that nobody has yet tested. Where the paper describes regulatory or rule-making material, it reports what the documents say and offers no view on the law.

Legal and regulatory notice. Parts of this paper describe legislation, regulation, regulator activity or court decisions. 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 I draw a practical implication that a source does not state, the text says so.

HOW TO READ THE EVIDENCE TAGS

Every substantive claim carries a tag stating what kind of evidence supports it. Two tags are combined where a claim rests on both.

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

REVIEW A systematic review, meta-analysis or narrative review.

CONSENSUS A governing body position, consensus statement or official guidance.

PREPRINT A working paper or preprint not yet peer reviewed.

INDUSTRY A company announcement, press report or industry source.

SCHOLARLY A scholarly book, edited volume or law journal article.

PRACTICE Practitioner opinion, including the author's own.

EVIDENCE GAP The available evidence does not answer the question.

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

CONTENTS

Contents

FRONT MATTER

A note from the author	5
Executive summary	6

PART ONE · THE TERMS OF THE QUESTION

1	Two programmes in the same week	9
2	Technology, fairness and the governing body	11
3	The grassroots coaching gap	13

PART TWO · TWO FORCES AT ONCE

4	The democratisation case	16
5	The stratification case	18

PART THREE · GOVERNING WHAT YOU DO NOT OWN

6	Institutional capability and lag	22
7	Choices for governing bodies	24

END MATTER

References	27
About the author	28

ILLUSTRATIVE MATERIAL

Every organisation, person and scenario described as illustrative in this paper is invented. Published research and official documents are cited and numbered, and invented material is never presented as evidence.

Research closed 24 July 2023. Sources and events are as they stood at that date.

CITATION

Forrest, P. (2023) *Democratised or Stratified? AI, Competitive Fairness and the Grassroots Coaching Gap*. Athleet.AI Working Papers, AI in Coaching and Athlete Development, Paper 6. Athleet.AI, 26 July 2023. Available at: athleet.ai.

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

PAPERS IN THIS STRAND

- Paper one. *Predicting the Unpredictable*. Why Algorithmic Talent Identification Struggles With Non-Linear Athlete Development. 25 January 2023.
- Paper two. *The Coach in the Loop*. Tacit Knowledge, Professional Judgement and the Risk of Deskilled Coaching. 27 February 2023.
- Paper three. *The Measured Child*. Datafication, Consent and Autonomy Across the Athlete Pathway. 19 April 2023.
- Paper four. *Maturation, Bias and the Training Set*. Developmental Fairness in AI-Supported Athlete Development. 24 May 2023.
- Paper five. *Load, Readiness and False Precision*. The Limits of Algorithmic Training Prescription in Long-Term Endurance Development. 21 June 2023.
- **Paper six. *Democratised or Stratified? AI, Competitive Fairness and the***

Grassroots Coaching Gap. 26 July
2023.

A NOTE FROM THE AUTHOR

Why the village club keeps coming back to me

I have spent a good part of my working life on both sides of this question without always noticing that they were the same question. In advisory work I sit with organisations deciding what to buy, what to build and how to govern systems they cannot see inside. On a track on a weekday evening I am one of the many volunteers trying to put together a sensible session for a group whose ages, abilities and attendance change from week to week, armed with a stopwatch and whatever I took from my last course.

When general-purpose chatbots reached the public late last year, both halves of that life reacted at once. The volunteer in me saw a patient assistant that would help draft a progression at ten o'clock at night without complaint. The adviser in me saw a service whose behaviour could change without notice, owned by companies with no particular interest in athletics, and sold into a market where the programmes with the most data and the most staff would get the most out of it. I do not think either reaction is wrong, and this paper is my attempt to hold the two together long enough to say something useful.

I should be plain about the limits of what follows. I have found no evidence that AI tools improve coaching outcomes at grassroots level, and I have found nobody who claims to have it. What I do have is a strong impression, from watching institutions respond to earlier technologies, that the choices made in the first few years tend to set the pattern for the following decade, and that sport is at risk of making those choices by default. If this paper persuades one coach development lead to put the question on a board agenda before the market has answered it, it will have done what I wanted.

Paul Forrest
Founder, Athleet.AI
26 July 2023

TWO PAGES THAT CIRCULATE ON THEIR OWN

Executive summary

AI could give every volunteer coach something close to expert planning support, and it could also hand the best-resourced programmes a further advantage that compounds over time. Both effects are likely to arrive together, and which one dominates will depend on decisions governing bodies make in the next few years about coach education, competition rules and their own capacity to judge tools they do not own.

The early productivity evidence gives the democratisation case real support, because two studies of generative AI at work found the largest gains among less experienced or lower-performing workers. Sport's own history supports the opposite case. Technology has repeatedly moved faster than governing bodies, and its benefits have tended to reach those who could afford them first, so my conclusion is that leaving AI access to the market will probably produce stratification, although governing bodies already hold the precedents and levers they would need to push the other way.

1

The grassroots coaching workforce is large, mostly unpaid and thinly supported.

Sport England's coaching plan estimated that 3.1 million adults in England had coached sport or physical activity in the previous year, and that 74% of coaches did so voluntarily.¹ **CONSENSUS** The barriers reported by people from under-represented groups, including cost, scarce mentoring and isolation, are the gaps an AI assistant would be marketed to fill.

3

Sport has a long record of technology outrunning its governing bodies.

A systematic review found 31 controversial cases of sports technology across three decades and proposed that long periods without intervention by a governing body tend to increase the peer-reviewed attention paid to such cases.⁴ **REVIEW** AI used in preparation is likely to follow that pattern, and it will be harder to see.

5

The models beneath many coaching tools are not stable.

A preprint posted on 18 July 2023 reported that GPT-4's accuracy on a prime-number identification task fell from 97.6% in the March 2023 version to 2.4% in the June 2023 version.⁶ **PREPRINT** A tool approved once is a moving target, and few governing bodies employ staff who would notice it move.

2

Early evidence suggests generative AI helps weaker performers most.

