Towards Understanding People's Experiences of AI Computer Vision Fitness Instructor Apps

ANDREW GARBETT, Samsung AI Center, Cambridge, UK
ZIEDUNE DEGUTYTE, Samsung AI Center, Cambridge, UK
JAMES HODGE, Samsung AI Center, Cambridge, UK
ARLENE ASTELL, Samsung AI Center, Cambridge, UK

The recent rise in on-device AI computer vision and dialogue systems has facilitated a growing number of AI fitness related instructional apps. However, these technologies have yet to be explored within the HCI community. To investigate this domain we recruited 12 participants and asked them to engage with five recently launched AI fitness instructor apps. We interviewed participants and thematically analysed transcripts to understand their experience and expectations of these technologies. Our qualitative analysis outlines five main themes focusing on; *limitations of computer vision*, *visual feedback*, *dialogue with the AI*, *adapting to the user*, and *working out with the instructor*. Based upon our findings we present five design considerations for designers that relate to three key areas: *feedback and motivation*, *personalising the experience*, and *building a relationship with the AI*. We contribute a first look into people's initial experiences with on-device AI fitness instructor applications and we provide design considerations to guide future contextually-aware AI research in this domain.

CCS Concepts: • **Human-centered computing → User studies**.

Additional Key Words and Phrases: Computer vision, contextual AI, fitness, mobile devices, dialogue systems, qualitative methods

ACM Reference Format:
Andrew Garbett, Ziedune Degutyte, James Hodge, and Arlene Astell. 2021. Towards Understanding People's Experiences of AI Computer Vision Fitness Instructor Apps. In *DIS '21: ACM Conference on Designing Interactive Systems, July 28-2, 2021, Nowhere and Everywhere.* ACM, New York, NY, USA, 29 pages. https://doi.org/10.1145/3461778.3462094

1 INTRODUCTION

There is a growing interest in managing our fitness and personal health through the use of smartphone apps and activity trackers [65, 81, 86]. With the help of fitness tracking systems, users have been able to be more self-aware and monitor their health and control their fitness goals and motivations all in the comfort of their home [78, 88]. In these unprecedented times of a Global COVID-19 pandemic, the demand for home fitness applications [31] has increased worldwide. Through a combination of physical distancing measures forcing the closure of instructor-led fitness classes in shared spaces such as gyms, many people have resorted to engagement and motivation through online classes and adopting fitness apps into their daily routines [29]. Additionally, while it is well-known that the benefits of exercise to wellbeing are positive [26], knowing what type of activities can improve fitness and overall health has seen a surge in users seeking out instruction, feedback, and guidance through health and fitness technologies [68].

One promising technology development that aims to provide instructional feedback and personalization is the rising prevalence of computer vision-based AI for pose detection [13, 15, 18, 19, 70, 90], and more advances explicitly in mobile-based *on-device AI* computer vision for pose detection. Recent Software Development Kits (SDKs) such as Apple MLKit [5], and Google ML Toolkit [35] are supporting a new generation of on-device AI fitness technologies [23] that have the potential for wider adoption, given the prevalence of mobile devices, than existing sensor-based devices such as the Microsoft Kinect [93] that require extensive up-front hardware costs.

Inspired by computer vision advances, fitness apps have begun to adopt AI features resulting in several AI fitness instructor applications that incorporate both computer vision for pose detection and interactive dialogue systems. These new additions to fitness apps aim to understand a user's movements (accuracy, intensity, workout repetitions). By combining computer vision features with conversational agents, these fitness apps facilitate dialogue between the user and the AI instructor to offer verbal feedback, motivational statements, and allow the user to utilize voice commands for media playback control. Within the HCI community, research has yet to explore these forms of mobile-based computer vision AI instructor applications to understand people's experiences of these technologies and inform design principles. Understanding what features are important to users, what they value and what deters them from getting started or continuing with AI-powered fitness technologies is vital to inform and improve future development and meet the growing demand from an expanding, more heterogeneous user population. *Understanding* and *designing with* these advancements of AI is an emerging domain within HCI [4, 89] and industry where both are beginning to establish AI design guidelines [11, 36, 44, 75] to better understand how to work with this medium. This paper investigates the emerging trend in AI fitness instructor apps and offers preliminary insights into people's experiences and expectations of using these technologies. We invited 12 participants to workout with three out of five iOS AI fitness instructor apps over eight days to understand their experiences. We held one-to-one semi-structured interviews with each participant to understand the features they wanted and their reported accounts of the challenges and opportunities during their time with the AI fitness apps. From our thematic analysis, we identify the following five themes: *limitations of computer vision, visual feedback, dialogue with the AI, adapting to the user*, and *workout with the instructor*. Through engaging with this work, it became apparent that although fitness instructor technologies are somewhat capable of providing real-time feedback base on user's posture, they still fell short of an engaging and personalised instructional experience. In this paper, we present two novel contributions: a series of findings describing participant's experiences and expectations of AI fitness instructor apps, and propose five design considerations to guide future contextually-aware AI research in this domain.

2 RELATED WORK

Identifying users priorities for AI fitness apps aligns with the third wave of HCI, which embraces experience and meaning-making of technology user(s) [10]. User Experience (UX) which has emerged as central to the design process, requires an understanding of the individual's motivation to start and continue to use a product [40]. In the following section we identify existing HCI research focused on barriers for engaging in fitness, virtual trainers, and designing with AI.

