Paradoxical Tensions Related to AI-Powered Evaluation Systems in Competitive Sports

Elena Mazurova[1] · Willem Standaert[2,3] · Esko Penttinen[1] · Felix Ter Chian Tan[4]

Abstract

Judging in competitive sports is prone to errors arising from the inherent limitations to humans' cognitive and sensorial capabilities and from various potential sources of bias that influence judges. Artistic gymnastics offers a case in point: given the complexity of scoring and the ever-increasing speed of athletes' performance, systems powered by artificial intelligence (AI) seem to promise benefits for the judging process and its outcomes. To characterize today's human judging process for artistic gymnastics and examine contrasts against an AI-powered system currently being introduced in this context, an in-depth case study analyzed interview data from various stakeholder groups (judges, gymnasts, coaches, federations, technology providers, and fans). This exploratory study unearthed several paradoxical tensions accompanying AI-based evaluations in this setting. The paper identifies and illustrates tensions of this nature related to AI-powered systems' accuracy, objectivity, explainability, relationship with artistry, interaction with humans, and consistency.

Keywords Paradoxical tensions · Artificial intelligence · Sports judging · Bias · Explainability · Artistry

1 Introduction

Human judging plays an important role in sports: figure skating, gymnastics, snowboarding, and many others (Stefani, 1998). Well-timed, accurate, and reliable information and evaluations are key factors that can contribute to more reliable judging and, indirectly, better athletic performance (Harding & James, 2010). However, human-based judging is susceptible to error due to a host of factors, from fatigue to various biases of judges (Perederij, 2013). Research attests that competitive-sports judging is influenced by judges' prior knowledge and values, earlier experience, training, iterative reflection, and cognitive and sensory limitations (Plessner & Haar, 2006). This presents a problem: "In sport, the accuracy of the results of a game or competition is important in order for the sport to be deemed valid, but in many cases in sport, humans cannot always provide reliable results" (Kerr, 2018, p. 116). Moreover, the time consumed by human judging can render a competition more tiring for athletes and less spectator-friendly for live and televisual audiences. Some shortcomings of human judging can be especially costly, leading to judging scandals, retarding athletes' development, and decreasing the sport's overall attractiveness.[1] In summary, they put the legitimacy of a human-judged sport at risk.

To overcome issues of subjectivity and bias in judging, improve the decisions' objectivity and accuracy, and expedite the judging process, AI-powered systems have been introduced in recent years as an aid to refereeing, performance assessment, and judging for sports.[2] Such changes

✉ Willem Standaert
willem.standaert@uliege.be

Elena Mazurova
elena.mazurova@aalto.fi

Esko Penttinen
esko.penttinen@aalto.fi

Felix Ter Chian Tan
f.tan@unsw.edu.au

1   Aalto University, Espoo, Finland

2   HEC Liège – Management School of the University of Liège,
    Rue Louvrex 14, B-4000 Liège, Belgium

3   Ghent University, Ghent, Belgium

4   School of Information Systems and Technology Management
    at UNSW Business School, Sydney, Australia

---

[1]   An example is provided by the Reuters piece "Judging scandal still haunts Korean Yang," available at https://www.reuters.com/article/us-olympics-rio-gymnastics-yang/judging-scandal-still-haunts-korean-yang-idUSKCN10205W.

[2]   The introduction of AI in soccer is particularly high-profile; see https://www.espn.com.au/football/fifa-world-cup/story/4237186/fifas-semi-automated-offside-var-tech-to-go-into-development-in-2021.

in officiating or judging are often met with objections and controversy (Kolbinger & Lames, 2017). As a revelatory case in point, we studied the relatively complex context of AI in artistic gymnastics, wherein Japanese technology company Fujitsu and the International Federation of Gymnastics (or Fédération Internationale de Gymnastique, FIG) are collaborating to develop an AI-powered judging system.[3] Such systems hold potential to circumvent human sensory and cognitive limitations and to offset or eliminate human biases by using a combination of AI technologies to support (or even replace) the actions of human judges (Benbya et al., 2021). A core element of Fujitsu's computer-vision-based judging-support system (JSS) is a three-dimensional computer-generated image of proceedings that is examined in light of set definitions of gymnastics elements, for determination of a performance score.

We conducted a two-year exploratory case study examining the introduction of the Fujitsu system and how the various stakeholder groups perceive it. With AI-powered performance judgement remaining in its infancy, the possible positive and negative consequences are not obvious yet. Hence, we examined the complex form of human judging involved and how an AI-powered system might ameliorate or compound the issues posed by such human-based systems. Pursuing clarity that could inform the development of AI-based performance-judging technology, we employed the analytical lens of the *paradox* to identify, understand, and explain the *tensions* experienced by particular stakeholders (Dubé & Robey, 2009; Smith & Lewis, 2011) in the introduction of an AI-powered system for sports. We formulated the research question accordingly: "*What are the paradoxical tensions related to the use of an AI-powered system for competitive-sports judging?*"

To address this question, we conducted an in-depth case study of the application of the aforementioned AI-powered system for judging in artistic gymnastics. Taking an approach similar to one employed in prior research (Calabretta et al., 2017), we used several rounds of coding to develop a data structure (Gioia et al., 2012) that covers six "paradoxical tensions": 1) accurate AI is *too* exact, 2) ostensibly objective AI can be biased, 3) black-boxed AI provides a *sense* of explainability, 4) AI-based judging for artistic gymnastics cannot judge artistry, 5) much AI intended for humans lacks human interaction, and 6) consistency requires AI's adaptability. We believe that acknowledging these tensions should provide a step toward better understanding and,

thereby, fundamentally better design of AI-powered systems' introduction for supporting or replacing humans' evaluations (Benbya et al., 2021).

We begin with a literature review discussing the use of electronic support for judging in sports, then turn to the use of AI for (expert) evaluations. The section concludes with a review of the notion of paradoxical tensions and how prior research has addressed them with regard to technologies and AI in particular. Then, in Section 3, we outline our methodology, describing the collection and analysis of data. Further vital background is supplied in the fourth section, dealing with the case setting of artistic gymnastics and details of the *as-is* situation of human judging alongside what the use of an AI-powered system is expected to be like. Then, we proceed to present the paradoxical tensions identified through our empirical study. The body of the paper concludes with a discussion of our findings' implications.

2 Literature Review

A review of the scholarly work on the key concepts utilized in our research aids in positioning this study. The key branches of literature have examined electronic judging systems in sports, the use of artificial intelligence for expert evaluation in particular, and the notion of paradoxical tensions.

2.1 Electronic Systems in Judging of Sports

Humans' sports-performance judging is influenced by individual judges' prior knowledge and values, which are, in turn, based on experience, training, and iterative reflection (Schön, 1983). Therefore, this judging is liable to suffer from human error and to manifest various types of biases (Plessner & Haar, 2006). Designed to address these concerns, electronic systems for assisting with judging have entered widespread use in the realm of competitive sports, since the mid-1990s. Often, the implementation has been triggered by a push for professionalization in the sport, with further impetus and fuel sometimes added by the relevant sport's inclusion in the official program of the Olympic Games (Taymazov et al., 2013).

Humans' processing of information is inherently slow, and this bottleneck prolongs gymnastics competitions. The protracted decision-making also makes the sport less inviting for spectators. Moreover, the same judging panel must evaluate all participants in any given competition, for the sake of consistency, though this presents the risk of panel members being tired and less able to concentrate by the end of a 12-h session, for instance (Perederij, 2013). Anecdotal evidence suggests that another time-related factor is at play too: gymnasts who perform early in the day are evaluated

more harshly and hence have lower chances of advancing in the competition (Mazurova et al., 2021). Moreover, various studies have indicated that gymnastics judges make errors connected with the people involved or other aspects of context (Flessas et al., 2014; Kerr, 2018; Plessner & Schallies, 2005). The former may entail scoring that is biased in relation to the nationality, body morphology, or reputation of the gymnast (Duong, 2008; Heiniger & Mercier, 2018). Further contextual factors may be more mundane – e.g., the judge's vantage point or viewing angle (Plessner & Schallies, 2005) – but no less relevant. Such factors have given rise to several controversial moments in sports history related to judging decisions (Dumoulin, 2020).

Electronic systems can assist in overcoming errors of human judgement (related to accuracy, fairness, validity, and reliability) and, thereby, preventing inquiries and complaints by coaches and athletes (Can et al., 2011; Omorczyk et al., 2015). For instance, judges may exploit technologies such as video replay (slow-motion and time-lapse in particular) and time measurements (Omorczyk et al., 2015). Also, if the athletes are fitted with wearables, sensor data from their movements can inform the judging (Harding & James, 2010). The use of such support technologies may be expected to improve judging systems' reliability significantly and to reduce both conformity bias and arithmetic errors in the scoring of athletes' performance.

The use of electronic systems of various sorts to decrease human factors' influence on the judging process may improve the quality of competitions from the audience perspective too, making them more understandable and, through more real-time feedback, exciting for spectators (Ferger & Hackbarth, 2017). Looking specifically at the use of technology for taekwondo judging at the 2012 Olympics, Leveaux (2012) found evidence supporting this conclusion. Having struggled for some time to provide transparency in the judges' decision-making and render it attractive to spectators, the sport's key actors chose to address these issues by embracing technological advances. Leveaux concluded that the technologies not only improved the correctness of the decisions greatly but, by doing so, also contributed to a more attractive competition from the fans' perspective (Leveaux, 2012).

Many suggest that the use of electronic judging systems in sports, gymnastics in particular, possesses potential to make a positive impact on the technical development of performance evaluation, through greater objectivity and clarity of the judging process and outcomes (Can et al., 2011; Ferger & Hackbarth, 2017; Omorczyk et al., 2015; Taymazov et al., 2013). However, researchers have expressed concerns at the same time, in that no technology can interpret and assess the myriad of situations that competitions present, especially with regard to athletes' artistry and creativity (Leveaux, 2010). Also, various stakeholders strongly oppose technology in the judging process on such grounds as freedom, individuality, and aesthetic focus (Harding et al., 2008). People are attached to these values and do not want to see them eroded through technology. Finally, there are concerns about job losses: human judges might get replaced altogether at some point (Mazurova et al., 2021).

2.2 AI for (Expert) Evaluations

Artificial intelligence (AI) has been defined as technology that offers "the ability of a machine to perform cognitive functions that we associate with human minds, such as perceiving, reasoning, learning, interacting with the environment, problem-solving, decision-making, and even demonstrating creativity" (Rai et al., 2019, p. iii). The key strengths of AI lie in pattern recognition, probability work, consistency, speed, and efficiency (Dellermann et al., 2019). For numerous cognitive and perceptual tasks, it already outperforms humans or is expected to do so in the near future (Benbya et al., 2019).

A key application of AI for business lies in generating insight and making decision in (narrow, thus far) task domains (Benbya et al., 2021; Raisch & Krakowski, 2021), since this technology "enables the creation of new information and predictions from data" (Shrestha et al., 2019, p. 67). Several research themes related to human versus AI decision-making have been identified, among them accountability, transparency, biases, ethics and associated values, efficiency, replacing intuition with rationality, and adaptability (Benbya et al., 2021). Of particular interest for the present paper is the specific field of decision-making related to expert evaluations and the assessment of human performances. Much of the recent work on the use of AI technology in this arena has been conducted in a human-resources/organizational context.

For instance, in the framework they synthesized for AI in the context of organizational control. Kellogg et al. (2020) found two mechanisms for algorithmic evaluation: recording and feedback (based on fine-grained behavior) and rating and ranking (based on aggregated quantitative and qualitative data). Also examining the literature to pinpoint new affordances of AI technology in this domain, they identified comprehensiveness (use of a variety of data), instantaneity (high velocity), interactivity (interfaces for participation), and opacity (abstraction connected with technical literacy). Finally, they observed resistance against algorithmic control among workers, who expressed concerns related to data accuracy, surveillance and loss of privacy, algorithms' discrimination, and absence of non-algorithmic assessment. Among the tactics of resistance emerging among workers, for which the authors coined the term "algoactivism," were ignoring algorithm recommendations, obfuscating data, and hacking the algorithm (Kellogg et al., 2020).

In another human-resources context, van den Broek and colleagues (van den Broek et al., 2021) examined the introduction of a machine-learning-based system aimed at better hiring decisions. The authors found that AI systems "promise to mitigate the human biases, inefficiencies, and path-dependencies that have plagued experts' work for decades, discovering 'truthful' knowledge on their own instead" (p. 3). While there is a common belief, accordingly, that AI can yield superior insights efficiently and consistently, the authors concluded that it remains crucial to keep domain experts "in the loop," via a human–AI hybrid system. In particular, they noted that the various developers and domain experts, as they engage in the AI development process's distinct phases, reflect on several factors: when constructing training data, they reflect on AI's objectivity and experts' bias; when building a model, they consider the novelty of AI and the criteria for the choice of experts; and when using a model, they reflect on AI's efficiency and practical utility (van den Broek et al., 2021). Such reflections prompt cycles of mutual learning, to exclude human experts' knowledge for purposes of independence and to include it for its relevance. The end result is a human–AI configuration that, while not purposefully designed as a hybrid, emerges through this dialectical process.

2.3 Uncovering Paradoxical Tensions Related to Technology Use

Lack of conceptual clarity and consistency characterizes research on paradoxes across contexts (Smith & Lewis, 2011). In line with recent paradox research in the management, organization studies, and information systems (IS) fields (Calabretta et al., 2017; Raisch & Krakowski, 2021; Wimelius et al., 2021), we draw on the work of Smith and Lewis (2011) on paradoxical tensions, which they define as consisting of "contradictory yet interrelated elements (dualities) that exist simultaneously and persist over time; such elements seem logical when considered in isolation, but irrational, inconsistent, and absurd when juxtaposed" (2011, p. 387). The two essential elements of a paradoxical tension are contradiction and interdependence – it is their combination that extends beyond a mere tradeoff perspective, to a systemic relationship (across time and space) (Schad et al., 2016). In organizational and other contexts, paradoxes are fueled by technological advancements, growth in complexity, and diversity of perspectives (Lewis, 2000; Raisch & Krakowski, 2021).