In a preregistered experiment with 453 professionals, access to ChatGPT reduced average time taken by 40%, raised output quality by 18% and reduced inequality between workers.² **PEER-REVIEWED** A field study of 5,179 customer support agents found the greatest gains among novice and low-skilled workers.³ **PREPRINT** Neither study involved coaching, and whether the effect carries over is untested.

EVIDENCE GAP

4

Governing bodies already have a precedent for acting on unequal access.

World Athletics announced in January 2020 that, from 30 April 2020, a shoe would need to have been on open retail sale for four months before use in competition, and that a shoe not openly available would be treated as a prototype and not permitted.⁵ **CONSENSUS** The logic of that rule, which ties legitimacy in competition to availability, could be extended to proprietary AI systems used in competition. **ATHLEET.AI**

6

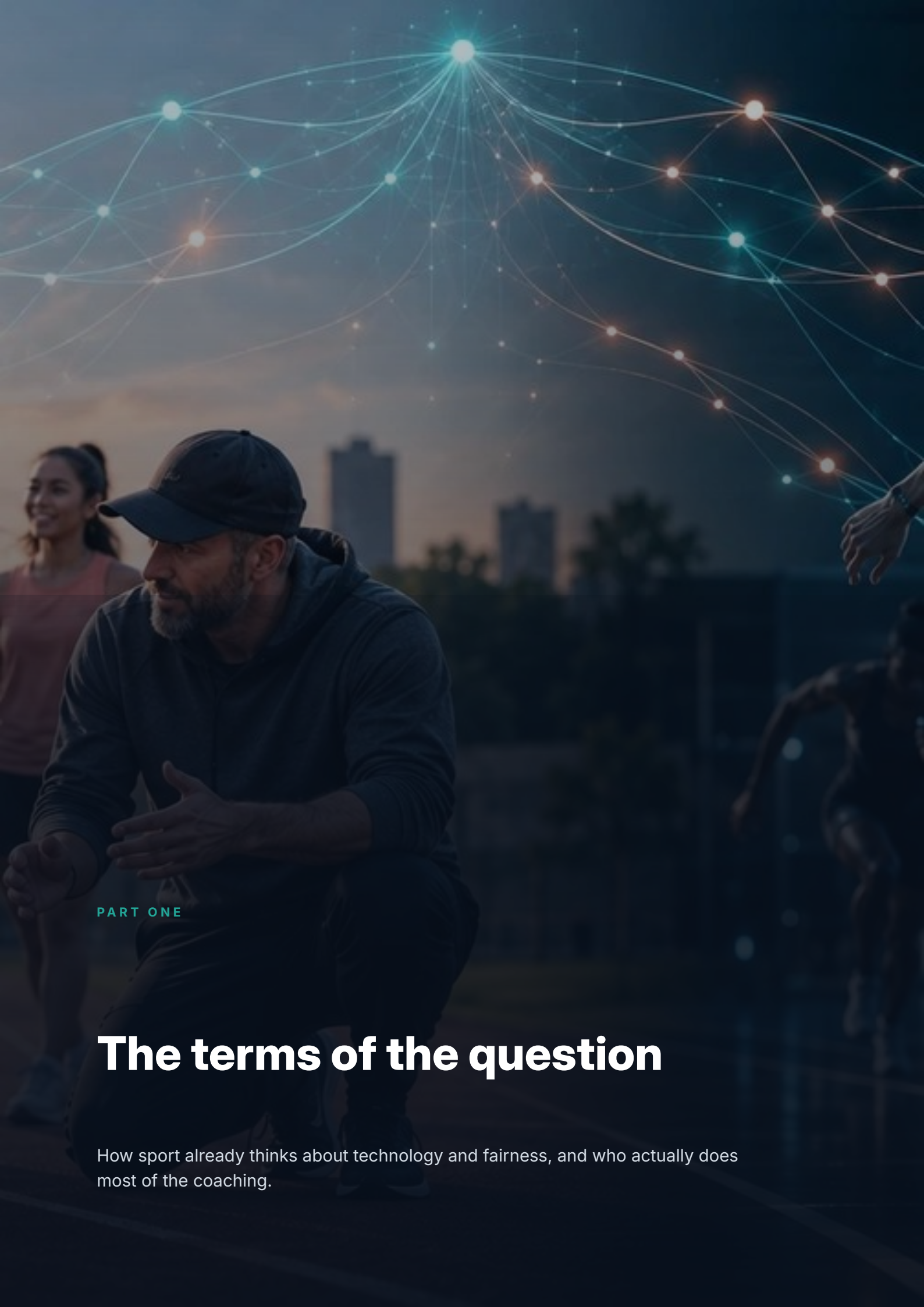
General AI policy has, so far, nothing specific to say about sport.

The United States secured voluntary commitments from seven AI companies on 21 July 2023,⁷ the European Parliament adopted its position on the proposed AI Act on 14 June 2023,⁸ and the UK government proposed a context-specific approach led by existing regulators in March 2023.⁹ **CONSENSUS** I have found nothing in any of these documents that addresses coaching or competition, which leaves governing bodies to set their own terms. **EVIDENCE GAP**

What follows for a club or governing body

A governing body should treat access to AI support as part of coach development and decide whether licensed volunteer coaches receive a shared and governed tool. It should decide, before a dispute forces the issue, at what point a proprietary AI system becomes an equipment question under its competition rules, and it should build or pool enough technical capability to assess tools it did not build and cannot inspect.

Clubs can ask their governing body for a lead before adopting whatever arrives first, and can insist that any tool their volunteers use comes with a mentor behind it.



PART ONE

The terms of the question

How sport already thinks about technology and fairness, and who actually does most of the coaching.

CHAPTER ONE

Two programmes in the same week

Two groups of athletes will train with the help of AI this week. The distance between their experiences is the subject of this paper.

The question this paper addresses is simple to state and hard to answer. Within months, generative AI has put a capable text assistant within reach of almost any volunteer coach. The same wave of technology is being absorbed, more quietly and at much greater depth, by performance programmes that hold years of athlete data and employ the specialists to interpret it. Whether the net result is a narrower gap between elite and grassroots coaching or a wider one is not yet determined, and I argue that institutional choices will largely decide it.