2.1 Motivators and barriers for engaging in physical activity

To date, research from HCI has produced a large body of work on the capabilities of how technology can be used as a tool to motivate and maintain users in their physical activity. Prior user-centred studies have explored wearable systems [60], the considerations and motivations of different marginalised populations [39, 52], use cases for rehabilitation [34], leisure activities and long-term goals [88]. This prior work has called attention to several challenges in this domain such as a lack of motivation or meaning to workout [16], accessibility [58], social support through intergenerational communities for older adults [30], and how technology can be used to independently support users to achieve their short and long term goals [32]. A key consideration across this body of work is the importance of individual needs and agendas that must be considered when ensuring engagement in physical activity that echoes the typical approach of a fitness instructor who would traditionally personalise and tailor the clients training programme [22]. In response to these barriers to engagement, design and technology solutions have considered adopting gaming mechanics and sonification to entertain the support of rehabilitation movements [67, 79], auditory and haptic feedback for people who are blind or low-vision [74], producing a variety of personalised wearable sensors to increase self-awareness and safety, and social challenges to encourage conversations between students about their physical activity data [33].

Regarding the growing popularity of fitness and HCI research, we have seen a substantial increase in the development of commercially ready activity trackers such as Fitbit, Samsung Galaxy Watch, and Apple Watch [84] that are supported by a range of fitness apps such as Strava, Fitness Buddy, MyFitnessPal and many more [86]. To maintain motivation and continued use of these apps, they have adopted approaches similarly seen in research such as engagement through achievements to enhance accomplishment [63], social sharing features, notification reminders, and goal setting [92]. In Aladwan's [2] "expectations and experiences framework", they break the adoption and motivation of fitness apps into three elements: 'content' - 'utilities' - and 'character'. Content refers to specific app features such as progress tracking, feedback and goal setting which are key features for motivating users to achieve their fitness goals [3]. Utilities refer to app features and functions that help deliver the app fitness purpose. Apps that are difficult to use or do not assist users to meet their fitness goals are abandoned [42]. Character describes the users' psychological experiences from engaging with fitness apps, such as motivation and sense of achievement [2]. Moreover, while these studies have revealed the key considerations for motivation and adoption, recent research has considered the pitfalls and challenges that are present in long-term uses of fitness Apps. For instance, Stoica et al. [82] describes users abandoning apps among a sample of people in regular training for more than four years included limited customization and feedback, limited functionality, personal (e.g. manual entry) and financial (e.g. subscription or premium versions) costs, commitment required to follow a plan, expectations of the app and comparison of app with personal coach.

A key limitation of the current setup of mobile fitness apps is they take a one-solution-fits-all approach [86] that contradicts prior HCI research around the importance of considering the individuality of the user. While some fitness apps have added tailoring features at the start of their apps [92], these features are still a distant comparison to that of a fitness instructor who will personalise and adapt your exercise programme during the user's fitness journey [37]. To tackle the challenge of personalised feedback and fitness programmes in fitness apps, a series of AI fitness apps have been developed that aims to use AI to provide real-time feedback, repetition counting, personalised workout plans and long-term stat tracking. In the following subsection, we explore the current state of the art in Artificial Intelligence and how it may enhance the experiences of fitness apps.

2.2 Virtual Trainer Technologies

Within HCI, research around the factors that motivate people to seek and then adopt exercise technology began with the advent of systems that could be set up in the home. These initial systems were mainly gaming platforms offering the user opportunities for more active interaction with their screens by performing sport moves which were reflected back through actions on the screen. The popularity of these systems attracted research into the both the potential benefits of technology-delivered exercise [62] and what motivated users to engage with them [80]. Furthermore the creation of the

Special Interest Group for sports [64], workshop [59] and novel prototypes for technologies for instruction in specific sports such as; cycling [9], skiing [61] and swimming [91] demonstrate the growing interest in this domain.

The Microsoft Kinect [93] substantially advanced gesture control and pose estimation with inclusion of a depth sensor to detect distance and movement as well as four microphones to calculate the direction the voice is coming from. The Kinect also had voice recognition capability to aid in identifying which player is speaking. The home exercise genre became well established with the launch of the Wii (2006) and Xbox Kinect (2010) games consoles with titles such as Wii sports [56], Wii fit [57], and Kinect Sports [83]. In addition to the home fitness market, both the Wii and Xbox have been widely adopted in healthcare for rehabilitation [87] and physical therapy [43]. The limitation of these systems is the associated upfront costs of purchasing a games console and sensing peripherals. Although these original game-based systems are still popular, apps on mobile devices are widening access and participation in fitness at home.

While computer-vision AI is expected to provide deeper insights into the user's posture, Fitness AI apps also incorporate Conversational Agents (CA) to simulate meaningful and engaging conversational feedback. In a study by Clark [20] to gather user opinions on conversations with humans and CA, participants viewed conversations with the CA in practical and transactional terms. This contrasted with conversation with humans which was viewed as a social bonding experience. For example, participants felt that finding common ground with another person is a vital part of conversation. In contrast, when discussing common ground in relation to CA, participants generally disliked the idea of being likened or compared to a machine. However, participants noted that if the CA was able to learn user likes and dislikes over time, the conversation could become more personalised which could be likened to finding a common ground. As these system improve in their capabilities and user experience, they offer the potential for enabling more real-world interaction with AI. Furthermore there seems to be a general lack of research looking into user perceptions of the corrective feedback they have received. Pan [69] interviewed seven participants using a weightlifting application with wearable sensors that provides corrective verbal feedback. The participants preferred positive and encouraging feedback and found repetitive corrective feedback, i.e. repetition of the same phrases, during exercise upsetting and distracting. However, in a study exploring the experiences of 16 visually impaired participants using Microsoft Kinect 'Eyes-free Yoga' system providing auditory-only feedback from skeletal tracking on six yoga moves [73], most participants preferred to receive extra corrective feedback. This reflects the importance of the auditory channel for this particular population. In addition the researchers found that the participants became more confident with yoga poses and needed less verbal feedback over time.