Foundational management-science literature highlights identifying and exploring paradoxes as potentially fertile ground for theory development or a starting point to refinement (Lewis, 2000; Poole & van de Ven, 1989; Smith & Lewis, 2011), pointing out that contradictions aid in challenging assumptions and expectations of consistency,

stimulate new insight, force creative thinking to emerge, and reveal underlying mechanisms (Majchrzak et al., 2013; Robey & Boudreau, 1999). In practice, paradoxes are considered crucial for explaining dramatic change; hence, rather than stamp them out, managers are advised to "embrace paradoxical thinking as a stimulus for more complex and creative action" (Robey & Boudreau, 1999).

IS scholars have identified multiple paradoxical tensions when looking at the use of emerging technologies or examining technology in novel contexts. For example, Jarvenpaa and Lang (2005) explored paradoxes related to the use of mobile technology and identified tensions in terms of freedom, dependence, needs, and planning. Another example is research by Dubé and Robey (2009) into virtual team work, wherein tensions were pinpointed with regard to the type of task, trust, and structure. Beyond any specific technology or context, Wimelius et al. (2021) identified paradoxical tensions that organizations may encounter when updating their technology (e.g., adopting new platforms and infrastructure solutions) – namely, between "established and renewed technology usage, deliberate and emergent renewal practices, and inner and outer renewal contexts" (p. 220).

A few scholars have explored specific paradoxes and tensions connected with the use of AI (Benbya et al., 2021). One of these is Moravec's paradox, which articulates the fact that AI, while able to handle various cognition-related tasks that humans deem to require adult intelligence, cannot perform some of the tasks (e.g., more social and emotional ones) that a one-year-old human baby can complete (Dellermann et al., 2019). Another relevant paradox, referred to as Polanyi's paradox, involves inability to transfer knowledge from humans to machines and vice versa (Brynjolfsson & Mcafee, 2017). Initially, the core problem tackled in computing was how to codify tacit knowledge of humans, but focus has shifted to the problem of not knowing how the AI arrives at any given decision or outcome (Brynjolfsson & Mcafee, 2017). In other work, Raisch and Krakowski (2021) examined the dual-application paradoxical tension related to AI (automation–augmentation): a line cannot be drawn between AI taking over humans' tasks and AI collaborating with humans to perform a task. While many handbooks for practitioners recommend prioritizing augmentation over automation, these two authors argue that focusing on just one of the two in the management domain produces negative-reinforcement cycles. They argue that instead taking a perspective that comprises both applications leads to virtuous cycles of benefits for organizations and society as a whole (Raisch & Krakowski, 2021). Similarly, van den Broek et al. (2021) have argued that human–AI hybrids offer great value in the context of knowledge systems, for overcoming the tension related to producing knowledge that is both independent and relevant.

In prior work, technology-related paradoxical tensions have been recognized for their impact not just at the level of functioning or productivity but also at that of user emotions, from joy, empowerment, and belonging for positive experiences to anger, fear, and depression for negative ones (Jarvenpaa & Lang, 2005). Once a technology paradox has been identified, it can be tackled. An appropriate response to the tension, while seldom obvious or easy, is considered critical for the success of technology-related initiatives (Wimelius et al., 2021). In broad terms, there are two approaches to coping with paradoxical tensions: *avoiding* them, by minimizing or abandoning the technology's use, and *confronting* them, by bringing understanding of the technology's features to bear for changing how it is used (Dubé & Robey, 2009; Jarvenpaa & Lang, 2005; Lewis, 2000).

The discussion above suggests that paradoxical tensions are likely to emerge with AI-powered judging systems' implementation in gymnastics. In the absence of studies specific to tensions connected with such systems for competitive sports, we undertook an exploratory study in this particular context. To explore and identify foreseeable contradictions and interdependencies, then formulate them as clear paradoxical tensions, our empirical research applied the qualitative approach described below (per Walsham, 2006).

3 Methodology

Given the exploratory nature of our research, into a still poorly-understood phenomenon – paradoxical tensions related to AI-powered systems in the judging process – a case-study method is especially appropriate (Siggelkow, 2007;Walsham, 2006). In addition, we adopted a qualitative orientation with a predominantly interpretive stance (Klein & Myers, 1999). For our case study, we ascertained that a relevant and revelatory context for tackling our research question (Yin, 2013) would be one with a transition in progress from human-based judging to an AI-powered system. International competitive artistic gymnastics offered a suitable setting of this nature, thanks to its current process of transition. Further, we employed purposeful sampling to support a relevant, information-rich empirical setting and select suitable informants (Patton, 2002).

3.1 Data Collection

We applied the following criteria in selection of the case and data sources for our study: 1) informants should be stakeholders affected by the transition to an AI-powered system (i.e., judges, gymnasts, coaches, technical directors and representatives of international and national artistic-gymnastics federations, and fans), and 2) they should be involved with international competitions at senior level (i.e.,

the European Championships, World Championships, and Olympic Games), the first competitions envisioned for the AI-powered system's deployment. Since we aimed to take diverse viewpoints into account, we also interviewed Fujitsu representatives whose work dealt with developing and promoting the system in question. This input rounded out the picture with the perspective of commercial players.

We conducted semi-structured interviews with 52 informants, from May 2019 to March 2021. The breakdown of the stakeholder group is as follows: 20 international judges, 11 gymnasts competing at the international level, eight coaches of international teams, two technical directors with national federations, two FIG representatives, two representatives of Fujitsu, and seven artistic-gymnastics fans. In the interests of our informants' anonymity, each has been assigned a pseudonym, as shown in Table 1.

We paid special attention also to the initial interview protocol, to guarantee its thoroughness, focus on the research question, and avoidance of leading questions. The protocol entailed open-ended questions, for collection of a broad base of empirical results without researcher influence on the responses. When developing the interview questions, we were guided primarily by a wish to encourage the participants to share their opinions and perceptions of both the human-based judging system and the new, AI-powered one. We avoided the words "paradox" and "tensions," lest these influence the interviewee (Gioia et al., 2012; Jarvenpaa & Lang, 2005). We adjusted the interview questions slightly as the research progressed, guided where the informants led us in our investigation of the overall research question. The interview topic guide was prepared with general open questions for all informants, alongside questions tailored for each stakeholder group (see Appendix 1). The questions' design focused on eliciting open sharing of opinions on topics such as the accuracy and biases of both systems, explainability, the training process, and related challenges. All interviews were audio-recorded and transcribed (after which non-English-language interviews were translated into English). All told, the resulting corpus of empirical data consists of 113,449 words.

3.2 Data Analysis

The research team analyzed all of the interview material, with support from software designed for qualitative data analysis (ATLAS.ti). To display the evidence for our assertions, demonstrate a systematic way of gathering and analyzing the data, and make the data's presentation more structured, we used the Gioia method to produce a representation of the first- and second-order data analyses (see Appendix 2). Organizing the data via two orders of categories allows one not only to "facilitate their later assembly into a more structured form" but also to "enhance qualitative rigor" of

Table 1 Informants in the case study

	No.	Pseudonym		No.	Pseudonym		No.	Pseudonym	Role
Judges	1	Abby	**Gymnasts**	1	James	**Others**	1	Steven	Tech. director
	2	Bella		2	John		2	Mary	Tech. director
	3	Charlie		3	David		3	Simon	FIG
	4	Edward		4	Thomas		4	Adam	FIG
	5	Felicity		5	Mark		5	Joona	Fujitsu
	6	Harry		6	Lauren		6	Caleb	Fujitsu
	7	Lilly		7	Reece		7	Henriette	Fan
	8	Nick		8	Fabian		8	Megan	Fan
	9	Norman		9	Nathan		9	Fabian	Fan
	10	Sarah		10	Jacob		10	Thomas	Fan
	11	Ulla		11	Damian		11	Jo Allen	Fan
	12	Tracy					12	Isla	Fan
	13	Isabella	**Coaches**	1	Michael		13	Bruno	Fan
	14	Mia		2	William				
	15	Sofia		3	Daniel				
	16	Emily		4	Kyle				
	17	Bob		5	Margaret				
	18	Don		6	Jessica				
	19	Katarina		7	Paul				
	20	Josh		8	Kevin				

the research; here, the first-order analysis represents "analysis using informant-centric terms and codes" and the second-order one employs "researcher-centric concepts, themes, and dimensions" (Gioia et al., 2012, p. 18). This technique allowed us to demonstrate the links between the data and indicate the main paradoxical tensions related to the introduction of an AI-powered judging system.

With the first-order concepts, we highlighted codes that reflect informants' opinions. The process, which resembled an open coding technique, resulted in 96 distinct codes reflecting interviewees' ideas and opinions (Appendix 3 presents a summary of the codes with relevance for our study and their related content). This coding process utilized cross-reading and comparison of the interview transcripts. The questions' topic-based organization in our interview-topic guide facilitated summarizing the participants' opinions, perceptions, and predictions for both of the judging systems and carrying out their comparative assessment, generally and by informant role. In the interest of validity and reliability, researchers took turns checking the results of the data analysis (Patton, 2002).

We chose the first-order constructs conscientiously to use the informants' terms, not ours; this helped the concepts reflect their points of view. At this stage in the data analysis, we made little attempt to distill wider categories. As the research progressed, for formulation of the second-order constructs (themes), we started seeking intergroup similarities, correlations, and patterns among the many codes. This process was similar to that of axial coding. We formed

17 code groups or categories, then assigned them phrasal descriptors and attempted to analyze them for their description and explanation of the phenomenon under examination. In the second-order analysis, we sought higher-level perspectives, necessary for informed theorizing aligned with themes. After this, we assessed the possibility of distilling the emergent themes even further, into aggregate dimensions, whereupon we uncovered paradoxical tensions. Together, these elements form the building blocks of the data structure presented in Appendix 2.

In parallel with the data-gathering work and after the initial stages of analysis, we iterated between the emerging dataset, themes, concepts, and dimensions and, on the other hand, the relevant literature. At this phase in the process, "the research process might be viewed as transitioning from 'inductive' to a form of 'abductive' research, in that data and existing theory are now considered in tandem" (Gioia et al., 2012). In this, we faced the inevitable issue of authors varying in their interpretation of some informants' comments and chosen terms. Therefore, as the data analysis neared its completion, we reviewed the source material, engaged in group discussion, and reconciled divergent interpretations, thereby reaching agreement on how to code various terms or phrases. This refinement to our analysis enabled clearly identifying the main challenges and paradoxical tensions related to implementing a new, AI-powered judging system in the gymnastics context.

Building the data structure (see Appendix 2) constituted a key step in our qualitative research. It aided in crystallizing

Fig. 1 The disciplines in men's artistic gymnastics
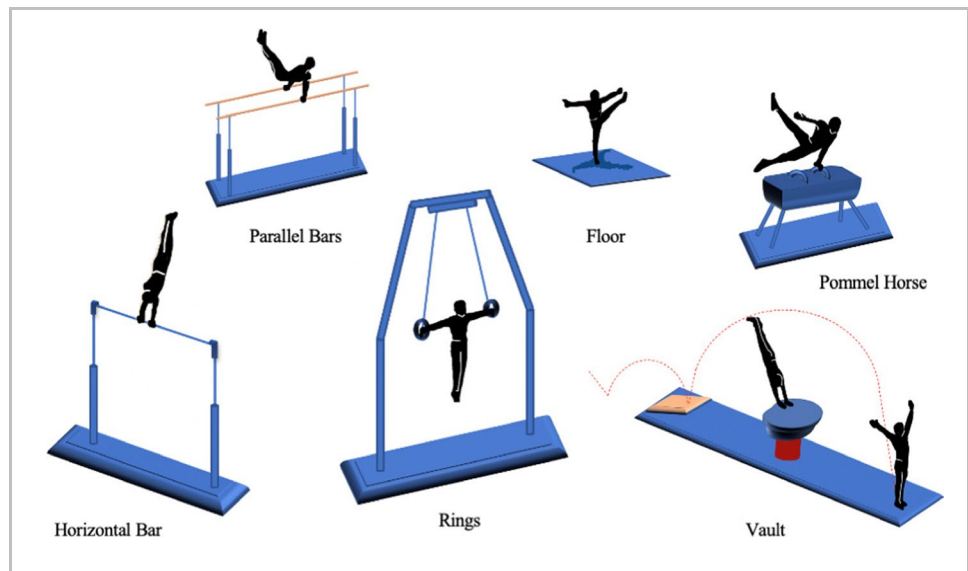
a graphical representation of our progress from raw data to terms and themes as we conducted the analyses and in configuring them for further theorizing (Gioia et al., 2012). We strove to form an inductive model that not only is a data-grounded one capturing the informants' experience but also reflects the dynamic interrelationships between them. Another part of our goal was to represent the essential concepts, themes, and dimensions encompassed by the data structure clearly and render "the relational dynamics among those concepts" (Gioia et al., 2012) more transparent.