ILLUSTRATIVE SCENARIO

Two programmes, one week

On Monday, the performance analyst at a national sprint programme loads a season of timing-gate, force-plate and wellbeing data into a model the programme commissioned last year, part-trained on its own archive of former athletes. By Wednesday a biomechanist and the relay coach have checked the model's suggestions against video, discarded two of them, and adjusted a block of speed endurance work for three athletes returning from minor injuries. On Thursday evening, about 40 miles away, a volunteer at a village athletics club types a question into a free chatbot on her phone. She has 14 athletes aged 11 to 15 arriving within the hour, several of whom have grown noticeably since Easter, and she wants a session that will not overload them. The chatbot returns a tidy and confident plan. It has never met the athletes and it cannot see the track, and nothing in its answer warns her that one of the plyometric drills is a poor choice for a child in a growth spurt. The club's only qualified coach moved away in the spring, so there is nobody for her to ask.

Both coaches in the scenario are using AI, and it would be easy to read the story as the familiar gap between rich and poor programmes with some software attached. That reading is partly right. What it misses is that the volunteer's position has also changed, because until recently she would have had no assistant at all, and the plan on her phone may well be better than the one she would have written alone. The same technology has moved both coaches forward. The question is how far, in what direction and with what risk.

1.1 The question sport has not yet asked

Sport has argued about technology and fairness for decades, over swimsuits, prosthetic limbs, and golf equipment,⁴ **REVIEW** and most recently over running shoes.⁵

CONSENSUS It has also, separately, worried about the state of its grassroots coaching workforce, which in England is overwhelmingly voluntary and reports thin support.¹

CONSENSUS These two conversations have not yet met on the subject of AI. I have found no national governing body in the United Kingdom that has published a posi-

tion on whether AI coaching support should be treated as part of coach development, or on whether proprietary AI systems have any bearing on fair competition.

EVIDENCE GAP Given how recently the tools arrived, that silence is understandable, though it will not stay harmless for long.

1.2 What this paper argues

Two terms need fixing before the argument can proceed. By democratisation I mean a reduction in the difference between the quality of coaching support available to a well-resourced programme and that available to a volunteer coach, measured in what reaches the athlete. By stratification I mean the opposite, a widening of that difference, including the case where everyone improves but the best-resourced improve by more.

My argument is that both forces will operate at once, because the technology that makes a volunteer more capable also rewards those who already hold data, specialist staff and money. Which force dominates depends on four choices governing bodies can make now, namely whether AI access becomes part of coach education, whether competition rules address data and models, whether grassroots coaches get shared and governed tools, and whether governing bodies build the capability to assess what they approve. Left to the market, my expectation is stratification.

CHAPTER TWO

Technology, fairness and the governing body

Sport already has a theory of which inequalities between competitors matter, and a long record of acting on them late.

Two sources frame this chapter: Loland's philosophical account says which inequalities between competitors matter, and Dyer's systematic review shows how governing bodies have handled contested technology, and between them they explain why AI poses an unfamiliar version of an old problem.

2.1 Fair play as a theory of relevant inequality

Loland treats fair play as a system of moral norms, and at its centre sits the idea that competitors should have a fair opportunity to perform.¹⁰ **SCHOLARLY** On his account, the purpose of a contest is to measure and compare the performances the rules define as relevant, so inequalities that bear on those performances are the point of the exercise, while inequalities that athletes cannot influence and that the contest does not set out to measure are candidates for elimination or compensation. Sex categories, weight classes and age groups are the familiar institutional responses. Loland also discusses inequalities in what he calls system strength, meaning the financial, technological and scientific resources that stand behind an athlete, and treats them as a difficult case for his principle. **SCHOLARLY**

Read against that framework, a proprietary AI system looks much more like system strength than like talent or effort. An 18-year-old in a national programme did not earn the programme's data science team, and a 15-year-old at a village club did not choose to be coached by a volunteer with a free chatbot. My own view, which goes beyond anything Loland wrote, is that AI increases the share of performance that system strength can buy, because it turns data and specialist time into better-targeted training at a rate that was not previously possible. **ATHLEET.AI** If that is right, the old hard case becomes harder.

2.2 A long record of late intervention

Dyer's review searched the peer-reviewed literature for case studies of controversy around sports technology and found 56 articles covering 31 cases over three decades.⁴ **REVIEW** Golf supplied the largest number of cases, while the most attention went to Oscar Pistorius's use of prosthetic legs in athletics and to full-length swimsuits in swimming. Academic interest peaked between 2008 and 2010. The finding that matters most here is Dyer's proposal that long periods without intervention or resolution by a governing body tend to increase the peer-reviewed attention paid to such cases. **REVIEW** My own inference is that delay has a cost, paid in controversy and in the credibility of the governing body. **ATHLEET.AI**

AI differs from every case in that review in one respect that makes delay more likely. A swimsuit can be seen on the blocks and a shoe can be measured at a call room, whereas an AI system used in preparation leaves no physical trace at the competition itself. Its effect is felt in training, selection and tactical preparation, none of which a technical official inspects. (Tools used during an event are the exception, and I return to them in the final chapter.) My inference is that governing bodies will not be forced to act on AI by a visible incident in the way they were forced to act on swimsuits, and that without deliberate attention the period of non-intervention Dyer warns about could be long. **ATHLEET.AI**

2.3 The shoe rules as precedent

The most recent large-scale intervention on equipment offers a useful model. On 31 January 2020 World Athletics announced modified rules on shoes, recommended by an internal working group that included technical, scientific and legal experts as well as athlete representatives.⁵ **CONSENSUS** The announcement placed an indefinite moratorium on any shoe with a sole thicker than 40mm or containing more than one rigid embedded plate or blade. It also provided that, from 30 April 2020, a shoe must have been available for purchase by any athlete on the open retail market for four months before it could be used in competition, and that a shoe not openly available would be deemed a prototype and not permitted. **CONSENSUS**

The availability provision is the part that transfers. It treats access as a condition of fair use, so that an innovation is acceptable in competition once every athlete could, in principle, obtain it. Applied to AI, the same reasoning would ask whether a system used in competition is available to all competitors on reasonable terms, and would treat a bespoke system available only to one programme as a prototype. The analogy has limits, which I would not want to understate. World Athletics was regulating what athletes wear in a race, and no governing body regulates who a programme may hire, what its physiologists may read or how its coaches plan a training block. Most AI use falls on that unregulated side of the line, and the precedent only reaches the part of AI use that touches competition directly.