The integration of vision-based fitness instruction with an everyday device such as the smartphone provides a unique opportunity for an affordable and widely adopted technology that we interact with frequently to become a personal fitness instructor. Breakthroughs in computer vision has removed the associated costs of led to the ability to estimate the human pose, 2D or 3D [13, 15, 18, 19, 48, 70, 90] using cameras as sensors and computer vision deep learning methods. These approaches have proven to be significantly more versatile and robust than prior methods, subsequently enabling tasks such as recognizing human activities [17, 55]. In addition, such methods have been suggested for use in sports science as the means for athlete evaluation [27, 28, 54]. As these technologies permeate into the mass market there is a proliferation of mobile (on-device) AI fitness technologies that are yet to be understood by the HCI community.

2.3 Designing with AI

According to Xu [89], advancements in AI since 2006 have kick started a third wave of AI. The improvements in AI applications such as speech and pattern recognition, deep learning models in addition to powerful computers, are helping to solve problems to meet human needs in their everyday work and personal lives [38]. The Human-centered

AI (HAI) framework defined by Xu [89] proposes technologies should be designed with human intelligence in mind to further enhance the AI, such as reflecting on human psychology and behavioural drivers. Finally, new technologies should consider the human factors of design, that refer to AI decisions and solutions being easily understandable and practical to the user.

Recently within HCI research, Amershi and colleagues [4] proposed design guidelines for human-AI interactions based on two decades worth of research and validated them through a user study evaluating a variety of AI technologies. The proposed guidelines [4] consist of 18 considerations separated into four key categories: *'Initial'*; *'During interaction'*; *'When wrong'*; and *'Over time'*. The researchers explained that during the *'Initial'* phase, the AI should make clear to the user what it can or cannot do. *'During interaction'* phase, the AI should be able to evaluate and produce contextually relevant information to the user. Furthermore, the AI needs to ensure it avoids undesirable social biases when making decisions or providing solutions to the user. *'When wrong'* phase refers to a user having an option to invoke, dismiss or reconfigure the AI-suggested solutions, which in turn promotes a sense of control. In addition the AI should be able to explain why it made any given decision. Finally, the *'Over time'* phase suggests that the AI should continuously learn from the user and adapt to the user's needs and preferences and always notify the user of any system updates. The technology industry has since developed similar design guidelines for working with AI; Microsoft, Google, IBM, and Adobe [11, 36, 44, 75], that define the capabilities and expectations of AI systems with an emphasis on the experience being personalised to the user through transparent and explainable AI to facilitate trust with users. Prior to Human-AI interactions guidelines [4], Brdiczka [11, 12] offered a conceptual definition of what human-centric AI *is* and *how* it should integrate with humans. Brdiczka borrowed from Context-Aware Computing [7] to define a set of requirements for what is described as *Contextual AI* that is capable of *intelligible, adaptable, customizable*, and *context aware AI* intelligence. Bradiczka defines *'Intelligible AI'* as AI that is transparent with a user and is able to explain its decisions in the context to what it knows, how it knows, and what it is doing about it [6, 11, 12]. *'Adaptable AI'* refers to AI that is able to learn and adapt to user's preferences and then can provide a personalised experience in different environments and situations [11, 12]. *'Customizable'* refers to AI that is fully configurable by the user, giving the user full control over AI decisions [11, 12]. Context is defined as *"any information that can be used to characterize the situation of entities (i.e., whether a person, place, or object) that are considered relevant to the interaction between a user and an application, including the user and the application itself. Context is typically the location, identity, and state of people, groups, and computational and physical objects"* [25]. According to [8, 11, 12] the *'Context-aware AI'* should have capabilities to perceive user's environment and situation and have adequate reasoning abilities to make intelligent decisions based on these contextual inputs.

The capabilities of Contextual AI as portrayed by Brdiczka [11] provides a definition with which to describe what is referred to as human-centered AI by [89] with the expectations of how to work with this material as proposed by the Human-AI Interaction guidelines [4]. In our discussion we return to Brdiczka's concept of Contextual AI [11] and associated principles of intelligible, adaptable, customizable, and context-aware AI.

Health and fitness technologies range from more clinical offerings to entertaining and engaging experiences that help motivate sustained use in efforts to improve health benefits more generally. AI has begun to permeate into fitness technologies for detecting pose [87, 87] and offer conversational agent systems for verbal feedback [69]. Advances in *on-device* computer vision [5, 35] in the past 18 months has led to the growth of AI fitness instructor mobile apps. As such, while prior work has explored AI verbal feedback, and the accuracy of posture detection, and current commercial fitness apps, the experiences and expectations of how AI contributes to the fitness app experiences has yet to be explored.