4 Background on the Case: Competitive Judging in Artistic Gymnastics

Artistic gymnastics comprises several disciplines, each involving a specific apparatus. Men's artistic gymnastics features six distinct disciplines (see Fig. 1), and women's competitions have four (see Fig. 2). Both international and other events involve competitions specific to each discipline, overall ("all-around") competitions, and team (e.g., country-based) competitions. The events' order of execution for routines usually follows this apparatus-based order: (for men) floor exercises, the pommel horse, the rings, the vault, the parallel bars, and the horizontal bar and (for women) the vault, the uneven parallel bars, the balance beam, and floor exercises (Appendix 4 provides a brief explanation of each discipline). Before starting the routine, the gymnast raises his or her arms, thus saluting the judges and declaring readiness. In response, the judges salute back or nod to acknowledge that they too are ready. When the routine is over, a final raise of the athlete's arms signals this.
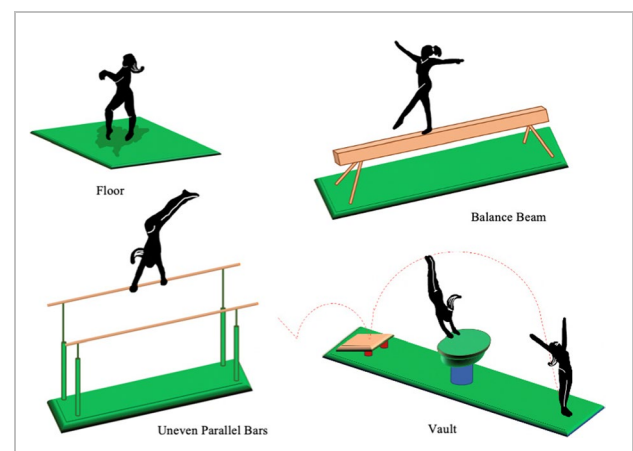

Fig. 2 Women's artistic-gymnastics disciplines

4.1 Human Judging in Artistic Gymnastics

Since the same set of judges evaluates all gymnasts in a given competition, judging panels for artistic gymnastics sometimes are active for stretches of 10–12 h on several consecutive days. They follow an elaborate scoring system created by the FIG, the organization that oversees World Championships competitions and the Olympic Games.

In this system, known as the Code of Points (CoP), the various skills shown in a routine get assigned particular numeric values.[4] A gymnast's final score is composed of a Difficulty (D) score and an Execution (E) score, both of which may suffer from human judging error. The total

---

[4] The CoP rulebook, covering these point values, can be downloaded from http://www.fig-gymnastics.com/publicdir/rules/files/en_WAG%20CoP%202017-2020.pdf.

D-score is the sum for the eight most difficult elements in the routine coupled with an evaluation of the variety expressed by the elements and how they were combined. When a gymnast includes an element never seen before, it must be assigned a difficulty score (usually, this is done before the new element is presented at a competition). At competitions, two judges evaluate the difficulty of the routine, independently from each other, and then must reach consensus. The total *E-score*, in turn, rates the performance for execution and artistry. The base E-score is 10 points, with points being deducted for errors in the technique and artistry exhibited. Small errors bring minor deductions (0.1, for example); for large errors, such as falling, the deductions are larger (such as a full point). Six judges participate in determination of the E-score: the highest and lowest of the six scores given are not taken into account, and the average of the remaining four becomes the final E-score. The D- and E-score, when added together, make up the final score of the gymnast.

Each athlete's routine includes compulsory and optional elements. The former are specific actions that all gymnastics must perform if wishing to compete against each other at a certain level, where the official levels at which a gymnast may compete are Level 1 to (the most advanced) Level 5. The optional elements cover additional strengths, elements, and advantages that the athlete displays and the artistic aspect of the performance. For example, the music choice and choreography for floor exercises put the athlete's personality and charisma on display. A gymnast among our informants reflected thus on balancing the D- and E-score:

> It's always better to do something less creative, a little bit easier, with cleaner execution, than something extremely difficult that you might fail on. But it depends on the situation: if you're trying to maximize your score, you will have to do something a little harder, more creative, more difficult in, say, the finals [...] where you're trying to win, essentially. You might have to perform something more difficult. It's risk management. (David, gymnast)

When a gymnast and coach jointly create a routine for execution, they have to check the CoP guidance for their level of competition and make sure the routine is designed to achieve maximum points for the athlete's skill range and meets all the requirements. In other words, they establish the D-score by preparing a set of elements, and executing the corresponding actions as well as possible maximizes the E-score. If an athlete leaves an element out or improvises in the middle of the routine, the D-score will change.

As the sport itself does, the judging of artistic gymnastics involves exceptional skills and competition. Only the best judges are allowed on panels for international competitions. Judges need to remain well versed in the CoP, and they have regular meetings to learn about updates in the evaluation

system. The FIG tracks and evaluates their performance, in terms of accuracy and consistency (Mercier & Heiniger, 2018), in a process that judges described thus:

> Judges are [...] checked for what they are doing. They can get a yellow or red card. (Nick, judge)
> We have a very strict program; it's called "the judges' evaluation program," and it looks at your scores. It looks at the scores you give your own country and the score you've given to a competing country, the country that is closest in rank to yours. And if you unreasonably give a higher score to some athletes or a lower score to others, you will actually get sanctioned. (Abby, judge)

In artistic gymnastics, athletes and coaches do not receive any official clarification of the results and the deductions made by the judges. Neither do they get explanations of the mistakes made by the gymnast during the routine. Also, they are not allowed to talk to the judges at all during international competitions, let alone ask questions about the routine. Competitions' time constraints are cited as justification for the lack of an explanation accompanying the scores, alongside human cognitive and sensory limitations that rule out the amount of precision expected.

Currently, there are very few opportunities for verifying the judgement process applied by the panel or for protesting against the outcome. The only recourse is an official inquiry, which is a request by the coach or gymnast for revision of the D-score. In response to this appeal, the head/supervisory judge replays video of the routine (in slow motion) and may decide to adjust the D-score – possibly downward. Even here, no clarification or explanation gets supplied.

Perhaps unsurprisingly, the lack of clarification of the results often provokes conflict between coaches and judges.

> People are emotional. Coaches are [...] emotional, even more emotional than athletes... Very often it happens that an athlete works, trains well, and at the competition is also trying hard – of course. And for the coach it may seem like the athlete did everything great! And then suddenly the judges punish and make deductions... and it happens that the coach comes, yelling emotionally, [...] "What are you judging here?! How can you do it?! My gymnast just did a great job!" [...] Very often, later on they calm down and often come back, asking for forgiveness. (Ulla, judge)

In the opinion of several interviewees, resolving the problem of low explainability of humans' judging could never be easy, for reason of its complexity.

> In reality it's not possible, because every judge has a different way to make the record of deductions. The FIG requires the judges to use the [prescribed set of]

symbols to record, but only a few judges can use 100% [of the symbols]. (Harry, judge)

While explanations and guidance from judges during the competition remain scarce, there is ample opportunity for this before and after it. For instance, top-level judges typically counsel top gymnasts from their country, in a somewhat informal manner, as one judge explained:

Sometimes we recommend [to] a gymnast, based on their body type, what kind of element may work better for them. Or we advise them to change the order of the elements, such that they are less tired when executing them. (Bob, judge)

It is important to note that such sharing of expertise and knowledge among judges, coaches, and gymnasts trickles down from the highest level of (professional) gymnastics all the way down to the most local, grassroots level. In addition, "podium training" is held right before an international competition. This provides an opportunity for judges to get a preview of what the gymnasts will do and for gymnasts and coaches to obtain some feedback beforehand, as judge Harry explained:

During the podium training, gymnasts show their routine and they are allowed to ask the judges what would be the amount of deduction on the execution.

4.2 AI-Powered Judging for Artistic Gymnastics

In November 2018, the FIG and Fujitsu unveiled a judging-support system intended for real-time "judging support that is fair and accurate."[5] This JSS captures gymnasts' movements, then analyzes them and provides a score. The FIG president clarified that the primary aim behind the planned introduction and use of the AI-based judging system was to obtain accurate assessment of gymnasts' performance, without any biases or errors.[6] The International Olympic Committee too had expressed support for development and implementation of a new judging-support system of this sort, citing the importance of Olympic sports being judged fairly and transparently, and Fujitsu began development of the JSS in 2016, with about 100 people working on the project full-time since. Two of them are collaborating with the FIG to develop the CoP such that it is compatible with the JSS. This entails formulating more precise rules for a digital form of the CoP.

By means of 3D sensing technology, the JSS captures a multi-angle view of the gymnast's movements during the routine. It would then analyze them against the digitalized

element definitions from the CoP artifact. On the basis of this analysis, the JSS then would indicate mistakes, deductions, and a final score. The Fujitsu-proposed judging system uses 3D laser sensors, combined with AI-enabled joint-position-recognition software (Fujiwara & Ito, 2018). Via depth images obtained via these sensors, it derives the posture and position of the human body from the joint positions.

This 3D sensing technology oscillates many lasers on a scale of about 2 million points per second, detects the reflected light, and calculates the distance to the target object (point cloud). It then recognizes the joint positions from this shape; calculates hands and feet positions [sic], bending of joints, etc.; and finally compares those results with model data of human movement in a database to derive differences in movement. (Fujiwara & Ito, 2018)

The JSS can provide a visual representation of each element along with all technical details of how it should be executed (for instance, the length and height of a jump, the number of steps taken before and after it, the angle of the turns, etc.). Figures presenting Fujitsu's conceptualization of how the skeleton-recognition technology works can be found on the website of Fujitsu.[7]

At the moment, Fujitsu's efforts are focused on ascertaining D-scores via the JSS, though the company foresees the JSS supporting the decision-making process connected with execution scores too. At present, the E-score poses an obstacle in that the point-deduction system currently in place is described in vague terms. For instance, what constitutes "straight" or "slightly bending" can be interpreted differently from one judge to the next. In their current form, the rules cannot readily be transplanted to an AI-powered judging system. This is why Fujitsu and the FIG are jointly developing a refined rulebook that specifies precise angles and ties them in with the execution score (Fujiwara & Ito, 2018).

One of the judges among our informants expressed the following vision for the execution score:

In the future, the system will tell you the difficulty and show you all your execution marks and where [why] you got it... And it allows you to select from about 10 different icons that [represent] different [...] angles of the body. (Nick, judge)

The expectation is that, thanks to its technical capabilities, the JSS will resolve some of the challenges related to the process of human judgement in artistic gymnastics. For instance, the JSS may improve the accuracy of judging by

---

providing a multi-angle view of the gymnast's routine, which goes beyond what the human eye can capture from a single human's point of view. It could also significantly speed up the process of judging. Furthermore, the system is expected to manifest neutrality and impartiality, affording elimination of human biases and subjectivity from the judging. There are hopes connected with another vexing issue too: visualization and recording of all steps in the evaluation process should enable provision of explanations and clarification of the scores for the gymnasts.

Thus, while supporting the judges' work, the system may simultaneously assist gymnasts and coaches with their training since, according to its developers, it "can accurately inform athletes regarding their stability and the exact angle between their joints, so they can make appropriate adjustments and improvements."[8]

While the current objective is for the AI-powered system to support judges with their judging process during official international competitions, exactly how it will be made part of that process is less clear. Multiple options can be readily envisioned: giving input for joint decision-making with human judges, providing a reference value for each judge's decisions, serving as a backup in the resolution of inquiries, etc. Down the line, the system might even become the main source of scores, with the human judges there only to serve as a backup when something goes wrong with the system.

While the FIG and Fujitsu assure that deploying a fully autonomous system that replaces the human judges is not part of the plan, many judges raised associated concerns nonetheless, such as these:

> Of course, you should understand that some [of the] atmosphere among the international judges is fear, and laughing. Is it so that in the future we're not needed? Some people are totally against this system now. It's coming. We can be against it, but anyway it comes. We can't deny it [...] I would want personally to continue judging and not to be replaced myself by some [...] robot. (Felicity, judge)
>
> I think I would be concerned if I felt like artificial-intelligence technology is going to replace the judges entirely. I think that I would be concerned about the future of the sport. It's not fair. (Abby, judge)
>
> For the judges, this system is not good, because one day we will not be necessary anymore. Maybe in 10 years, we will not have judges. But I'm a judge: I love to do what I do, just afraid of not being needed anymore. I just want to keep my job. (Bella, judge)
>
> The system has potential. Of course, this could be very helpful and useful if it's needed. But if it's not needed,

please let humans stay in the judges' panel. I do hope it will not replace humans. (Nick, judge)

Moreover, some participants highlighted further implications of top-level judges' removal, ultimately leading to the collapse of the entire knowledge-sharing system, from the top all the way to the grassroots.

> If it's replacing human judges altogether, that would be catastrophic for the sport. Why do these athletes train? Why do they do 30 hours a week and keep on with injuries and delay school and everything? That's for the Olympics. If you take that out and there's no international travel anymore, it's over, and that's the same for judges. Because you really have to invest. I take at least two or three weeks, either unpaid or vacation time, to go to those events. If you take international competitions out of the equation because there are AI systems, I'm sure you're going to take [away] the real motivation for many people to keep on judging and do really hard work. And all those guys that have very much experience and that can help coaches and athletes, they just are not going [to] do it anymore. So I think you could lose the top of the pyramid. (Bob, judge)
>
> If the system can provide all final scores in the future, then you don't need the judges anymore [...]. There's another point that you have to think about – the people that are now working here as judges: they are the main judges of their countries. So in their countries, at home, they are [running] the judges' courses for other people; they are working in the federations. If you cut them, all the federations […] will lose their best people. (Lilly, judge)

Our informants raised additional financial and practical concerns related to the implementation of an AI-powered system, referring to high financial costs for implementing such a system for training purposes and the new technical skills required for setting up and working with it. High costs for installing the system would result in unequal access to the system, arising from many local gymnastics federations' lack of financial resources.

> Obviously, it's going to be a very expensive system. At the very top, there [is sufficient] financial means for using the system, [but] how is it sloped down to each federation, to each area within each region? Because, obviously, you're not going to be able to have this system in every program. (Charlie, judge)
>
> I think it might be a really expensive program to replace what we have already. I'm from a very small federation with no money. We can't buy the system. We don't have money for it. So only big federations will get the system. (Lilly, judge)

---

[8] Information about the system can be found at https://medium.com/syncedreview/meet-fujitsus-ai-gymnastics-judges-8cb52613b2a.