SHOES AND SOFTWARE

The shoe rules governed an object that officials could see and measure. An AI system used in training has no equivalent inspection point, which is why a disclosure rule may be the more realistic first step.

CHAPTER THREE

The grassroots coaching gap

Most coaching in England is done by volunteers, and the support the coaching plan says many would-be coaches lack is close to the support AI tools claim to offer.

3.1 A mostly volunteer workforce

The most detailed public picture of the coaching workforce in England comes from Sport England's coaching plan for 2017 to 2021, which drew on a nationally representative YouGov study commissioned by Sports Coach UK in October 2016.¹ **CONSENSUS** It estimated that 3.1 million adults in England had coached sport or physical activity in the previous 12 months, that 3.3 million children received coaching each year, and that coaches delivered about 7.6 million coach-led participation experiences each week, two thirds of them in sports clubs. Of those coaches, 74% coached voluntarily, 14% received some payment for part-time coaching and 12% made their living from coaching full time. **CONSENSUS**

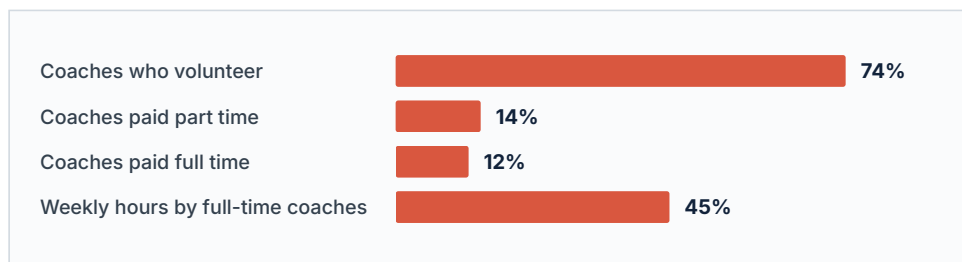


FIGURE 1 Employment status of coaches in England, and the share of weekly coaching hours delivered by full-time coaches. Source: Sport England (2016), drawing on a YouGov study for Sports Coach UK, October 2016.

The plan added that the 12% of coaches working full time delivered 45% of weekly coaching hours, and it acknowledged that more work was needed to understand what those coaches do and with whom.¹ **CONSENSUS** It also reported that 53% of coaches came from higher socio-economic groups against 25% from lower groups. Only 17% of qualified coaches were female. The figures are now several years old and predate the pandemic, so they indicate the structure of the workforce, while the counts themselves will have moved. My reading, which the plan does not state, is that full-time coaching is likely to be concentrated in performance, school and commercial settings, which would leave the volunteer majority carrying most of the community club coaching in which children first meet a sport. **ATHLEET.AI**

3.2 What the development evidence asks of coaches

Earlier papers in this strand argued that athletic development is non-linear and that early performance is a poor guide to later success. Two reviews underline why that matters for the grassroots coach in particular. Rees and colleagues reviewed what was known about the development of the world's best sporting talent under three head-

ings (the performer, the environment, and practice and training), rated the quality of evidence in each area, and concluded that future work needed to address the complex interactions they judged highly likely to exist across those domains.¹¹ **REVIEW** Till and Baker described talent as a complex and largely misunderstood phenomenon that lacks robust research evidence, pointed to the need for appropriate resourcing in the system to support the continued development of knowledge, and did not recommend talent identification and selection at younger ages.¹² **REVIEW**