Table 1. Study Participants

ID	Sex	Age	Work	Edu. Level	Fitness level	App 1	App 2	App 3
P1	M	30	Web Dev.	BSc	Weak	VAY	Kaia	Fitness Ally
P2	M	31	Video Editor	BA	Weak	Kaia	Fitness Ally	Onyx
P3	M	26	Cloud Arch.	MSc	V. Good	Fitness Ally	Onyx	Zenia
P4	M	27	Health Ass.	MSc	Weak	Onyx	Zenia	Peloton
P5	M	53	CompSci.	PhD	V. Good	Zenia	Peloton	VAY
P6	M	43	Student	MSc	Weak	Peloton	VAY	Kaia
P7	F	19	Student	A Levels	Weak	VAY	Kaia	Fitness Ally
P8	F	32	Teacher	PGCE.	V. Good	Kaia	Fitness Ally	Onyx
P9	F	39	Business Man.	BA	Weak	Fitness Ally	Onyx	Zenia
P10	F	47	Teacher	PhD	V. Good	Onyx	Zenia	Peloton
P11	F	33	Teacher	BA	Weak	Zenia	Peloton	VAY
P12	F	19	Student	A Levels	Weak	Peloton	VAY	Kaia

Given the potential opportunities and challenges that AI fitness apps may offer, our study progresses from a growing body of fitness work in HCI to provide an understanding of participants experiences of a set of fitness AI apps.

3 METHODOLOGY

Recognising the ever-changing nature of at-home fitness technology, the authors of the paper selected five of the most popular AI fitness instructor mobile applications from the Apple iOS App Store: Fitness Ally: AI Workouts Coach [66], Onyx Home Workout [45], VAY Fitness Coach [1], Kaia Personal Trainer [41], Zenia Yoga & Flexibility [46], that were available as of August 2020. Our motivation for selecting iOS Apps was that the iPhone was the predominant platform choice for these applications (six apps vs. Android's two apps). Furthermore, we added Peloton [47], a leading fitness instructor application that does not integrate AI, for participants to compare and contrast AI and non-AI experiences. we invited 12 participants to workout with three of the six apps over eight days to explore the opportunities and challenges that may arise when interacting with AI computer vision Fitness Instructors. We captured and reviewed these sessions to understand the interaction experience during real-time workout sessions. To fully grasp the user's experience and reflections after using the three apps, we conducted one-to-one semi-structured interviews focusing on their wants and needs from future AI Fitness Instructor implementations.

3.1 Participants

We recruited 12 participants (6 females) by word of mouth. The participants ranged in age from 19 to 53 years old (M = 33.3; SD = 10.5), were based in the UK, and varied in occupations, education and fitness levels (Table 1). Each participant completed a screening questionnaire to ensure that they did not have existing injuries and had not previously used any of the mobile fitness applications selected for the study. Participants were also asked to list fitness apps they were currently using or had tried in the past. This list included MyFitnessPal, Strava, Fitbit Coach, Google Fit, Nike Training Club and Shreddy apps. Only two participants (P4, P5), had no prior experience of using fitness apps. Participants also reported on their current and past experiences of using mobile devices. Three participants (P1, P4, P10) had no prior experience using an iPhone. To assess participants' fitness level the current study used a modified fitness level assessment questionnaire based on [51] and [78]. After receiving signed consent forms, an iPhone 11 was posted to each participant, with detailed instructions on using and navigating the phone. Before sending the phones, the research team set up user accounts, downloaded the apps for each participant and paid any necessary subscriptions, to reduce the users' burden. Before the study, all participants were instructed to watch each app's promotional videos to familiarise themselves with all six apps being used in the study. Each participant was assigned to use three of the six apps. Participants were required to try each app for two consecutive days and then have a rest day before commencing with the next app.

Our motivation for selecting three apps (used for two days) over the course of eight days in total was to prevent novelty bias in participants' responses, by providing participants with the experience to compare and contrast these novel technologies. This also ensured that participants had experience of the design language and functionality embedded within each app. The eight day study period was also intended to minimize participant dropout rates due to the high physical activity required in the study.

We asked participants to screen record their workout so that we could verify these had been completed. The order of usage of the apps was randomized, resulting in each of the six apps being used in first, second, and third-order across our cohort. Each morning the team emailed a short questionnaire to the participants to gather information about their experience (duration of exercise, who was present in a room, and the area in the home that participants selected to exercise). The participants used the apps over eight days and then returned the phone. A semi-structured interview with each participant followed this to talk about their experiences using these apps. Upon returning the iPhone, the participants were compensated with a £100 Amazon voucher.

3.2 AI Fitness Instructors

We selected the five available AI fitness instructor mobile applications from the Apple iOS App Store Fitness Ally: AI Workouts Coach [66], Onyx Home Workout [45], VAY Fitness Coach [1], Kaia Personal Trainer [41], Zenia Yoga & Flexibility [46], that were available as of August 2020[1]as well as Peloton [47], a leading fitness instructor application that does not integrate AI. The features and functionality is compared in table 2 (see appendix).

Fitness Ally: AI Workouts Coach (v1.7, first released May 2020): Fitness Ally tracks the user's movements and provides real-time verbal instruction, corrective pose feedback and motivational statements. The app has 30 different types of exercises mainly focused on High-Intensity Interval-Training (HIIT) workouts that can be configured by intensity and duration (workout period/rest period); easy (20sec/40sec), medium (30sec/30sec) or hard (40sec/20sec). Fitness Ally is a female animated avatar voiced by a human and delivers the class in a follow-along format displayed in portrait. Users see the instructor by default and can see repetition count and an intensity graph measuring the user's movement and receive 'kinetic points' based on their form, endurance and exercise pace. The app supports voice commands for playback controls (i.e. stop, continue, skip).

Kaia Personal Trainer (v1.6, first released Jan 2019): Kaia Personal Trainer tracks users movements and provides visual and verbal real-time corrective pose feedback. This app was developed for people with musculo-skeletal conditions and focuses on exercises at a slower and controlled pace. The app has 100 exercises for leg, back, side, chest, and arm workouts. The app uses a text-to-speech 'robotic' voice and offers short workouts consisting of three exercises displayed in landscape. Classes present users with a short video which then switches to the front-facing camera, showing a skeleton outline tracking the user's body.