Maybe it will work at European/World Championships or the Olympic Games. But when you gonna come to Africa, how you gonna bring all these technologies and artificial intelligence to Africa? We have no technical support or financial resources for it. (Sarah, judge)

In addition, starting to use the system will necessitate additional training. According to FIG representative Simon, plans are in place for special training camps where the judges can learn how to use the new system, which he acknowledges is going to bring further costs and resource needs. Our observations indicate that, in the meantime, the judges' level of awareness of the system's capabilities and their knowledge of how it works is relatively low.

I've already made the comment to Fujitsu that if you do the pilots within federations, the first one has to [be performed in] the countries where they organize open-training camps where people from other countries can come to get feedback. (Simon, FIG representative)
I think that the judges still need to be trained to make a decision supported by this system. You need to know how it works. If you have to work as a judge, then you need to know how to make it work. And you need to know what it can tell you, what information it can give you and how it gives you the information, and also how quickly it could give you the information. (Abby, judge)

5 Paradoxical Tensions Related to AI-Powered Judging

From the data analysis, we can posit six paradoxical tensions related to AI-powered judging as conceived of in our study. Below, we describe each tension in turn and detail the associated challenges related to developing and implementing the AI-powered system. The discussion draws on input from key stakeholders canvassed in our study of judging in artistic gymnastics.

5.1 Paradoxical Tension 1: Accurate AI Is Too Exact

The first tension identified arises from the accuracy of the AI-based judging system, its perceived excessive exactness in judging. Most informants stated that an AI-powered judging system will be more accurate than human judges, stating that a "machine" and technology are always more accurate than human beings. Moreover, it is clear that there is some sense of competition among human judges with regard to accuracy; they take pride in how accurately they can judge, and precision helps them advance to panels for international competitions.

We're like the athletes: we compete among ourselves; we want to get to really accurate scores and to score consistently, because when our scores are analyzed [and] you have too much variation, you're not going to be selected for the Olympics. (Bob, judge)

The AI-powered system may seem to be the ideal judge in this sense. However, according to some interviewees, one possible disadvantage of such a system is precisely its precision: excessive exactness in judging is not suitable for gymnasts, since they are not perfect – they cannot implement some elements to millimetric accuracy or to a specified degree of an angle.

The accuracy of judgements made by humans is a matter of some dispute. Variously, our interviewees pointed to accuracy shortcomings within our research context as possibly linked to limitations of people's physical capabilities, inability to maintain concentration, fatigue, sensory perception of information, cognitive capabilities, speed of processing information, and inappropriate levels of knowledge internalized by judges. Let us look at each of these before considering how an AI-powered system enters the picture.

Informants characterized the constraints of human sensory and cognitive capabilities thus:

The human eye and human brain can't work so fast and accurate. There are too many decisions to be taken, so for a human brain it's not possible to do it. In one second, you have to make maybe 8–10 decisions [...] this is almost impossible because it's all at the same time. (Norman, judge)
What a human eye sees is one thing, but what the machine sees is more accurate. I heard [...] that it's very useful, because there were a couple of inquiries especially on the rings: the human eye said "no," but when they saw the angle in three dimensions it was a very little difference. The human eye can't detect it, but the system did, so the [appeal] was accepted. (Edward, judge)

The second factor is the human fatigue and loss of concentration that stems from the sheer duration of international competitions. With the same panel of judges having to work for eight or 10 h straight, sustaining the same level of attention and concentration throughout the day is nearly impossible.

Of course, when you're sitting down and you have six sub-divisions in one day and you have it over two days so you're spending 14 hours a day in the gym, yeah, it's really hard to be fresh from the first moment of the first day until the last moment of the last day. (Charlie, judge)
I was sitting with the light in my eyes. Toward the end of the day, it was stressful with the lights. Yes, of

course, there's a human aspect. You compete at the beginning of the day [or] in the middle of the day, when the judges might be tired or thirsty or hungry, needing a break. All of these human components can injure [the results]. (Sarah, judge)

The issue of the level of internalized knowledge on judges' part may involve either lack of knowledge or, in contrast, too much influence of training and experience. Both may result in mistakes in judging.

And it's bad enough that you can make your own mistake anyway. And how do you tell the difference between somebody making a mistake because of an error and some little bit of incompetence (that's [considered] to be cheating)? It's very difficult to read this. (Sarah, judge)

Moreover, our informants indicated that the challenges plaguing human-based judgement could be resolved through an AI-powered system's advanced technical capabilities. In the stakeholders' opinion, the system under development is more impartial and should objectively yield high accuracy while also expediting the judging process. Indeed, almost everyone we interviewed cited the system's expected high accuracy as an advantage, especially for evaluation of the technical dimensions of a routine, such as the height of an element, angles, and speed.

It helps the accuracy. It also helps the speed of the program [...]. So, when there are doubts, the answers are already available [...]. The sky is the limit in its ability to help the judges feel secure about their evaluation and to be able to better educate the judges. (Simon, FIG representative)
I think that artificial intelligence can provide an accurate and detailed breakdown. (Charlie, judge)
The computer can do better, can better see angles, and it's more precise than [a] human. (Lilly, judge)
I believe that an electronic judging system can be more accurate than human judges. Especially for angles: if you make a turn, it would be perfect if the system could see it better than human eyes can see. (Felicity, judge)

However, tension is evident with regard to the accuracy advantage an AI-powered system might possess over human judges. While judges saw merit in exactness in judging, they also indicated that the current system of human judging provides some kind of equivalence between the inaccuracy of judges and that of the gymnasts – judges do not judge accurately, but neither do gymnasts perform accurately. Both work to a level of precision that is "good enough" and leaves room for artistry. In contrast, a judging system that demands exactitude would not provide such balance: there

were concerns about gymnasts being unable to provide high enough accuracy in their performance to satisfy the system's requirements.

My worry is that this system is too perfect. Right now, we're humans. Gymnasts are humans. We as judges note certain deductions, certain angle-based deductions. Sometimes 45° is very difficult to recognize for a human eye. But if a camera [shows] 44.9°, [AI] does not accept the exercise; it makes a deduction. But for a human eye, the normal eye, it may pass. The gymnasts will be mad at the judgement with the machine because it's going to catch every single mistake they make. It will ask perfection of the gymnasts. Too much perfection. (Edward, judge)
But the disadvantage is for the gymnasts themselves. For example, at the high bar you do something almost to a handstand and a judge will say, "Okay, it's a handstand." They won't be too strict, while a machine will see it's 97°; they will take one tenth [off]. (Fabian, gymnast)
I don't know if it is something interesting for the athletes. I think it is too precise and that perfection will be pursued and it will be too difficult. (Bruno, fan)
The exactness of a new electronic judging system may become a challenge for the gymnasts and may have an overall negative influence on the popularity of artistic gymnastics as it will lower the level of gymnasts' performances and scores. If they really measure everything – the angles and rotations – so precisely, the scores will be lower. So they could work with margins, but then you diminish the rationale for using computers. For example, the machine could say that you can deviate from 10°. (Kyle, coach)

The precision of a newly introduced electronic judging system could, hence, pose a challenge for the gymnasts while also exerting a generally negative influence on the popularity of artistic gymnastics through these effects on the gymnasts' performance and the visibility of lower scores.

5.2 Paradoxical Tension 2: "Objective" AI Can Be Biased

From our findings, a paradoxical tension became evident with respect to biases also, where key stakeholders perceived AI-powered systems as a way of addressing the challenges linked to biases in current human judging. Interviewees expressed a belief that the new system would be free from such biases and, therefore, more objective and neutral in its evaluations – after all, it should not be affected by such "human factors" as prejudice and personal preferences. In reality, however, AI may introduce new types of bias, which

could end up even harder to spot in a system that is assumed to be bias-free.

Judges' biases and subjectivity pose a vexing problem for artistic gymnastics, as is reflected by the fact that most informants highlighted this issue. Athletes are victims of biased and unfair judgements by the panel of judges, which may be rooted in any of various factors: emotions, personal preferences, initial expectations, familiarity with the routine or athlete, prejudice for or against a particular country, informal guidelines' influence on the judging process, etc.

> We have this subjectivity inside us. (Bella, judge)
> We don't want to be subjective, but sometimes it is how it is. (Nick, judge)

Studies have identified numerous sources of bias in humans' judging process. Memory biases (related to recall, testimony, and hindsight) may be present, as might similarity, desire, mere-exposure, order, rule, complexity, test, and anchoring and adjustment biases (Mazurova et al., 2021).

Memory biases (recall, testimony, and hindsight bias) could come into play in that judges may unconsciously rely on memory if unable to remember each particular detail of the routine. Our memory is not always a reliable source of information. A judge who cannot recall all the details of a routine may unconsciously try to "recall" the missing parts by filling in the gaps, possibly with unconscious reliance on experience. Hence, the judge may assume the athlete to have done certain things that were not actually done. Clearly, a certain level of approximation exists in the evaluation process. Judge Ulla referred specifically to the large amounts of approximation involved in judges' evaluation of heights, speeds, and angles in an athlete's performance, and she stated that scores are estimated on the basis of this approximation. As for anticipation, another judge, Sarah, stated, "When a gymnast runs [up], I can tell you if it's going to be a catastrophe or not. We anticipate. Anticipation helps you sometimes with your judgement. You have not only what you see at this moment; it's much bigger than just what you see."

Desire, similarity, and mere-exposure bias manifest themselves in relation to judges' expectations or personal preferences. For instance, familiarity with an athlete and his or her earlier success/failure may predispose a judge to expect a certain quality of performance. Judges' personal expectations may cloud their objectivity and cause them to perceive the routine as better or worse than it truly was. Likewise, certain countries' preeminence may steer judges toward giving higher scores to those nations' athletes.

> If you're really familiar with the routine, it can influence your judgement positively or negatively. Maybe you don't see a mistake, because you see it all the time and you get used to it. Or maybe you see every lit-

tle mistake that they make better. Familiarity with the routine can move your judgement up or down. (Abby, judge)
> Sometimes judges and coaches [...] set up a good relationship. Even though the judge here is supposed to be neutral and working for all the countries, she still has a little affiliation with some country. (Sarah, judge)

Even if no bias actually exerts an influence, judging a gymnast from one's own country may induce greater stress, as judge Bob highlighted: "It happens sometimes that, because you know all the mistakes and you are not surprised to see them, you can end up giving a lower score than the average judge. If that happens for an important routine, the gymnast can lose because of that, so there could be repercussions from your country." If instead assigning too high a score, "you're going to get flagged by the FIG."

In addition to the factors mentioned above, we found a significant influence of routines' order on the judging process. Judges often give undue weight and attention to the first and the last things encountered; this corresponds with the notion of *order bias*. In a related phenomenon, an athlete who competes in the morning is likely to receive a lower score while one performing in the evening gets a higher score. Our informants confirmed this.

> It's always like this: if you compete in the morning, judges are harder on you; they easily take away many more points. They want to be good/strict and do their job properly. Thus, if you compete in the morning, they can make a bigger […] deduction, and in the evening if you do exactly the same mistake, they will not take so much from your total score. (John, gymnast)

Regarding *rule bias* in artistic gymnastics, some interviewees identified unofficial guidelines on keeping one's scoring "average" throughout the day and on not giving an excessively high score to a "perfect" routine.

> Judges have a certain average from a morning competition, and they need to keep this average between morning and evening scores. So they are afraid to give high scores from the start, as it will be harder for others to get a higher score in the evening, so they need to keep this average between the morning and the evening score. Thus, they don't give too good scores in the morning, and the better scores come in the evening. (Mark, gymnast)

As for perfect scores, gymnast James said, "Human judges, even if they see something perfect, like a perfect routine, they can't leave the papers empty. They need to find something [wrong] in the routine, to fill in the papers. That's why it's so hard to get 10.0 nowadays."

Complexity bias is clearly present, due to such factors as information overload, time pressure, human fatigue, lack of accuracy, the need to pay constant attention, and perceived importance of the judges' task and responsibility. Extensive approximation in the judges' process arises from the limits of the human brain, which can process only a certain amount of information within a given time, like it or not. Judge Felicity summed up the issue: "We don't want that, but we all make mistakes when we judge."

Under the influence of *anchoring and adjustment bias*, judges may tend to root their judgement either in the previous achievements of the athlete or in their first impression of the gymnast. The initial anchor for the judge might be the athlete's ranking or "usual" performance level. And the first impression, once made, is very hard to change or adjust.

> Especially in your home country, it's usually not fair when the judges know you and have seen you so many times during the training so they kind of know already where you will do your mistake. So if you don't do it in the competition, they think like, "Oh, he usually makes this mistake, so it will be a mistake now" even if you do it very well. I like to compete more internationally, where judges don't know me. I usually score higher points. (John, gymnast)

Regarding AI-powered systems, both athletes and judges stated that these systems would be fairer and capable of assessing all routines equally. Such a system should prove unbiased and more objective in that it has no "anticipations" or expectations for a given athlete's performance, and no other human factor should affect the system's judging process.

> *AI doesn't care which country you're from. It evaluates the technical side of the performance. Judges can hear very often from the coaches that we've been biased with their athletes, and if the routine is evaluated by the system, who can you blame for low scores? Nobody. Because AI is unbiased. It's objective.* (Ulla, judge)
> *We can make some mistakes about an objective thing, but the system can't.* (Bella, judge)

Others, however, were able to point to potential issues of assumptions embedded in AI. Biases might arise especially in relation to body morphology, skin color, and the style of performance.