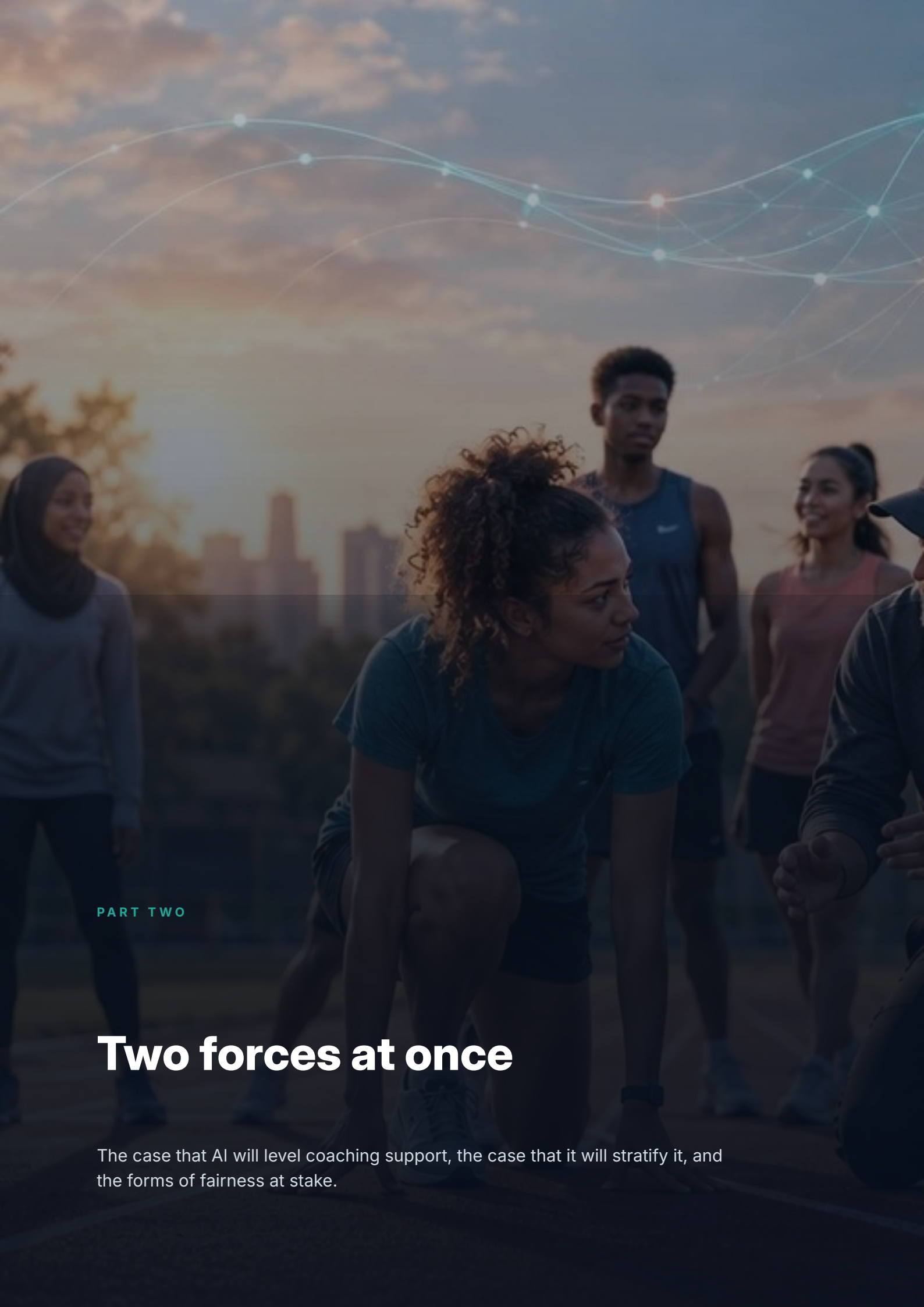
The practical consequence is that the coaching with the longest reach into an athlete's future is often delivered by people with the least support. A volunteer who keeps a 12-year-old in the sport, varies the child's practice and resists the urge to pick winners early is doing the work the evidence values. A tool that nudged that volunteer towards early selection, narrow specialisation or adult training loads would do harm at scale, and it would do so quietly, since the effects would appear years later and in athletes who had already left the sport. That is the developmental risk inside the democratisation case, and it recurs throughout what follows.

3.3 Where support runs out

The coaching plan also listed the barriers that people from under-represented groups reported when engaging with coaching.¹ **CONSENSUS** They included the cost of qualifications; a lack of mentoring, peer observation and experiential learning; a lack of continuing development alongside one-off courses; inaccessible and infrequent courses; a lack of support for coaches not connected to a club, resulting in isolation; and the format of assessment. Several of these read almost as a specification for an AI assistant, since a chatbot is available at any hour, costs little or nothing and asks no one to travel.

An assistant is not a mentor. The barriers recorded in the coaching plan are largely about isolation and access to experienced people. A chatbot can answer at any hour. It cannot, however, watch a session, know the athletes or carry any responsibility for its advice. An AI offer to grassroots coaches that deepens their isolation would meet only the informational half of the need.

Mentoring, peer observation and experiential learning are social activities, in which a less experienced coach is watched and corrected by someone who knows the context. An assistant that answers questions leaves that relational part untouched, and there is a plausible risk that easy answers will reduce the pressure on clubs and governing bodies to provide the people.

A group of diverse runners are on a track at sunset. In the foreground, a woman with curly hair is in a starting crouch. Behind her, a man and another woman are standing. To the left, a woman in a hijab is walking. On the right, a man in a cap is partially visible. The background shows a city skyline under a sunset sky. A digital network of glowing lines and nodes is overlaid on the top half of the image.

PART TWO

Two forces at once

The case that AI will level coaching support, the case that it will stratify it, and the forms of fairness at stake.

CHAPTER FOUR

The democratisation case

The best current evidence suggests that generative AI helps weaker performers more than stronger ones, which is the most persuasive argument that it could narrow the coaching gap.

4.1 What the compression evidence shows

Two studies published this year give the democratisation case its empirical footing. Noy and Zhang ran a preregistered online experiment in which 453 college-educated professionals completed occupation-specific, incentivised writing tasks, with half randomly given access to ChatGPT.² **PEER-REVIEWED** Average time taken fell by 40% and output quality rose by 18%, and inequality between workers decreased. Participants exposed to ChatGPT during the experiment were twice as likely to report using it in their real jobs two weeks later, and 1.6 times as likely two months later. **PEER-REVIEWED**

An earlier working paper version of the same study, released in March 2023, reported that ChatGPT compressed the productivity distribution by benefiting lower-ability workers more, and that it mostly substituted for worker effort, restructuring tasks towards generating ideas and editing and away from rough drafting.¹³ **PREPRINT** That detail matters for coach development, and I come back to it below.

The second study, by Brynjolfsson, Li and Raymond, followed 5,179 customer support agents.³ **PREPRINT** As a generative AI conversational assistant was introduced to them in stages, access to the tool increased productivity, measured as issues resolved per hour, by 14% on average, with the greatest effect on novice and low-skilled workers and minimal effect on experienced and highly skilled ones. The authors offered suggestive evidence that the model spread the potentially tacit knowledge of more able workers to newer ones, and they reported improvements in customer sentiment and employee retention alongside fewer requests for managerial intervention. **PREPRINT**

Writing experiment	Customer support field study	Not tested
• 453 professionals • time taken down 40% • quality up 18% • inequality reduced	• 5,179 agents • 14% more issues resolved • largest gains for novices	• coaching tasks • outcomes over years • effects on children

FIGURE 2 What the two workplace studies found, and what they did not test. Source: Noy and Zhang (2023); Brynjolfsson, Li and Raymond (2023).