[1]With fitness technology being an ever-growing industry, many of the apps and the features have changed or improved from conducting our research study. While this is a limitation for papers that focus on the user experiences of apps, our findings and discussion remain relevant and timely towards the future development of contextual AI assistants, including fitness apps.

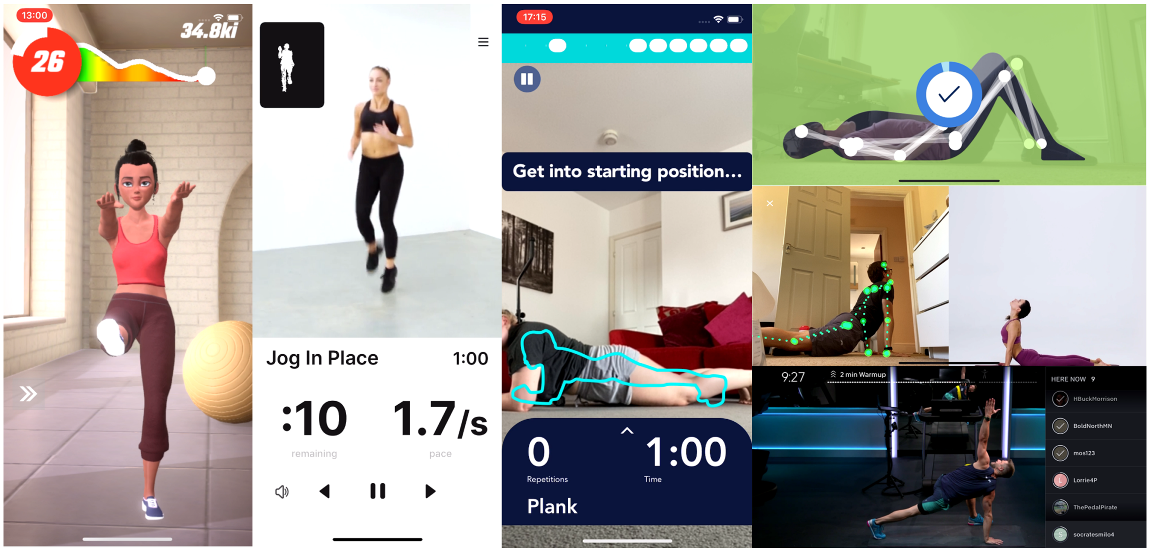

Fig. 1. Fitness Apps included in the study: *(left to right)* Fitness Ally, Onyx Home Workout, VAY Sports, *(top to bottom)* Kaia Personal Trainer, Zenia Yoga & Flexibility, Peloton

Onyx Home Workout (v1.7.32, first released April 2020): Onyx tracks users movements and provides visual and verbal real-time corrective pose feedback. The app claims to learn user's abilities and provide personalised programs. The app has 30 body-weight exercises and a selection of workouts (e.g. cardio, upper-body, full body). The user can choose one of four human instructor voices and can filter workouts based on available time, level and instructor. The class format consists of a series of video clips of different instructors working out with a small view of the front-facing camera showing the user's body outline in white on a black background displayed in portrait. The app tracks exercise repetitions, time remaining, and workout intensity displayed as number of repetitions per second.

VAY Fitness Coach (v1.2.0, first released June 2019): VAY Fitness Coach tracks users movements and provides visual and verbal real-time corrective pose feedback and motivational statements. The app has 11 body-weight exercises and offers six human instructor voices. Users are provided with a fixed body-weight workout selected by the instructor. The class format, similar to Kaia, presents a short video of an instructor and then switches to the front-facing camera to show a body outline starting position but does not provide skeletal overlay displayed in portrait. The app tracks exercise repetitions and time remaining.

Zenia Yoga & Flexibility (v3.8, first released Jan 2020): Zenia is a yoga-focused app that tracks user's movements and provides real-time visual feedback on pose. The app offers a choice of seven goals - improve posture, get in shape, improve flexibility, build strength, relax body and mind, relieve stress or get new experiences. The app has 35 beginner and 15 intermediate yoga poses and a selection of variable length workouts (e.g. cardio yoga, energy flow, standing sequences). The class format consists of a split screen in landscape showing a video clip of the instructor completing the exercise and the font-camera showing the user with a skeleton overlay that displays green or red dots to provide feedback on the user's pose. There are no metrics displayed in real-time however post-workout poses are scored out of 100 points based on the user's form. Users can review screenshots of exercises they complete and interact with a chatbot that provides workout feedback.

Peloton At Home Fitness (version 14.4.0, first released June 2018): Peloton Fitness contains a range of full body workouts, offering stretches, boot camp, HIIT, full body strength plus walking and marathon preparation. Users can join live-streamed classes with a choice of instructor or download recorded sessions. Peloton offers a selection of 34 instructors and classes that range in duration from five minutes to one hour. The classes are delivered in a continuous follow-along format. Users can see a predicted number of calories burnt and a countdown timer for the remaining workout.

3.3 Semi-structured Interviews

Three of the authors iteratively developed a semi-structured interview which was conducted using video conferencing software (Zoom). Our interview schedule consisted of open-ended questions on the following: 1) experiences of pre/during/post workout features, 2) experience with contextual AI features, 3) personalisation of workout experience, 4) visual and verbal feedback, and 5) creating the users perfect AI features. Two of the authors carried out the interviews, which typically lasted 1hr and were audio-recorded for later transcription.