> The system assesses the routine based on where it identifies the joints of the gymnast in the 3D skeleton model. So if the joints are positioned wrongly for certain body types, it will make an error. The system will always make this error, systematically: [...] it can be wrong. (Bob, judge)
> People do have structural differences, like, for example, bow-leggedness, someone whose legs will not

come together or where they make adjustments with how they bring their feet together or make it appear as if their feet are together. So there are structural concerns there, and there is potential for body biases, which is essentially like a racial bias because there are different body structures. (Katarina, judge)

> Physiologically, I don't think the technology can do that, take into account the different morphologies of gymnasts. They will standardize the body type of a gymnast. Physically, smaller gymnasts are better at gymnastics than taller ones. A taller gymnast has longer arms and legs, which make it more difficult to rotate. They already standardize more and more because they recruit especially small children. Other, taller gymnasts... it is not that they are bad, but they will have more difficulties and it will take more time. I think that the danger of this technology is that we will see all gymnasts with the same body type, height, musculature. (Fabian, gymnast)

> I know that they were saying that they had problems utilizing the Fujitsu program with exceptionally small bodies: […] the female gymnasts who are, like many of them, under five feet... I also heard that for people with darker-colored skin the system was not working in terms of being able to track them accurately. And maybe a bias based on gymnastics' [style in a specific] country. The Chinese tend to like a very, very straight line, whereas the Western countries tend to prefer a more open, aesthetic arching line. So whoever wins that argument, I guess, that could build a bias into the analytical program. (Anonymous stakeholder)

5.3 Paradoxical Tension 3: Even Black-Boxed AI Represents Explainability

The third tension is related to the perceived explanation capabilities of black-boxed AI. Our informants felt that, via the new technology represented by the AI-based system, the current judging system's lack of explainability and interpretability can be resolved, at least partially. This stands in contradiction with the challenges associated with AI black boxes' inherent *lack* of explainability and interpretability.

The lack of explanation in today's judging system poses a substantial problem for gymnasts and coaches, as section 4.1.2 attests. They identified providing explanations and clarifications of the results as important for advancing the training and development of the athletes, and they stressed that they would like to receive an explanation of the deductions and of the final scores given by the judges.

> It's really hard to get some explanation from the judges about what I did wrong. (Mark, gymnast)
> We need to figure out what was not perfect. The judges don't tell us what exactly we get the deduction for. (John, gymnast)

Such frustrations may lie behind these key stakeholders' perception of the AI-based judging system as a way to allow better-organized, systematic provision of explanations regarding scores. Judges too told us that supplying an explanation to athletes is important and that it would be very useful for athletes if an AI-powered system could do so. In the opinion of our informants, the electronic judging system holds potential to offer such explainability benefits as more explanation and greater transparency of the judging process, pinpointing of details that could inform the training process, and easier verification of the judging process. Cited in particular were the advantages of some explanation and clarification of the results over what the human judges provide and new possibilities for effectively gaining and accumulating in-depth knowledge.

Additionally, the system's ability to present the results immediately, alongside an indication of how it has come to these conclusions, may be of great use. Also, those inquiries not precluded via the system's use might see more rapid resolution and be handled more accurately, at the competition itself.

> When they have something appealed – I mean an inquiry for the superior judges and also for the Technical Committee members – they will use this system to help them to evaluate, again, the whole routine. (Harry, judge)
> That could be helpful when there's an inquiry. We had some cases here [when the Fujitsu system was being piloted] where the [appeal] was accepted. Then, it's really helpful. (Nick, judge)

Again, while supporting the judges, the system was described as able to assist athletes with their training too. For instance, Ulla stated that the new system, if able to provide information about how the decision on the final score and deductions was reached, could definitely help coaches improve the training process. This benefit could be expected to exert a positive influence, in turn, on the athletes' perception of such AI-powered systems. Moreover, it could facilitate drawing new fans to the sport.

> When you see a routine as a layman and you think it's great but then the gymnast gets a low score, you respond, "I don't understand this. Why would I watch it?" So using this technology could make you understand better why a score was given. (Jacob, gymnast)

At the same time, some interviewees recognized that parts of the AI-powered system are not going to provide explainability. They cannot. For instance, judge Katarina commented, "Apparently, there's some machine learning that's going on there, and that brings about the whole black-box issue."

5.4 Paradoxical Tension 4: While Used for Artistic Gymnastics, the System cannot Assess Artistry

The fourth tension involves the consequences of removing the human element by implementing an AI-powered system.

While reducing the "human factor" and human emotionality may enable the judging system to avoid certain biases and subjectivity and to yield highly accurate and objective judgements, some interviewees stated that, precisely because of its lack of "humanity," the AI system lacks the ability to evaluate a crucial feature of artistic gymnastics – namely, the artistry of the gymnasts' performance. This inability could eventually lead to elimination of the artistic component of artistic gymnastics. Informants worried about this, emphasizing the importance of the artistic aspect of a gymnastics performance. One put it thus: "Taking artistry out of competition is the same as taking the soul out of gymnastics" (Josh, judge). These concerns point to new challenges in the process of judging, which bring about a corresponding value–costs tension.

> It's called artistic gymnastics. And artistic is the key part of it, how it looks. I don't think the machine really can take up this part. We have artistry, we have a lot of things. Beautiful things. (Norman, judge)
> The computers don't understand what is artistic. If in artistic gymnastics judging is completely done by the computer, it's not artistic gymnastics. This is then something like online games. (Harry, judge)
> The computer will not see the faces of the gymnasts. This is the most important – the face, eyes, smiling. How can the computer analyze this? And then what about the music? In the end, we will have every exercise look the same. And the personal style of athletes will be lost. (Lilly, judge)
> Gymnastics is the sport of emotion. Artificial intelligence has no emotions to this point. (Charlie, judge)
> Yes, the technology can probably measure the amplitude of the legs perfectly, but I cannot imagine that technology can evaluate the gracefulness or whether the performance is in synch with the rhythm of the floor music. (Lauren, gymnast)

5.5 Paradoxical Tension 5: A System Intended for Humans Lacks Human Interaction

Similarly, our interviewees expected the technology to be unable to provide the same level of human interaction that gymnasts experience today with the human judges. Some judges stated that human interaction between the judges and athletes during routines is a crucial part of the performance. This interaction, while important, is not always visible – it is formed of details: the athlete greeting the judges before starting the routine, a slight nod of approval after completion of a routine, a fleeting smile from a judge. All of these make the athletes feel more confident and afford a strong sense of real human presence and interaction at the competition. Every judge interviewed assumed that this

integral part of competitions will be reduced or eliminated if an AI-based system ends up judging the athletes' performance. Accordingly, the fifth tension identified in our study is that of an AI-powered system that is expected to evaluate human gymnasts and interact with human judges yet devoid of the human-interaction element. Some judges stated that the lack of humanity in the AI-powered system could become a stumbling block to this system's large-scale implementation in artistic gymnastics.

> Gymnasts standing in front of a computer and saying, "Hi, I'm starting my exercise." That's kind of weird for me. We're part of the competition, and there should be always a human aspect of judging at the competition. (Nick, judge)
>
> I'm not quite sure how the athletes will feel. When an athlete does a good exercise [routine] and looks over to present to the judge and sees the reaction of the judge, I think that's something that is a human emotion that gives that athlete a good feeling's worth. And if the judge [offers] a sympathetic look even though the routine was not good, maybe the athlete still knows that there's someone who is cheering about the performance. Well, I'm not sure if artificial intelligence will be able to provide that type of feedback to the athlete. (Charlie, judge)

5.6 Paradoxical Tension 6: Consistency Requires AI Adaptability

The final tension identified becomes evident as the AI-powered system's introduction brings a need to codify the CoP as a digital artifact. This artifact must be much more specific in how the rules are specified (e.g., defining "straight" explicitly as 180° or as 179–181°). The expected result is consistency in judging: the same routine performed by the same gymnast should always get the same score.

> With human judges, you can't compare across competitions in terms of the scores obtained. At one competition, you perform a good routine without a fall and get 11.6. The next competition, you perform a routine with a fall and get 12.4. This is kind of ridiculous and should go away with this technology. (Damian, gymnast)

Codifying the rules so brings a danger of becoming static, however. One informant explained in the following words: "If humans are no longer as closely involved in applying and interpreting them, who will adjust the rules as the athletes and sport evolve? Changes to the rules over time are an important aspect of the sport, partly because gymnasts start 'playing the system.'" Judge Bob elaborated on the latter phenomenon.

> It's like doing your tax [return]: you optimize for the regulations. If you see that gymnasts take advantage of

the rules, then you try to fix it for the next cycle. In men's gymnastics, for instance, over time the importance of difficulty has increased and that of artistry has decreased.

He cited another major reason for human tweaking of the rules too: a desire to keep the sport both exciting and safe.

> If we systematize everything and we have no human judges anymore, how do you change the rules? Where do you take gymnastics? It takes human input from athletes and coaches [to answer questions] like "what is a dangerous exercise?" Some exercises are beautiful, but they are too difficult and dangerous for the sport […]. Unlike sports like running the 100 m, gymnastics evolves. If you lose the top guys, you have a catastrophe. Who can keep it entertaining and safe? With static rules, gymnastics would get more and more boring.

Summarizing our findings, Table 2, below, outlines the six paradoxical tensions identified in the empirical study.

6 Discussion

For the emerging body of literature on in-depth research into AI, our case study offers several contributions, to knowledge both of AI in its own right and of its implications for business operations and society at large. Below, we discuss the theoretical and practical implications of our research against the backdrop of prior literature.

6.1 Implications for Theory

Our articulation of six paradoxical tensions associated with AI applications in connection with judging in competitive sports contributes to scholarly understanding of tensions related to specific IT artifacts in their use context (Jarvenpaa & Lang, 2005; Majchrzak et al., 2013). While AI already outperforms humans in several cognitive and perceptual domains and is on the verge of doing so in many more, we recommend proceeding with a degree of caution. Prior research has highlighted a need to manage the tensions and contradictions, affecting stakeholders of diverse types, that are associated with an emerging IS artifact (Benbya et al., 2021; Dubé & Robey, 2009; Jarvenpaa & Lang, 2005). Building on the existing body of knowledge and with succinct examples from human vs. AI-powered judging in the sport of competitive gymnastics, we were able to put the spotlight on these tensions and consider the multiple stakeholder perspectives that are at play in the decision-making processes. The paradoxical tensions exhibit implications for theory on four fronts: configurations of human–AI hybrids, reversal to digital objects, multifacetedness of biases, and transparency of decision-making.

In the first of these arenas, our analysis of how stakeholders perceive the tensions and contradictions brought

Table 2 Summary of the core paradoxical tensions in AI-powered judging

	Paradoxical tension	Description
P1	Accurate AI is too exact	While human judges strive to be as accurate as possible, they must be content with approximations. In contrast, AI – on account of digital properties – can judge a gymnast's routine in an exact manner. However, gymnasts are not perfect; they cannot implement their routines with the level of accuracy to which AI-powered systems judge them. Hence, while the inaccuracy of human judges and that of gymnasts are in balance, there is imbalance between gymnasts' and AI-powered systems' accuracy.
P2	"Objective" AI can be biased	The removal of human judges' bias is no guarantee that biases stemming from the AI implementation are not going to emerge.
P3	Even black-boxed AI represents explainability	The AI-powered system and its algorithm are still black-boxed for many experts and require better explainability and interpretability. Yet the AI per se is assumed to present gymnasts with an explanation for the judging process, in contrast against the current, opaque-seeming judging system.
P4	While used for artistic gymnastics, such systems cannot assess artistry	Artistry, which is fundamental to artistic gymnastics, is bound up with human emotion and preferences. An AI-powered system employed for judging artistic routines is inherently incapable of evaluating artistry.
P5	A system intended for humans lacks human interaction	Interaction between the gymnast and the audience (including judges) is an essential part of artistic gymnastics. With AI, there is not the level of human interaction that gymnasts are used to experiencing at competitions.
P6	Consistency requires AI adaptability	While scoring by human judges shows a lack of consistency and an AI-powered system should prove consistent, improving the AI and adjusting to the disciplines' evolution requires adaptation. For instance, in the longer term, adaptation of the rulebook will be required. However, this becomes improbable if no humans keep up with the rules anymore.

on by the AI-based system points to a human-supporting rather than human-replacing role for AI. This is in line with suggestions from the literature review that such a hybrid human–AI configuration may involve cycles wherein the focus alternates between developing for automation (without human experts' involvement) and augmentation (informed by experts) (Raisch & Krakowski, 2021; van den Broek et al., 2021). Also dovetailing with recent work, our framework of paradoxical tensions points to the importance of engaging multiple categories of stakeholders when developing an AI-powered system, for supporting its relevance and ongoing utility (van den Broek et al., 2021).

Secondly, with regard to the articulation of the role of AI, we found that several of the six tensions stem from incongruities perceived by the various stakeholder groups in relation to what they saw as the characteristics of AI and digital objects. This ties in with the notion of ontological reversal, from physical objects to digital ones (Baskerville et al., 2020). Indeed, the AI-powered system draws on a digital representation of the athlete's performance, as opposed to a physical one. This transition, or reversal, from physical objects to digital objects (Kallinikos et al., 2013) and digital processing (Salovaara et al., 2019) seems to be generating pushback from stakeholders as the digital representation's exactness and improved accuracy push them out of their comfort zone – however desperately needed they know these to be. Furthermore, the ontological reversal precludes the evaluation of artistic notions of gymnastics and also brings

tensions related to the role of domain experts for creation, sharing, and use of knowledge (van den Broek et al., 2021).

Additionally, by demonstrating how AI can at the same time reduce and introduce bias, we engage with current discourse on bias of AI systems. While we identified expectations among informants that the AI-powered system will be objective and free of bias (van den Broek et al., 2021), other perceptions were evident too. This is consistent with mounting evidence that AI-powered systems may acquire, replicate, and even amplify (implicit) human biases present in the training data used for learning from past performance–evaluation pairs (Shrestha et al., 2019). For instance, some performance-evaluation systems have displayed race and gender biases (Aysolmaz et al., 2020; Nadeem et al., 2020), thus drawing attention to the importance of detail-level discussion of what type of AI and biases may be involved.