4.2 Why coaching might follow, and why it might not

The mechanism Brynjolfsson and colleagues describe is what the volunteer in the opening scenario lacks. She has no experienced colleague to absorb practice from, and a tool that carried some of the tacit knowledge of good coaches into her planning would address a real deficit. Session planning also has features in common with the writing tasks Noy and Zhang studied, since it involves structuring material for a known audience under time pressure. On that basis I think it plausible that AI planning support would raise the floor of grassroots session quality, and that it would raise it most for the least experienced coaches. **ATHLEET.AI**

Four differences counsel caution, and none of them has yet been studied. **EVIDENCE GAP** First, the quality of coaching shows itself in athlete development over years, whereas the writing tasks were graded within hours, so a tool could make plans that look better and develop athletes no better. Second, the customer support assistant was trained on records of past conversations and their outcomes, a feedback loop grassroots sport does not have, because nobody records whether a Thursday session worked. Third, the finding that ChatGPT mostly substituted for effort suggests that a coach who relies on it may produce better plans without becoming a better coach, which is a poor bargain for a coach education system whose purpose is the coach's own growth. Fourth, errors cost more. A weak paragraph in a work report is an inconvenience, while an unsuitable loading progression for a child in a growth spurt is an injury risk, and the volunteer is the person least equipped to spot it.

4.3 What a democratising offer would require

If the democratisation case is to be realised in practice, the offer to volunteer coaches has to meet several conditions at once. It has to be free or close to it. It has to rest on the development evidence summarised earlier, which a general-purpose chatbot trained on the open internet has no particular reason to favour. And it has to connect the coach to people, whether a club mentor, a regional coach developer or a peer group, so that the informational gains do not come at the expense of the relational support that coaches say they lack. My own view is that the market is unlikely to meet all these conditions unaided, because the commercial incentive runs towards the paying customer, and the paying customer in sport is usually the well-resourced programme. **ATHLEET.AI**

WHAT IS UNTESTED

No study I could find has measured whether AI planning support changes what athletes experience at grassroots level. Until one does, the democratisation case rests on analogy with office work.

CHAPTER FIVE

The stratification case

The same technology rewards those who already hold data, specialist staff and money, and in sport those advantages compound.

5.1 Four assets that compound

The performance programme in the opening scenario holds four assets that the village club lacks, and each of them multiplies the value of AI. The first is proprietary longitudinal data, often years of training, testing and competition records on athletes who went on to succeed or fail; the second is the ability to commission or adapt models to that data. The third is specialist staff who can judge a model's suggestions, reject the wrong ones and act on the right ones. The fourth is sensor and testing equipment that generates new data of the kind the models need. **ATHLEET.AI**

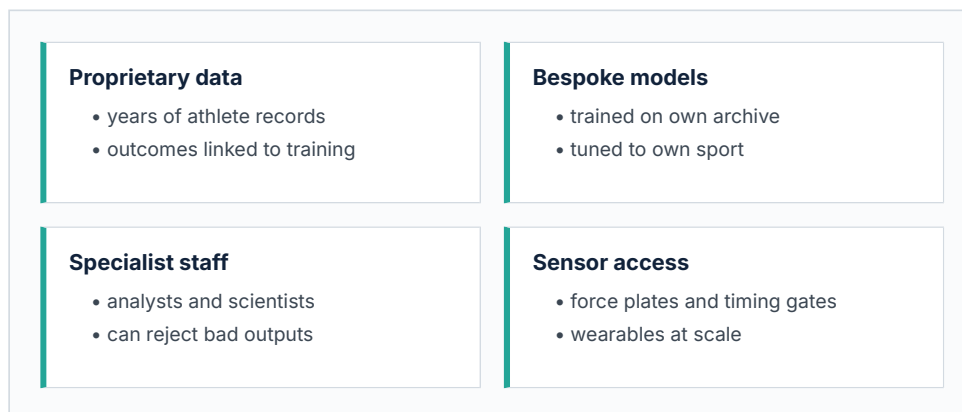


FIGURE 3 Four assets that let well-resourced programmes gain more from AI. Source: author's framework.

The customer support study illustrates the first two assets in a different industry, since the assistant there was, as the authors describe it, trained on records of past support conversations and their outcomes.³ **PREPRINT** A national programme can do something similar with its own archive. A volunteer can only use a general tool trained on whatever the internet happens to contain about coaching, much of it written for adults, and some of it written by people selling something.

The third asset works differently from the others. Specialist staff are what turn a model's output into a safe decision, and they are also what protects a programme from the model's errors. Friedman and Nissenbaum distinguished bias that predates a system, bias that arises from its technical design, and bias that emerges when a system is used in a context, or by a population, its designers did not anticipate.¹⁴ **PEER-REVIEWED** A general chatbot answering a grassroots coach's question about adolescent athletes is a plain instance of that emergent case, and the grassroots coach is the user least able to detect it. The elite programme has a biomechanist to catch the unsuitable drill; the volunteer has only her own judgement.

5.2 Why the gap can widen when everyone gains

The compression evidence in the previous chapter measured differences between workers using the same tool within the same task. The stratification question concerns differences between institutions using different tools with different complementary assets, and there is no reason to assume the same pattern holds. My inference is that where the gain from AI depends on data, staff and equipment, the programmes that hold those assets will gain more in absolute terms, even if the grassroots coach gains more relative to her starting point. **ATHLEET.AI**

For grassroots participation and long-term development, an absolute improvement for the volunteer may be the outcome that matters most, and a governing body could reasonably count it a success. In competition the arithmetic is different, because a podium is a ranking and every advantage is relative. A sport could find its community coaching improving while its competitive structure becomes more stratified, with the programmes that already dominate moving further ahead. Both outcomes could hold in the same year. That possibility is part of why the question is so hard to see from inside any single organisation, where only one of the two trends is usually visible. (Whether the same logic applies between nations, where system strength already differs hugely, is a question this paper leaves open, though I suspect it does.)