3.4 Data Collection and Analysis

Our data comprised 759 minutes of audio transcriptions from interviews with 12 participants collected over a two week period. Our analytic approach followed Braun and Clarke's guidelines to Thematic Analysis (TA) [21], where we analysed our data in NVivo. Using a combination of TA and NVivo, our approach was particularly effective in identifying patterns as we had multiple coders and a large dataset. Our first step, was to come to an agreed initial codebook to ensure consistency of the labels throughout the entire dataset. This consisted of two authors independently generating codes and notes for two interviews (15% percent of the transcript data). Our second step saw the authors of the paper having multiple meetings to compare codes and organise them into potential themes which were then used to independently code the remaining data. Disagreement about the themes were resolved through discussion with the third author. Our last step was to give an overall structure to our analysis by agreeing on the final identified themes from our dataset.

For context on study engagement we report that participants used the apps for a total of 1242 minutes with an average workout duration of 17.25 minutes per session of use (MIN=6.6mins, MAX=37.5mins, SD=1.9) 1. Workouts were screen recorded by participants with duration calculated as the time between launching a workout and ending a workout, and does not include general use i.e. browsing the app. Each participant completed six workouts (two workouts per app using three apps) except for one participant who failed to complete one of the six requested workouts.

4 FINDINGS

Our findings are derived from the thematic analysis of 12 interview transcripts and we situate them within five themes: *Limitations of Computer Vision, Visual Feedback, Dialogue with the AI, Adapting to the User, and Workout with the Instructor.*

4.1 Limitations of computer vision

Computer vision is key to providing visual and auditory feedback within AI fitness apps. In some apps, participants received real-time verbal feedback to correct their posture or dialogue to motivate them. The participants identified and expressed their frustrations at technical issues relating to motion tracking and spatial limitations. These are described in the following sub-themes, where computer vision fails to accurately detect the participant or its functionality is limited due to the confined spaces participants are working out in. As a reminder, Peloton [47] is not included in this theme as the app does not incorporate computer vision or AI.

Detection issues: All 12 Participants highlighted a range of experiences with pose detection during the study. The majority of participants reported that the AI failed to recognise that they were in the correct position. This would often result in incorrect real-time feedback or repeatedly asking participants to return to a visible position for the pose-detection to work. Participants expressed that these detection issues became infuriating and patronising. For example, P1 mentioned how their experience with VAY and Kaia *"was just so much more annoying because it was like, "Please align yourself within the frame." Just over and over again while I'm desperately trying to move bits of furniture, lie back down, look across the phone to see if I'm in the frame then I'm not (in the frame)"*. Similarly, P2 commented on how the AI *"was having a bit of difficulty identifying between me and stuff in the background"*. However, instead of the AI assisting the participant to resolve the vision problem, P2 had to rely on trial-and-error to realise the detection issues eventually was due to the AI having *"a bit of trouble seeing the curves of the sofa"*. Additionally, P5 expressed frustration in the sensitivity of pose detection and its narrow margin for errors *"it just kept on saying, "You're squatting too deeply. You're squatting too deeply."*.

Pose detection errors often left participants frustrated in cases where the AI either failed to count the correct number of repetitions or the participant felt that the AI inflated the number of repetitions. For example, P10 stated *"I quite liked [Onyx] because of the counting thing, although, as I said earlier, I was a bit annoyed when it didn't count properly."*. Similarly, P5 reported attempting to subvert the AI. This resulted in the AI counting sit-up repetitions when the participant simply *"mov[ed] my legs... They were so bad it's comical"*.

When the computer vision was not able to correctly detect the participant, P5 wanted to be able to ask *"what do I need to do for you to see me? If I could talk to it"*. Likewise, P6 considered AI should communicate factors that may affect its ability to accurately detect the user such as *"adapt in the sense that it might have an opinion about [the space] like, "It's too dark," or, "It's too small," or, "It's too noisy," or, "It's too crowded, I can see so many people here"*. The participants' experiences and insights raise questions about the AI's ability or lack of, to problem-solve about its own detection capabilities. The participants often felt the problem-solving relied too heavily on their own trial-and error instead of getting useful contextual advice from the AI apps.

Spatial limitations: Participants frequently commented that the amount of space required to accommodate the viewing angle of the camera and ensure they fit within the frame for the AI to track their movements, was greater than they had at home. P12 was irritated as they *"had to do [the exercises] in the hallway"* due to lack of larger spaces throughout the rest of their home. Similarly P3 commented *"I was thinking I'm going to have to rearrange my bedroom in order to use this app"*. Participants also reported how the spatial issues seemed more prominent based on the position of the phone (portrait or landscape) and the change in positions between specific exercises (standing or floor exercises). For example, P1 expressed that the pose-detection *"worked well for the standing up ones"*, it was *really annoying for any of the sit-ups, push-ups type thing"*. P3, went on to explain that while working out with Zenia, they found the camera having to be *"landscape and being sideways"*, resulted in the participant being *"quite far away from the phone in order for [the app] to catch your full height"*.

In addition, the need to place the phone at a considerable distance from the user to accommodate the camera tracking, made it difficult to view any visual corrections or easily interpret the real-time metrics. As P10 explains*"the problem with the Zenia one is, the person on the screen was really small, and because you had to be quite far back to be in the screen, you couldn't really see what they were doing"*. P1 reported similar experiences where they are *"six foot away trying to fit on the screen. I don't want to have to go up to the screen and go 'skip' and then go back over and do another exercise."* .