Fourthly, our work enriches understanding related to conceptions of intractability associated with AI-powered systems that utilize machine learning (Faraj et al., 2018). What we add to the emerging literature on AI-powered systems (Rai, 2020) is connected mainly with an apparent increase in transparency of evaluation processes relative to human judging and with lack of interpretability in *human* decision-making. Our work responds to calls for research into the implications, for multiple stakeholders, of explainable AI (Asatiani et al. 2021), which is "a prerequisite for fair, accountable, and trust-worthy AI" as well as for managing, justifying, evaluating, improving, and learning from AI (Meske et al., Forthcoming, p. 1). Our study can be considered in light of the notion of AI and fairness also, which is a crucial

consideration in implementing AI. As van den Broek et al. (2019) did in a human-resources context, we found that it is important to consider the accuracy of the information and the consistency visible in decision-making. A related paradox we identified lies in the view that AI increases transparency and explainability. The literature suggests that eliminating bias and making the process more fair and transparent can be expected to yield positive outcomes for multiple stakeholders, including acceptance, trust, satisfaction, a sense of commitment, and engagement behavior (Heiniger & Mercier, 2018; Konovsky, 2000). Our data analysis indicated that as the decision-making process grows more (seemingly) transparent, the various parties get more willing to accept potentially adversarial outcomes. Our findings mesh with those of Rzepka and Berger (2018), who revealed that the transparency of an AI system's decisions or actions significantly influences users' behavior, and of Xu et al. (2014), who found a positive correlation between system transparency and users' satisfaction with recommendation systems. Our findings point to judges having more positive perceptions of a system if it seems more transparent. Transparency in the decision-making process leads to a sense of greater informativeness and enjoyment too, thereby also enhancing the perceived decision quality and increasing system acceptance (Meske & Bunde, 2020).

6.2 Managerial Implications

Our study contributes to IS scholars' practical knowledge of the digitalization of sports (Goebeler et al., 2021; Xiao et al., 2017). The work builds on prior understanding of how IT artifacts have changed the face of decision-making in competitive and professional sports, by articulating the role of AI-powered systems for judging in competitive gymnastics. Complementing contributions from earlier literature, we have captured key relationships and tensions expressed by multiple stakeholders in connection with AI use in competitive sports. These shape the impact of the decisions made in the design of future systems. Hence, awareness of them could help to reduce human error and otherwise inform effective rollout and implementation of the AI systems now under development for such domains as competitive sports. In particular, acknowledging the paradoxical tensions could lead to development of mechanisms for avoiding or confronting them. The latter would be preferred, since it involves fuller understanding of the technology's features in aims of optimizing the system's use in a particular context.

Also, the study has broader implications for fan engagement in competitive sports. Research has already shown that fans become more involved when data are shared (Cortsen & Rascher, 2018). This is highly relevant from a marketing point of view, in that some sports (gymnastics among them) suffer from a lack of interest between Olympics years. For spectators, an AI-based system could enrich competitions' streaming services by providing not just real-time, understandable-seeming judging results but also quantitative indicators such as height and stance stability, to guide the audience to greater appreciation for the sport.[9] Thus, the technology can add some concrete numbers to stadium viewing experiences and broadcasts both, making such sports more attractive to watch. Prior work indicates that technologies that greatly improve the correctness of judging decisions can contribute to a more attractive competition (Leveaux, 2012). Another factor worthy of attention is that, in Fujitsu's words, "the technology also could add entertainment value to TV broadcasts and phone apps. One day, it could represent potential revenue for gymnasts themselves, who will be able to market and monetize their data."[10]

7 Conclusion

Our study shed light on paradoxical tensions accompanying AI-based evaluations. The contributions notwithstanding, it has its limitations, though. The first, associated with single-case-study designs, raises the issue of generalizability of the findings (Walsham, 2006). Nonetheless, single-case studies are a "typical and legitimate endeavor" (Lee & Baskerville, 2003, p. 231) in IS research, permitting one to gain rich insight into the interplay of various issues in application of AI systems and the accompanying tensions. This was possible only via in-depth work on articulation of the contradictions from the case's empirical reality. Against this grounding, future research can validate the findings from our research and generalize them beyond the single context of AI use in competitive gymnastics.

The study's most crucial contribution lies in identifying paradoxical tensions in complex socio-technical relationships amid the work to introduce an AI-powered judging system, in light of the perceptions of multiple stakeholders. Our findings suggest that the adoption of AI technologies for decision-making is by no means straightforward, and our outline of the related tensions lights the way for future research on AI adoption and management across diverse settings and stakeholder landscapes. In summary, we hope that we have evoked the energy and power related to exploring paradoxical tensions (Lewis, 2000), which may open avenues for fruitful debate of the organizational and societal implications of using AI, and that we have enriched the emerging body of knowledge on its transformative potential (Raisch & Krakowski, 2021).

---

[9] M. Sarazen highlighted this aspect in the *Synced* piece "Meet Fujitsu's AI gymnastics judges," published on January 26, 2019; see https://syncedreview.com/2019/01/26/meet-fujitsus-ai-gymnastics-judges/.

[10] This statement, made by a Fujitsu official, was quoted by L. Clarck in the October 9, 2019, *Washington Post* article "Computer-assisted judging is being tested at gymnastics world championships," available at https://www.washingtonpost.com/sports/olympics/computer-assisted-judging-is-being-tested-at-gymnastics-world-championships/2019/10/09/214a515e-eab9-11e9-9306-47cb0324fd44_story.html (accessed on October 10, 2019).

Appendix 1: Guide to the interview topics

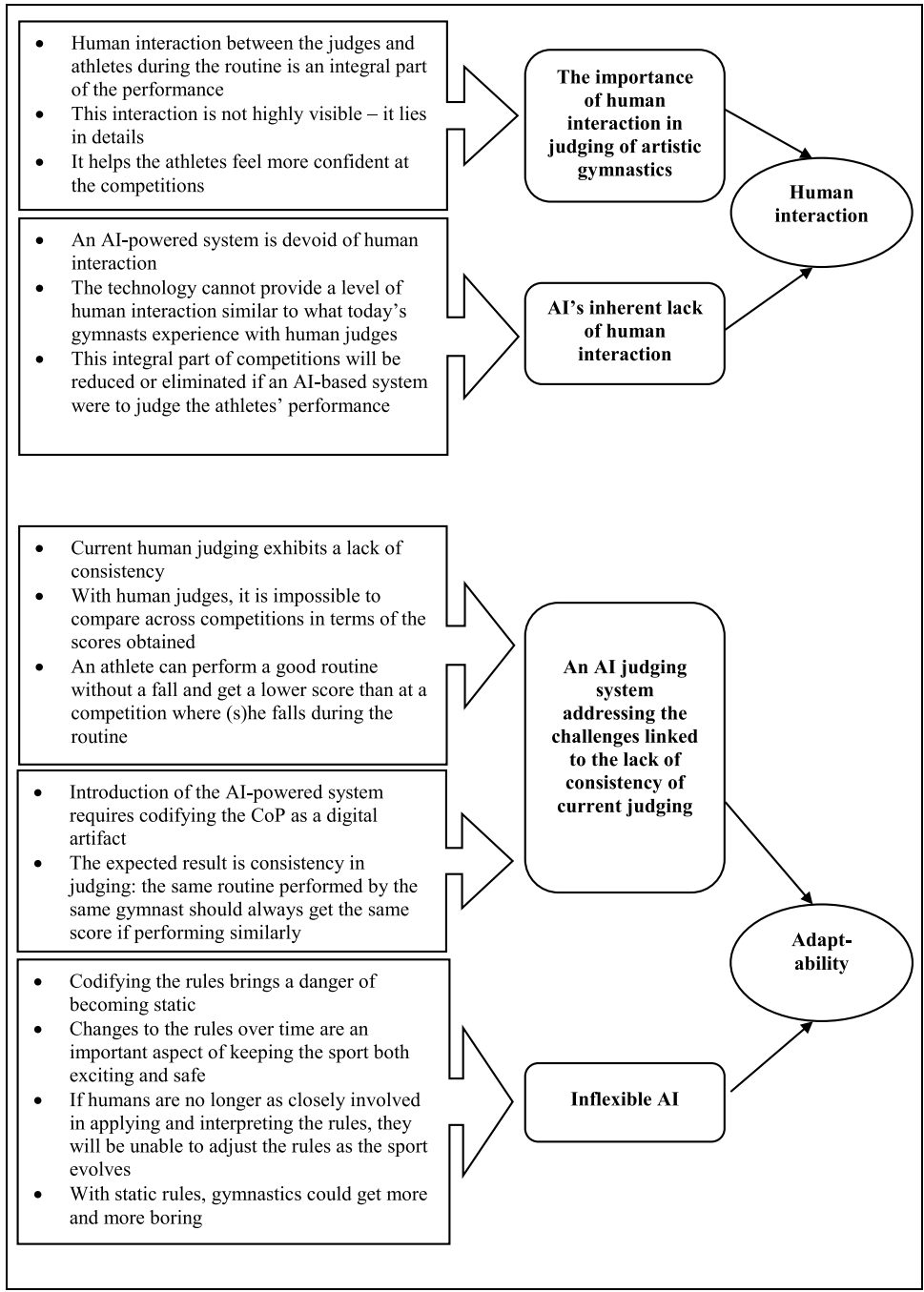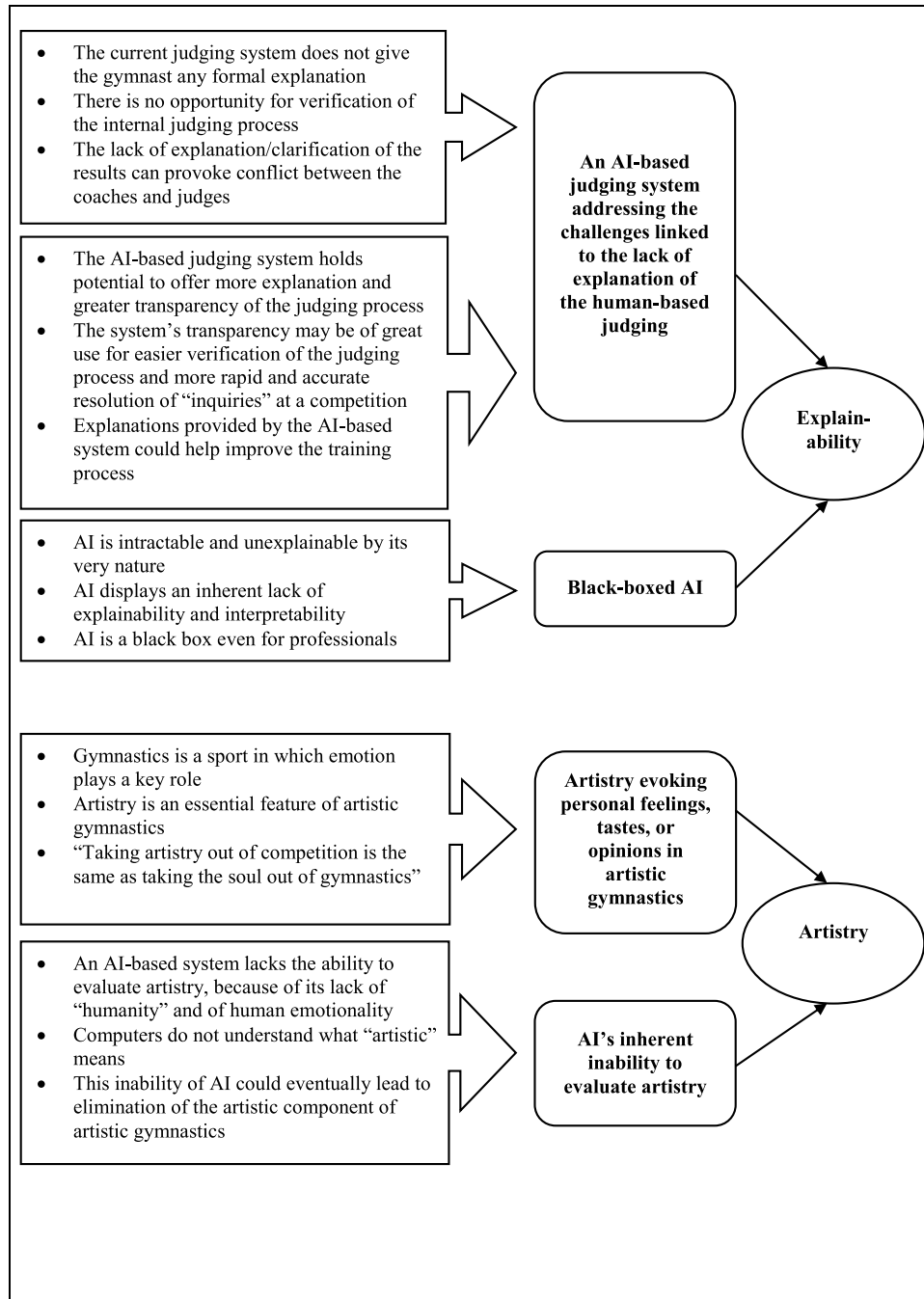| Sub-focus of research	General questions for all stake-holders
The stakeholders' general aware-ness of the AI-powered system	– Have you heard about a new, AI-powered system? How do you feel about it? – Have you experienced using any AI-powered system at a competition? – Do you understand how the AI-powered system works / how it comes up with a score?
Stakeholder understanding and awareness related to such systems' technical capabilities	– What do you know/think about the technical capabilities of an AI-powered system? – What do you think about... – the fairness of human judging and AI-powered systems? – biases in human judging and AI-powered systems? – the accuracy of human judging and AI-powered systems?
Comparison between the judging systems – the AI-powered one and the human panel	– What are the main advantages and disadvantages of each type of judging system – human- and AI-based? – In your opinion, what is the main difference? – Which system type is prefer-able?
Stakeholder perceptions of the new system	– What is the opinion of the gym-nasts/judges/coaches about the new, AI-powered system? – Could you summarize your opinion on the AI-powered system? – Would you trust the AI-powered system? – Are you concerned or excited about the new, AI-powered system?
	Questions for specific groups of stakeholders
Introduction of the technology	– How did the idea of introducing an AI-powered system arise? – Who is taking the lead in the system's development and implementation? – How will the system be imple-mented?