5.3 Six forms of fairness

Fairness is asked to do too much work in discussions of AI in sport, and it helps to separate at least six questions that a governing body could answer differently. The table below sets them out, drawing on Loland for competitive fairness,¹⁰ **SCHOLARLY** on Till and Baker for developmental fairness,¹² **REVIEW** and on Friedman and Nissenbaum for the representational and epistemic questions.¹⁴ **PEER-REVIEWED**

Form of fairness	The question it asks	How AI could affect it
Representational	Whose athletes are in the data?	Tools trained on adult or elite data may misread girls, late developers or disabled athletes
Allocative	Who gets the tools and the support to use them?	Access may follow money, club size and existing staff
Competitive	Do athletes have a fair opportunity to perform?	Proprietary systems may add to inequalities of system strength
Developmental	Does the tool serve long-term growth?	Plans may favour early selection, specialisation or adult loads
Procedural	Can AI-informed decisions be explained and challenged?	Selection or funding choices may rest on reasoning nobody can inspect
Epistemic	Whose coaching knowledge is encoded?	Tools may reproduce the views of a narrow group of coaches and writers

TABLE 1 Six forms of fairness at stake when AI enters coaching. Source: author's framework, drawing on Loland (2002), Friedman and Nissenbaum (1996) and Till and Baker (2020).

The epistemic question is the least discussed and, in my view, the most consequential for a national grassroots tool. **ATHLEET.AI** The coaching plan's figures, in which more than half of coaches came from higher socio-economic groups and fewer than one in five qualified coaches was female, suggest that the body of formal coaching knowledge a tool might be built on reflects a particular slice of the coaching population.¹

CONSENSUS A tool distributed to every licensed coach would carry that slice's assumptions into every club, and a governing body that commissions such a tool would, whether it intended to or not, be choosing a coaching philosophy for the whole sport.



PART THREE

Governing what you do not own

The capability governing bodies need, and the choices open to them now.

CHAPTER SIX

Institutional capability and lag

Governing bodies are being asked to judge tools that can change without notice, under general policy frameworks that say little or nothing about sport.

6.1 Tools that change underneath you

Most AI coaching products available to grassroots coaches are, or are likely to be, built on a small number of general-purpose models provided as services by large technology companies. The stability of those services cannot be taken for granted. Chen, Zaharia and Zou compared the March 2023 and June 2023 versions of GPT-3.5 and GPT-4 on four tasks (solving maths problems, answering sensitive questions, generating code and visual reasoning) and found that the behaviour of the same service could change substantially within a short period.⁶ **PREPRINT** GPT-4's accuracy in identifying prime numbers fell from 97.6% to 2.4% between the two versions, while GPT-3.5 improved on the same task; GPT-4 was less willing to answer sensitive questions in June; and both models made more formatting mistakes in generated code. The authors concluded that the findings pointed to a need for continuous monitoring of model quality. **PREPRINT**

The study is a single preprint, posted only days before this paper was completed, and its tasks have nothing to do with sport, so it says nothing about the quality of any particular coaching tool. As a warning about a general class of risk, it still carries weight. A governing body that reviews a supplier's tool in September and endorses it has endorsed the behaviour of the underlying model in September. If the model changes in November, the endorsement travels with the brand while the behaviour it vouched for may not, and the coaches relying on the endorsement will have no way to tell. My view is that approval of an AI tool should be treated as a continuing relationship with monitoring attached, and that a one-off certificate offers false comfort. **ATHLEET.AI**

6.2 General policy with little to say about sport

The wider policy environment offers governing bodies limited help. On 21 July 2023 the White House announced voluntary commitments from seven companies (Amazon, Anthropic, Google, Inflection, Meta, Microsoft and OpenAI), framed around safety, security and trust and including internal and external security testing of AI systems before release.⁷ **CONSENSUS** In Europe, the European Parliament adopted its amendments to the proposed AI Act on 14 June 2023, and the text now moves into negotiation with the Council and the Commission.⁸ **CONSENSUS** In the UK, the government's white paper of 29 March 2023 proposed a principles-based and context-specific approach, applied by existing regulators within their own remits, without a new AI regulator or new legislation at the outset.⁹ **CONSENSUS**

None of these documents, on my reading, addresses sports coaching or competition.

EVIDENCE GAP The UK proposal is the most relevant in principle, because it relies on sector bodies to interpret general principles for their own context. My own view, which goes beyond what the white paper says, is that sport sits awkwardly in that design, since the bodies that set the rules of most sports are governing bodies and not statutory regulators, and a context-led approach may therefore leave coaching tools with no one expressly charged with interpreting the principles for them. **ATHLEET.AI** Whether governing bodies wish to take on that role, and whether they would be recognised in it, is a matter for their own boards and advisers.

6.3 The capability gap

Even a governing body that wanted to assess AI tools would, in most cases, struggle to do so. From my own advisory and coaching work, few national governing bodies outside the best-funded Olympic and professional sports appear to employ people with the technical background to evaluate a machine learning system, test it against the sport's development principles or monitor it after approval. **PRACTICE** Suppliers, by contrast, arrive with specialist staff, polished evidence of their own choosing and a commercial interest in a quick decision. That asymmetry is itself a form of stratification, operating between institutions before a single athlete is affected.

Dyer's finding about the cost of delay applies with particular force here.⁴ **REVIEW** A governing body that lacks the capability to judge AI tools will tend to defer judgement, and deferral in this case is a decision to let the market set the terms. Pooling capability across sports, through a national agency, a shared panel or a partnership with universities, would spread the cost. Universities are an obvious partner, though their incentives run towards publication and their timescales rarely match a governing body's, and I have not resolved how that tension should be handled.

CHAPTER SEVEN

Choices for governing bodies

Five decisions, taken deliberately in the next few years, would tilt AI in coaching towards democratisation.

The options below are my own proposals. **ATHLEET.AI** They are addressed to national governing bodies, since those bodies control coach education and competition rules, and they are ordered so that the earlier options can proceed without waiting for the later ones.

1	Teach AI use in coach education	Cover what tools can and cannot do, at every licence level.
2	Offer a shared, governed tool	Give licensed coaches free access built on development evidence.
3	Set supplier assurance standards	Require evidence, monitoring and change notice from approved suppliers.
4	Require disclosure in competition	Ask programmes to declare AI systems that inform competition decisions.
5	Build or pool capability	Employ or share the technical staff to test and monitor tools.

FIGURE 4 Five choices that would tilt AI in coaching towards democratisation. Source: author's framework.