These experiences highlight the tensions that arise when the AI pose detection requires a considerable amount of space to capture the user's pose but also uses visual indicators as a way to deliver feedback to the participant.

While recognising that spatial limitations may be prominent in computer-vision AI apps, AI systems should consider the need for workouts that fit within the confinements of the user's available space as opposed to the user having to move and alter the camera to fit the AI's requirements, which ultimately results in the user prioritising the camera's viewpoint instead of their fitness and well-being.

4.2 Visual feedback

Participants reported on a variety of visual feedback features that assisted them in tracking their workout progress and expressed their views relating to visual corrective feedback, workout metrics and a preferred workout screen mode.

Seeing myself on screen: Participants reported differing preferences for what they saw on screen whilst working out. Some preferred to see the instructor only, whilst others wanted to see themselves on-screen throughout the workout. The third option that participants discussed was predominantly seeing the instructor on-screen with the camera intermittently showing *"something wrong [with the users posture]"* (P12) as this draws the user's attention as opposed to showing the user all the time where they may *"feel like I'm doing something wrong if I saw it all the time."* (P12). Two participants proposed that the exercise type should determine whether visual feedback is continuously presented on the screen with complex exercises, such as Yoga, providing constant visual feedback to the user to guide them through difficult movements, while high intensity workouts require less visual feedback. *"I think it's a good thing but for those slower, more relaxed in a way pose-based workouts. I didn't miss it for Fitness Ally [...] because that was more energetic and let's just go go go, it's things that were much more obvious and simpler"* (P1). P2 suggested that users should be able to configure what could be seen on screen *"The perfect app would probably provide options on your visual feedback [...] so you can just watch yourself, you can use the phone screen like a mirror, or you can watch your instructor do the exercises while you do the exercises as your visual feedback, or you can switch between them during your workout with a vocal cue, say, "Show me skeleton." [...] "Show me silhouette." or "Show me instructor.""*.

Skeleton and body outline: Participants found the visual skeleton overlay useful and were impressed by the system's ability to highlight body joints using different colours when the pose was correct or incorrect (Fig.1; Zenia; Kaia). *"On the screen with [Zenia], it was similar in that, there was me, it was actually filming me, but they have these green or red dots [...] mapping bits of my body, which it seemed to do quite well actually, I was quite impressed with how it did that."* (P10). However, two participants mentioned that highlighting the specific incorrect body part was more useful than the whole skeleton lighting up. The participants also reported that the high contrast body outline (Fig 1; Onyx) that mapped the participants was initially visually appealing however the novelty wore off as they realised that it did not provide corrective feedback. *"It was also a bit gimmicky in a way, because even with the Onyx, it's like they're not highlighting in red or anything on your silhouette where you're doing anything wrong"* (P8).

Real-time glanceable metrics: Overall, participants viewed real-time metrics such as rep counts, time remaining, workout progress, and levels of user movement favourably and often referred to them when tracking their progress during a workout. Importantly, these metrics were seen as *glanceable* in that they could quickly be assessed without requiring extensive comprehension, as P2 states *"[Onyx] presented the time and reps count very clearly... That was pretty easy to keep track of and good to see how long you had left in each workout"*. Participants expressed the need for the screen to be visually simplified *"The screen wasn't cluttered at all. There was just Ally herself, and then intensity and number of reps recorded. It's actually reasonably easy to glance at the number of reps just to make sure if it's counting up* *the reps as I'm doing them"* (P3). This allowed participants to quickly understand the exercise progress and ensure that the system was correctly capturing their movements whilst maintaining their workout flow.

Two visual feedback features were highly praised: the Fitness Ally visual intensity measure displayed on a continuously updating graph and Onyx's uses of numerical metrics representing repetitions per/second. The visual intensity measure (Fig1 (left, top center) provides an animated graph that responds directly to the user's increase or decrease in movement. The visual element also uses colour to signify high intensity in green and low intensity in red to help motivate the user. Indeed, participants resoundingly agreed that the swift responsiveness gave them confidence that they were being tracked and therefore motivated them to maintain their exercise work rate as P10 explains *"Ally has that intensity bar, which I thought worked really well. It seemed to almost always reflect what I felt my actual intensity was, and that was really important to feel like I had that connection with what was going on on the app, it kind of verified my feelings about how much effort I was putting into it. That was good."*. Participants who tried both Fitness Ally and Onyx, reported that the Ally intensity bar that used colour to indicate how the participants were performing was more useful for feedback than Onyx where the repetition count *"just [displayed] numbers like 2.4 and 2.7"* (P2). This lack of design consideration made it difficult for P2 to follow along while the intensity bar was *"easy to say like, "All right, I'm keeping pace." I didn't have to understand the numbers or calculate anything or remember what I'd done before"*.

Post-workout feedback: After completing a workout participants were presented with descriptive details such as; workout duration, repetition counts per exercise, as well as more abstract data such as body pose accuracy, workout scores, achievements, and leader board positions. Five participants highlighted that the scores they were receiving did not make sense and that the apps failed to explain how those scores were calculated, as P2 explains *"I think the kinetics measurement tool was quite cool and definitely made it like an easy way to rate your performance [...] but then also, you don't know really how they're calculated, or what they really actually mean, and how much value I attribute to that other than scoring yourself against other people."*. The participants identified the repetition count and the pose accuracy score to be the most useful metrics to understand their progress and to improve the pose over time. *"I suppose if they could say you're 99% more accurate than you were last time [...] maybe that would be good because if you initially start doing something and you're not doing it very well, but then later you could see an improvement in your accuracy, that would be good."* (P10). The participants expressed that post-workout metrics could become useful features for the AI instructor to assist the user in understanding their progress over time.