Sub-focus of research	General questions for all stake-holders
The system's development and technical characteristics	Fujitsu: – How did the idea of introducing an AI-powered system arise? – To what extent are judges/gym-nasts/federations involved in the development? – Could you specify the role of hardware, of software, and of data in this system? – How will the AI-powered system affect the panel of human judges? – Will the evaluation process be changed through the imple-mentation of the AI-powered system? – Which parameters affect the decisions produced by the AI-powered system? – What other applications for this technology exist, outside sports?
Explainability	Gymnasts: – Do you get any kind of explana-tion/clarification of the final result from the judges? ∘ If so, please explain how this works. ∘ If not, do you think explanation should be provided? ● If so, then what form would be preferable: oral/visual/both?
The training process	Coaches: – The developers state that an AI-powered system could improve gymnasts' training process. What do you think about this? Do you agree? Why / why not? – How could using AI-powered systems in the training process influence the performance of a gymnast? Would it benefit the gymnasts' performance?
Local implementation of AI-based judging support	Technical directors with local federations: – What is the position of the local branch of [specific federation] on the system's procurement and introduction? – Do you think the system could be used for every performance apparatus?
The system's attractiveness to spectators	Fans: – As a fan, what do you think of this AI-powered system? – How could the sport develop in terms of fan engagement on account of this system? – Do you think this system will attract more people to watch/fol-low the sport?

Appendix 2: The data structure

(First-order) concepts	(Second-order) themes	Aggregate dimensions
• Human-based judging is prone to errors • Human judges demonstrate a high level of approximation in their evaluation • Judges' evaluation accuracy is limited by their physical and cognitive capacities • The accuracy of human judgements is a matter of dispute	An AI judging system addressing the challenges linked to approximation and inaccuracy in human judging	Exactness
• The existing limitations in the accuracy of human-based judgement could be resolved through an AI-based judging system's advanced technical capabilities • "Machines" are more accurate than human beings • The AI-based system yields high accuracy (especially in evaluation of the technical dimensions of a routine)		
• The AI judging's disadvantage is its precision in gymnasts' evaluation • Gymnasts are unable to provide high enough accuracy in their performance to satisfy the system's evaluation criteria • An AI judging system's exact evaluation could pose a challenge for the gymnasts	Imbalance between the human performance and the AI's exactness	
• Athletes are often "victims" of biased and unfair judgements by the panel of judges • Judges' biases may be rooted in various factors: emotions, personal preferences, initial expectations, familiarity with the routine/athlete, prejudice for or against a particular country, informal guidelines influencing the judging process, etc.	An AI-based judging system addressing the challenges linked to biases of human judges	Bias
• The AI-based judging system is free from human biases and, therefore, more objective and neutral in its evaluations • The AI-based system is not affected by such "human factors" as prejudice, personal preferences, or prior expectations for a given athlete's performance • AI-based judging is fairer and capable of assessing all routines equally		
• AI may introduce new types of biases • Algorithmic biases of AI may be evident especially in relation to body morphology, skin color, and performance style • AI biases are hard to predict and spot in a system that is assumed to be bias-free	Biased AI	

- Human interaction between the judges and athletes during the routine is an integral part of the performance
- This interaction is not highly visible – it lies in details
- It helps the athletes feel more confident at the competitions

The importance of human interaction in judging of artistic gymnastics

- An AI-powered system is devoid of human interaction
- The technology cannot provide a level of human interaction similar to what today's gymnasts experience with human judges
- This integral part of competitions will be reduced or eliminated if an AI-based system were to judge the athletes' performance

AI's inherent lack of human interaction

Human interaction

- Current human judging exhibits a lack of consistency
- With human judges, it is impossible to compare across competitions in terms of the scores obtained
- An athlete can perform a good routine without a fall and get a lower score than at a competition where (s)he falls during the routine

- Introduction of the AI-powered system requires codifying the CoP as a digital artifact
- The expected result is consistency in judging: the same routine performed by the same gymnast should always get the same score if performing similarly

An AI judging system addressing the challenges linked to the lack of consistency of current judging

- Codifying the rules brings a danger of becoming static
- Changes to the rules over time are an important aspect of keeping the sport both exciting and safe
- If humans are no longer as closely involved in applying and interpreting the rules, they will be unable to adjust the rules as the sport evolves
- With static rules, gymnastics could get more and more boring

Inflexible AI

Adapt-ability

Appendix 3: Summary of the codes used in the analysis and related example quotes

Code	Example
The lack of accuracy of human judges	
– Cognitive limitations of the human brain	"The human eye and human brain can't work so fast and accurate. There are too many decisions to be taken."
– Human fatigue	"When […] you're spending 14 h a day in the gym, yeah, it's really hard to be fresh from the first moment of the first day until the last moment of the last day." "[T]he judges might be tired or thirsty or hungry, needing a break."
– Human error	"[Y]ou can make your own mistake anyway."
The higher accuracy of an electronic system	"What a human eye sees is one thing, but what the machine sees is more accurate." "The computer can do better, can better see angles, and it's more precise than [a] human."
The AI-based system being too exact	"[T]his system is too perfect […]. It will ask perfection of the gymnasts. Too much perfection." "[I]t is too precise and […] perfection will be pursued and it will be too difficult." "The exactness of a new electronic judging system may become a challenge for the gymnasts […]. If they really measure everything – the angles and rotations – so precisely, the scores will be lower."
Subjectivity and prejudice of human judges	"We have this subjectivity inside us."
– Expectations of the judges	"We anticipate. Anticipation helps you sometimes with your judgement."
– Familiarity with the gymnastics routine	"If you're really familiar with the routine, it can influence your judgement positively or negatively."
– Friendship between the judges and coaches	"Sometimes judges and coaches [...] set up a good relationship."
– The athletes' order of performance	"[I]f you compete in the morning, judges are harder on you."
– Unofficial guidelines	"Judges have a certain average from a morning competition, and they need to keep this average between morning and evening scores."
– Familiarity / one's expectations of a given athlete	"[I]t's usually not fair when the judges know you and have seen you so many times during the training so they kind of know already where you will do your mistake."
The greater objectivity and impartiality of an electronic system	"Judges can hear very often from the coaches that we've been biased with their athletes, and if the routine is evaluated by the system, who can you blame for low scores?" "We can make some mistakes about an objective thing, but the system can't."
New biases caused by AI	

Code	Example
– Body morphology	"People do have structural differences […]. So […] there is potential for body biases, which is essentially like a racial bias, because there are different body structures." "Physiologically, [...] the technology [...] will standardize the body type of a gymnast [...]. I think that the danger of this technology is that we will see all gymnasts with the same body type, height, musculature."
Lack of explanation	"The judges don't tell us what exactly we get the deduction for."
Potential increase of explainability and transparency of the judging process	
– Handling of inquiries/appeals	"When they have something appealed [...] they will use this system to help them to evaluate, again, the whole routine."
Black-boxed AI	"Apparently, there's some machine learning that's going on there, and that brings about the whole black-box issue."
The artistic aspect of the gymnastics	"It's called artistic gymnastics. And artistic is the key part of it, how it looks."
Inability to evaluate artistry	"The computers don't understand what is artistic."
Human interaction between judges and gymnasts	"We're part of the competition, and there should be always a human aspect of judging at the competition."
Lack of human interaction on the part of the electronic system	"Gymnasts standing in front of a computer and saying, 'Hi, I'm starting my exercise.' That's kind of weird for me."
The lack of consistency of the current judging process	"With human judges, you can't compare across competitions in terms of the scores obtained."
Gymnastics becoming static through codification of the rules	"If humans are no longer as closely involved in applying and interpreting them, who will adjust the rules as the athletes and sport evolve? Changes to the rules over time are an important aspect of the sport."
Inflexible AI	"If we systematize everything and we have no human judges anymore, how do you change the rules? […] With static rules, gymnastics would get more and more boring."

Appendix 4: Brief explanation of each artistic-gymnastics discipline

Floor exercises (men's and women's) are performed on a special gymnastics mat of 12 × 12 m. The discipline combines individual elements (somersaults, splits, handstands, etc.) and various dance elements. In the performance, athletes should make the most of the entire mat space. The complexity of the routine and its elements is evaluated, as are the purity and confidence of the execution. Also important are originality

and the artistry of the performance, especially with regard to the women's discipline, since these exercises include specifically chosen music and individual dance steps and rhythmic movements. The duration of floor performances is limited to 1 min, 10 s for men and 1 min, 30 s for women.

The vault (men's and women's) is a jump performed with a running start and use of additional support. The apparatus is 1.6 m long and 0.35 m in width. The athlete runs along a special track 25 m long and 1 m wide, pushes off with his or her feet from a shock-absorbing device, and makes an additional push from the apparatus with the hands (for men, pushing with one hand is allowed). The jump may be straight, somersault, overturn, etc. Men have one attempt at this jump, while women have two attempts and the average score for the exercise is displayed. For evaluation, the important parameters are the height and distance of the jump, its complexity (the number of revolutions around the longitudinal and transverse axes etc.), purity of execution, and the precision of the landing.

The apparatus for *the Parallel Bars (for men) and Uneven Parallel Bars (for women)* consists of two wooden poles fixed to a metal base. For men, the apparatus height is 1.75 m; for women, the bars are 1.65 and 2.45 m above the surface of the safety mats. Men's routines using the parallel bars combine dynamic elements (rotations, swinging movements, etc.) and static ones (horizontal stops and handstands), and the gymnast must use the full length of the apparatus, exercising above and below the bars. A woman's routine with the uneven bars includes turns in both directions around the upper and lower bar, using one hand or both, and various technical elements completed above and below the bars, with rotation around the longitudinal and the transverse axis.

The Pommel Horse discipline (for men) involves an apparatus with handles at a fixed height of 1.05 m (as measured from the surface of the safety mats). The routine is a combination of handstands with swinging and rotational movements of the legs. The athlete must use all parts of the apparatus.

For *the Rings (men)*, movable wooden rings are attached to special cables. They are at a height of 2.55 m from the surface of the safety mats. Exercises using the rings (lifts, turns, and twists) should demonstrate the athlete's plasticity and physical strength. The static elements of the routine are no less difficult than the dynamic ones. The rules dictate that jumping from the rings apparatus at the end of the routine must be an acrobatic element, while the athlete may use the aid of a coach or other assistant to take up the starting position on the rings.

The Horizontal Bar (men's discipline) uses a steel bar with a diameter of 27–28 mm and a length of 2.5 m fixed at a height of 2.55 m (as measured from the safety-mat surface). The rules specify that the athlete must not touch the bar with his body when performing rotations in any of various directions. In the course of the performance, the athlete should demonstrate several types of grip and an ability to shift from one type to another.

The Balance Beam (for women) involves an apparatus 5 m long and 0.1 m wide that is rigidly fixed at a height of 1.25 m from the floor. The balance-beam routine is a composition that comprises dynamic elements (jumps, turns, "jogging," somersaults, dance steps, etc.) and static ones (twine, swallow, etc.) completed while the gymnast is standing, sitting, and lying on the apparatus. The gymnast must use the full length of the beam. The judges evaluate her plasticity, balance, and artistry. A routine should be no longer than 1 min, 30 s.

References

Asatiani, A., Malo, P., Nagbol, P., Penttinen, E., Rinta-Kahila, T., & Salovaara, A. (2021). Sociotechnical envelopment of artificial intelligence: An approach to organizational deployment of inscrutable artificial intelligence systems. *Journal of the Association for Information Systems, 22*(2), 325–352.

Aysolmaz, B., Iren, D., & Dau, N. (2020). Preventing algorithmic Bias in the development of algorithmic decision-making systems: A Delphi study. *Proceedings of the 53rd Hawaii international conference on system sciences*, *3*, 5267–5276. https://doi.org/10.24251/hicss.2020.648.

Baskerville, R. L., Myers, M. D., & Yoo, Y. (2020). Digital first: The ontological reversal and new challenges for information systems research. *MIS Quarterly, 44*(2), 509–523.

Benbya, H., Pachidi, S., Davenport, T., & Jarvenpaa, S. (2019). *Call for papers JAIS-MISQE joint SI on artificial intelligence in organizations – JAIS-MISQE SI on artificial intelligence*.

Benbya, H., Pachidi, S., & Jarvenpaa, S. (2021). Special issue editorial: Artificial intelligence in organizations: Implications for information systems research. *Journal of the Association for Information Systems, 22*(2), 281–303.

Brynjolfsson, E., & Mcafee, A. (2017). The business of artificial intelligence. What it can — And cannot — Do for your organization. *Harvard Business Review*, 4–11.

Calabretta, G., Gemser, G., & Wijnberg, N. M. (2017). The interplay between intuition and rationality in strategic decision making: A paradox perspective. *Organization Studies, 38*(3–4), 365–401. https://doi.org/10.1177/0170840616655483

Can, H., Lu, M., & Gan, L. (2011). The research on application of information technology in sports stadiums. *Physics Procedia*, 604–609. https://doi.org/10.1016/j.phpro.2011.11.093

Cortsen, K., & Rascher, D. A. (2018). The application of sports technology and sports data for commercial purposes. In: *The use of technology in sports - Emerging challenges*. https://doi.org/10.1016/j.colsurfa.2011.12.014.

Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid Intelligence. *Business and Information Systems Engineering*. https://doi.org/10.1007/s12599-019-00595-2

Du, W. (Derek), Pan, S. L., Leidner, D. E., & Ying, W. (2019). Affordances, experimentation and actualization of FinTech: A

blockchain implementation study. *Journal of Strategic Information Systems, 28*(1), 50–65. https://doi.org/10.1016/j.jsis.2018.10.002.

Dubé, L., & Robey, D. (2009). Surviving the paradoxes of virtual teamwork. *Information Systems Journal, 19*(1), 3–30. https://doi.org/10.1111/j.1365-2575.2008.00313.x

Dumoulin. (2020). Accuracy and National Bias of figure skating judges: The good, the bad and the ugly. *MIT Sloan Sports Analytics Conference*.

Duong, A. (2008). Bias in the 2008 Beijing Olympics (gymnastics).

Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization, 28*(1), 62–70. https://doi.org/10.1016/j.infoandorg.2018.02.005

Ferger, K., & Hackbarth, M. (2017). New way of determining horizontal displacement in competitive trampolining. *Science of Gymnastics Journal, 9*(3), 303–310.

Flessas, K., Mylonas, D., Panagiotaropoulou, G., Tsopani, D., Korda, A., Siettos, C., … Smyrnis, N. (2014). Judging the judges' performance in rhythmic gymnastics. *Medicine and Science in Sports and Exercise, 47*(3), 640–648. https://doi.org/10.1249/MSS.0000000000000425.

Fujiwara, H., & Ito, K. (2018). ICT-based judging support system for artistic gymnastics and intended new world created through 3D sensing technology. *Fujitsu Scientific & Technical Journal, 54*(4), 66–72.

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2012). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods, 16*(1), 15–31. https://doi.org/10.1177/1094428112452151

Goebeler, L., Standaert, W., & Xiao, X. (2021). Hybrid sport configurations: The intertwining of the physical and the digital. *Proceedings of the 54th Hawaii international conference on system sciences*.

Harding, J., & James, D. (2010). Performance assessment innovations for elite snowboarding. *Procedia Engineering, 2*(2), 2919–2924. https://doi.org/10.1016/j.proeng.2010.04.088

Harding, J., Toohey, K., Martin, D., Hahn, A., & James, D. (2008). Technology and half-pipe snowboard competition — Insight from elite- level judges. *7th ISEA conference*. https://doi.org/10.1007/978-2-287-99056-4.

Heiniger, S., & Mercier, H. (2018). *National Bias of international gymnastics judges during the 2013-2016 Olympic cycle*. arXiv preprint arXiv:1807.10033.

Jarvenpaa, S. L., & Lang, K. R. (2005). Managing the paradoxes of Mobile technology. *Information Systems Management, 22*(4), 7–23. https://doi.org/10.1201/1078.10580530/45520.22.4.20050901/90026.2

Kallinikos, J., Aaltonen, A., & Marton, A. (2013). The ambivalent ontology of digital artifacts. *MIS Quarterly, 37*(2), 357–370.

Keller, R., Stohr, A., Fridgen, G., Lockl, J., & Rieger, A. (2019). Affordance-experimentation-actualization theory in artificial intelligence research - a predictive maintenance story. *40th international conference on information systems*, 1–17.

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals, 14*(1), 366–410. https://doi.org/10.5465/annals.2018.0174

Kerr, R. (2018). Technologies for judging, umpiring and refereeing. In: *Sport and technology* (pp. 114–134). https://doi.org/10.7765/9781784995201.00010.

Klein, H., & Myers, M. (1999). A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS Quarterly, 23*(1), 67–93.

Kolbinger, O., & Lames, M. (2017). Scientific approaches to technological officiating aids in game sports. *Current Issues in Sport Science (CISS), 2*, 1–10. https://doi.org/10.15203/ciss_2017.001

Konovsky, M. a. (2000). Understanding procedural justice and its impact on business organizations. *Journal of Management, 26*(3), 489–511. https://doi.org/10.1177/014920630002600306

Lee, A. S., & Baskerville, R. L. (2003). Generalizing generalizability in information systems research. *Information Systems Research, 14*(3). https://doi.org/10.1287/isre.14.3.221.16560

Leveaux, R. (2010). Facilitating Referee's decision making in sport via the application of technology. *Communications of the IBIMA*. https://doi.org/10.5171/2010.545333

Leveaux, R. (2012). 2012 Olympic games decision making Technologies for Taekwondo Competition. *Communications of the IBIMA*. https://doi.org/10.5171/2012.834755

Lewis, M. W. (2000). Exploring paradox: Toward a more comprehensive guide. *Academy of Management Review, 25*(4), 760–776. https://doi.org/10.1080/15358593.2015.1116592

Majchrzak, A., Faraj, S., Kane, G. C., & Azad, B. (2013). The contradictory influence of social media affordances on online communal knowledge sharing. *Journal of Computer-Mediated Communication, 19*(1), 38–55. https://doi.org/10.1111/jcc4.12030

Mazurova, E., Penttinen, E., & Salovaara, A. (2021). Stakeholder-dependent views on biases of human- and machine-based judging systems. *Hawaii International Conference on System Sciences*.

Mercier, H., & Heiniger, S. (2018). Judging the judges: Evaluating the performance of international gymnastics judges. *ArXiv*, 1–12.

Meske, C., & Bunde, E. (2020). Transparency and trust in human-AI-interaction: The role of model-agnostic explanations in computer vision-based decision support. *Proceedings of the international conference of AI in HCI 2020, 12217 LNCS*, 54–69. https://doi.org/10.1007/978-3-030-50334-5_4.

Meske, C., Bunde, E., Schneider, J., & Gersch, M. (Forthcoming). Explainable Artificial Intelligence: Objectives, Stakeholders, and Future Research Opportunities. *Information Systems Management*, 1–11. https://doi.org/10.1080/10580530.2020.1849465

Nadeem, A., Abedin, B., & Marjanovic, O. (2020). Gender Bias in AI: A review of contributing factors and mitigating strategies. *ACIS 2020 proceedings*.

Omorczyk, J., Nosiadek, L., Ambroży, T., & Nosiadek, A. (2015). High-frequency video capture and a computer program with frame-by-frame angle determination functionality as tools that support judging in artistic gymnastics. *Acta of Bioengineering and Biomechanics, 17*(3), 85–93.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Sage Publications. https://doi.org/10.1177/1035719X0300300213

Perederij, V. V. (2013). The problem of the quality of judging in rhythmic gymnastics. *Pedagogics, Psychology, Medical-Biological Problems of Physical Training and Sports, 3*, 43–46. https://doi.org/10.6084/m9.figshare.647325

Plessner, H., & Haar, T. (2006). Sports performance judgments from a social cognitive perspective. *Psychology of Sport and Exercise, 7*(6), 555–575. https://doi.org/10.1016/j.psychsport.2006.03.007

Plessner, H., & Schallies, E. (2005). Judging the cross on rings: A matter of achieving shape constancy. *Applied Cognitive Psychology, 19*(9), 1145–1156. https://doi.org/10.1002/acp.1136

Poole, M., & van de Ven, A. (1989). Using paradox to build management and organization theories. *The Academy of Management Review, 14*(4), 562–578.

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science, 48*(1), 137–141. https://doi.org/10.1007/s11747-019-00710-5

Rai, A., Constantinides, P., & Sarker, S. (2019). Next-generation digital platforms: Toward human–AI hybrids. *MIS Quarterly, 43*(1), iii–ix.

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review, 46*(1), 192–210. https://doi.org/10.5465/AMR.2018.0072

Robey, D., & Boudreau, M. (1999). Accounting for the contradictory organizational consequences of information technology:

Theoretical directions and methodological implications. *Information Systems Research, 10*(June), 167–185. https://doi.org/10.1287/isre.10.2.167

Rzepka, C., & Berger, B. (2018). User interaction with AI-enabled systems: A systematic review of IS research. *International conference on information systems 2018, ICIS 2018*, (December).

Salovaara, A., Lyytinen, K., & Penttinen, E. (2019). High reliability in digital organizing: Mindlessness, the frame problem, and digital operations. *MIS Quarterly, 43*(2), 555–578. https://doi.org/10.25300/MISQ/2019/14577

Schad, J., Lewis, M. W., Raisch, S., & Smith, W. K. (2016). Paradox research in management science: Looking Back to move forward. *Academy of Management Annals, 10*(1), 5–64. https://doi.org/10.1080/19416520.2016.1162422

Schön, D. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.

Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. *California Management Review, 61*(4), 66–84. https://doi.org/10.1177/0008125619862257

Siggelkow, N. (2007). Persuasion with case studies. *Academy of Management Journal, 50*(1), 20–24. https://doi.org/10.5465/AMJ.2007.24160882

Smith, W. K., & Lewis, M. W. (2011). Toward a theory of paradox: A dynamic equilibrium model of organizing. *Academy of Management Review, 36*(2), 381–403. https://doi.org/10.5465/amr.2011.59330958

Stefani, R. (1998). Predicting outcomes. In J. Bennett (Ed.), *Statistics in sport* (pp. 249–275). Hodder Education Publishers.

Tan, F. T. C., Tan, B., & Pan, S. L. (2016). Developing a leading digital multi-sided platform: Examining IT affordances and competitive actions in alibaba.com. *Communications of the Association for Information Systems, 38*(1), 738–760. https://doi.org/10.17705/1CAIS.03836

Taymazov, V., Bakulev, S., Pavlenko, A., Simakov, A., & Chistyakov, V. (2013). To a question of electronic refereeing systems application in taekwondo (VTF). *Uchenye Zapiski Universiteta Imeni P.F. Lesgafta*, 155–160. https://doi.org/10.5930/issn.1994-4683.2013.04.98.p155-160

van den Broek, E., Sergeeva, A., & Huysman, M. (2019). Hiring algorithms: An ethnography of fairness in practice. *40th international conference on information systems, ICIS 2019*.

van den Broek, E., Sergeeva, A., & Huysman, M. (2021). When the machine meets the expert: An ethnography of developing AI for hiring. *MIS Quarterly, 45*(3b), 1557–1580. https://doi.org/10.25300/MISQ/2021/16559

Walsham, G. (2006). Doing interpretive research. *European Journal of Information Systems, 15*(3), 320–330. https://doi.org/10.1057/palgrave.ejis.3000589

Wimelius, H., Mathiassen, L., Holmström, J., & Keil, M. (2021). A paradoxical perspective on technology renewal in digital transformation. *Information Systems Journal, 31*(1), 198–225. https://doi.org/10.1111/isj.12307

Xiao, X., Chian Tan, F. Ter, Lim, E. T. K., Henningsson, S., Vatrapu, R., Hedman, J., … Van Hillegersberg, J. (2017). Sports digitalization: An overview and a research agenda. *38th international conference on information systems*, 1–21.

Xu, J., Benbasat, I., & Cenfetelli, R. T. (2014). The nature and consequences of trade-off transparency in the context of recommendation agents. *MIS Quarterly, 38*(2), 379–406.

Yin, R. K. (2013). *Case study research: Design and methods*. Sage Publications.

Elena Mazurova is a Ph.D. candidate at Aalto University in the Department of Information and Service Management with a specialization in Information Systems. Her research interests include user perception, human-computer interaction, AI-powered systems, virtual and augmented reality, and e-commerce. Her prior work has been presented at the following conferences and workshops in Information Systems: SCIS/IRIS 2018 and 2019, HICSS-53, HICSS-54, and AMCIS 2020. In addition, her research on consumer choice in e-commerce was published in the Journal of Information Systems Engineering and Management. Elena's main practical expertise lies in tourism management and hospitality.

Willem Standaert is an Assistant Professor in the Strategic Marketing Innovation Department at HEC Liège, Management School of the University of Liège, and a Visiting Professor at Ghent University. His research interests are in the areas of virtual and hybrid meetings, digitally-enabled buyer-supplier interactions, and digital innovation, with a focus on sports and financial services. Willem's work has been published in academic journals such as Information & Management, Journal of the Association for Information Systems, Business Horizons, Electronic Commerce Research and Applications, Psychology & Marketing, Information Technology and Management, and Journal of Information Technology-Teaching Cases. In 2020, Willem was awarded the ARES exceptional prize "Gender and Covid-19" for research on virtual meetings and gender inequality.

Esko Penttinen is Professor of Practice in Information Systems at Aalto University School of Business in Finland. He studies coordination of work between humans and machines, organizational implementation of artificial intelligence, and governance issues related to outsourcing and virtual organizing. His main practical expertise lies in the development, assimilation, and economic implications of inter-organizational information systems, focusing on application areas such as electronic financial systems, government reporting, and electronic invoicing. Esko's research has appeared in leading IS outlets such as MIS Quarterly, Journal of the Association for Information Systems, Information Systems Journal, Journal of Information Technology, International Journal of Electronic Commerce, and Electronic Markets.

Felix Tan is an Associate Professor of Information Systems at the University of New South Wales (UNSW) Business School. His research interests include Digital Platforms, Information Systems for Good, Sports Digitalization, E-Commerce and Enterprise Systems. His research work has been published in various academic and practitioner journals such as Information Systems Journal, Information and Management, MISQ Executive, Communications of the AIS, International Journal of Information Management and Australasian Journal of Information Systems. Dr. Tan is presently serving as associate editor of Journal of Strategic Information Systems, Information Systems Journal and Information and Management. Dr. Tan is presently serving as the Association of Information Systems Region 3 (Asia-Pacific) Diversity and Inclusion Taskforce representative. Dr. Tan is presently serving on the SAP Academic Alliance Board (Asia-Pacific and Japan).