7.1 Five options

The first option is to treat AI within coach education. Every licence level should include material on what AI tools can and cannot do, on the development principles a plan must respect, and on when a coach should seek a person instead of a prompt. **ATHLEET.AI** This is inexpensive and helps the volunteer who is already using a free chatbot.

The second is to provide a shared and governed tool for licensed coaches, free at the point of use. The governing body should decide what development evidence it rests on, which coaching philosophy it encodes and how its outputs link to human mentors and coach developers. **ATHLEET.AI** This is the option most likely to realise the democratisation case, and also the one that most directly raises the epistemic question from the previous part, so the process for deciding the tool's content should itself be open and should include volunteer coaches.

The third is to set assurance standards for suppliers the governing body approves or recommends. Those standards should require evidence that a tool respects the sport's development principles, disclosure of the underlying model, notice of material

changes and a commitment to continued monitoring after approval. **ATHLEET.AI** The instability documented by Chen and colleagues is the reason for the last two requirements.⁶ **PREPRINT**

The fourth is to require disclosure of AI systems that inform decisions in competition, such as in-event tactical analysis. Disclosure is the lightest available step, and it gives the governing body the information it would need to decide, on evidence, whether any system should be treated as equipment. **ATHLEET.AI** The fifth is to build or pool the capability to do all of this. In most sports that will mean sharing specialists across governing bodies.

7.2 When a model becomes equipment

The hardest question in this paper is the point at which a proprietary AI system should fall within competition rules. The World Athletics shoe rules suggest one test, which is availability, so that a system in use during competition would need to be available to all competitors on reasonable terms or be treated as a prototype.⁵

CONSENSUS A second test would look at proximity to the contest, distinguishing systems that shape training months in advance from systems that feed decisions during an event. My provisional view is that governing bodies should begin with disclosure for systems used during competition, apply an availability test only where disclosure shows a material and exclusive advantage, and leave training-stage AI to coach education and supplier standards for now. **ATHLEET.AI** I hold that view loosely, because the boundary between preparation and competition is porous and the tools are changing faster than any rule could.

7.3 Conclusion and open questions

AI will make some grassroots coaching better, and it will make some elite programmes stronger, and neither effect waits for a governing body's permission. The early workplace evidence supports the hope that less experienced coaches would gain most, while sport's own history supports the worry that advantage follows resources and that governing bodies act late. My conclusion is that the balance between these outcomes remains open and that governing bodies are the institutions best placed to shift it, provided they treat AI access as part of coach development, prepare their competition rules for proprietary systems and build the capability to govern tools they do not own. Several questions remain unanswered, and I would put the following to any board considering the issue.

- Should a governing body provide a shared, governed AI tool for every licensed coach, and who should answer for its advice?
- At what point does a proprietary AI system become an equipment question under competition rules?
- Whose coaching philosophy should a national grassroots tool encode, and who should decide?
- How would a sport know, five years from now, whether AI had narrowed or widened the gap between its best-resourced and least-resourced athletes?

The last question may be the most practical, since no governing body I know of currently collects data that could answer it.

END MATTER

References

1. Sport England (2016) *Coaching in an Active Nation: The Coaching Plan for England 2017-21*. London: Sport England.
2. Noy, S. and Zhang, W. (2023) Experimental evidence on the productivity effects of generative artificial intelligence. *Science* 381(6654), 187-192.
3. Brynjolfsson, E., Li, D. and Raymond, L.R. (2023) *Generative AI at Work*. NBER Working Paper 31161. Cambridge, MA: National Bureau of Economic Research.
4. Dyer, B. (2015) The controversy of sports technology: a systematic review. *SpringerPlus* 4, 524.
5. World Athletics (2020) Modified rules on shoes. Press release, 31 January 2020. Monaco: World Athletics.
6. Chen, L., Zaharia, M. and Zou, J. (2023) How is ChatGPT's behavior changing over time? arXiv:2307.09009, 18 July 2023.
7. The White House (2023) Fact sheet: Biden-Harris Administration secures voluntary commitments from leading artificial intelligence companies to manage the risks posed by AI. 21 July 2023. Washington, DC: The White House.
8. European Parliament (2023) Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Strasbourg: European Parliament.
9. Department for Science, Innovation and Technology (2023) *A Pro-Innovation Approach to AI Regulation*. White paper, 29 March 2023. London: DSIT.
10. Loland, S. (2002) *Fair Play in Sport: A Moral Norm System*. London: Routledge.
11. Rees, T., Hardy, L., Güllich, A., Abernethy, B., Côté, J., Woodman, T., Montgomery, H., Laing, S. and Warr, C. (2016) The Great British medalists project: a review of current knowledge on the development of the world's best sporting talent. *Sports Medicine* 46(8), 1041-1058.
12. Till, K. and Baker, J. (2020) Challenges and [possible] solutions to optimizing talent identification and development in sport. *Frontiers in Psychology* 11, 664.
13. Noy, S. and Zhang, W. (2023) Experimental evidence on the productivity effects of generative artificial intelligence. Working paper (not peer reviewed), Massachusetts Institute of Technology, 2 March 2023.
14. Friedman, B. and Nissenbaum, H. (1996) Bias in computer systems. *ACM Transactions on Information Systems* 14(3), 330-347.

END MATTER

About the author

Paul Forrest is the founder of Athleet.AI. He advises boards and leadership teams on the adoption and governance of artificial intelligence, drawing on a background in business process reengineering and on hands-on work building machine learning systems, including small, domain-specific language models. He has served as a special adviser to UK government on the Global Entrepreneur Programme and works as a board member and business adviser.

He is also a qualified athletics coach who coaches sprint, hurdle and endurance athletes as a volunteer, and much of his thinking about how people learn, decide and improve comes from the track.

This paper is part of the Athleet.AI Working Papers, AI in Coaching and Athlete Development. The views expressed are the author's own.

athleet.ai

Democratised or Stratified?

Athleet.AI Working Papers, AI in Coaching and Athlete Development
Paper six of six · 26 July 2023