Moving beyond the metrics, participants also expected the AI instructor to utilise the video captured during the workout to provide feedback on their form. Whilst some apps did provide a post-workout picture or video clip feature, they were intended more for social sharing than providing detailed information about user's posture corrections. When discussing preferred methods of displaying post-workout video feedback, participants reported that having a side by side picture comparing their posture with the instructor or a short video replay, would be useful to review the posture corrections in more detail, as P3 explains *"[...] if you had the option to have the front-facing camera record your session, and then afterwards, [...] go through and highlight bits for you or like suggest improvements at various parts, then yes, I would definitely use that"*. One participant suggested reviewing pictures over time, in order to track, for example, weight loss: *"Maybe partly almost as a montage that I can look at myself and go, "Yes, I look much better." [...] If I've also been losing weight [...] over a period of a month, that could be really fun."* (P6). Similarly, P4 discussed how the apps lacked information on what body parts have been exercised and expected the AI to assist with ensuring that workouts were appropriately balanced across the body with P4 expressing *"[...] if I was going to use them for a long period of time [...] how would I know that I'm exercising the parts of my body in an equal fashion [...] and not just ending up focusing on the same parts all the time. I feel like having some kind of balance check [...] some sort of bar that moves up and down and* *you could be like, "Oh, I need to do some more stuff on my legs." I feel that's, for me, what was missing"*. These examples demonstrate the desire for an AI instructor to provide meaningful summaries of the data using video and text-based metrics in order to help the user understand the the key moments within a workout and also throughout a wider exercise programme.

4.3 Dialogue with the AI

Participants shared their experiences of receiving corrective audio and verbal feedback including beeps, dings and pre-recorded feedback phrases from the AI Instructor. In addition, participants discussed different dialogue qualities and expressed their opinions about engaging in conversation with the AI.

Corrective feedback: Participants explained the importance of AI audible feedback for pose correction in instances when the user is actively engaged in an exercise and unable to see the device's screen. As P9 explains *"If you're in a pose with your head down, it's really important that the audio is instructing you really clearly what you have to do"*. In addition, many found corrective and confirmatory feedback helpful for keeping their form on track and informing the participant they achieved the requested movement. *"I wasn't squatting as properly and they said, "Oh, you need to make sure you go down." It's a good reminder. Then when you do go further down, then it says, "Oh yes, well done. Now you've gone further down."* (P10).

However, many participants also expressed their frustrations relating to repetitive corrective feedback asking participants to return to a visible position for the pose-detection to work. For example, P1 mentioned how their experience with VAY and Kaia *"was just so much more annoying because it was like, "Please align yourself within the frame." Just over and over again [...]"*. In addition, P2 explains: *"if it gives you the exact same feedback more than three times within a minute, it should know that the feedback's either not working or that's not an exercise that you can perform and take some corrective action. Maybe say, "Why don't we try something else or maybe this isn't working for some reason," trying to address it in some way, whereas I just find it was just throwing the same thing at me again and again. That would definitely annoy me to the point of not wanting to use it."*. Similarly, correcting the user too often, as explained by P8 *"If you're constantly on every jump, correcting something [it's] too much"*, was perceived unfavourably as it interfered with workout flow. The corrective and confirmatory verbal feedback is a useful AI feature giving participants tips on how to improve their form, especially in the situations when the participant's view is obstructed, mid exercise and unable to view the screen. However, at other times the repetitive corrective feedback was frustrating despite participant's efforts to correct their form. To this end, the AI lacked the ability to recognise user's efforts and to adapt to the situation by offering alternative feedback.

Conversational qualities: Participants criticized the feedback as being overly complimentary in such a way that motivational statements felt hollow and lacked purpose. This gave the impression that the AI lacked comprehension of temporality and context resulting in statements being delivered at inappropriate times. P5 explains *"They start a session and they say, "You're killing it," I haven't even started yet I can't be killing it right now. Just those, they have just got these stock phrases that I just found laughable."*. Similarly, P7 explained how verbal encouragements used in the wrong context made them feel demotivated *"[...] when it said, "Well done," when I was standing still [...] I think it's important to track what the person is doing [...] because if I'm getting told, "Well done," for standing still, I will stand still again"*. P5 expressed how the context of where they were in relation to the workout wasn't incorporated in the dialogue: *"All three of them [Zenia, Peleton, VAY] were very bad at telling you where you were in the session [...] "You're halfway there, well done," or something like that. Just the notion of time was missing from all of them."*.

Some participants reported feeling motivated by an upbeat AI personality with P1 describing their experience *"I thought this is an annoying, energetic trainer. Just I wouldn't want to meet this person in person because she would just drain your energy just by talking to you. That's how I thought going into it, but actually, it did work surprisingly well. Just having a real person's voice with that energy in their voice was just motivating in and of itself"*. When the AI provided verbal feedback participants discussed how the style of dialogue was important for motivating them. Participants highlighted the strengths of Fitness Ally and it's design with more playful dialogue components incorporated into the feedback: *"It was something like, "No, that's your right arm, silly. We're doing the left arm." It was just like that, it was much more conversational. [...] It was lighthearted."* (P1). In addition P3 described *"one of the ones that made me laugh with Ally was something to do with her saying that her boyfriend won't be happy with her or something like that which really made me laugh."*.

The apps in our study reportedly lacked temporal and contextual awareness when adapting dialogue to the overall workout structure and user's interactions. Indeed participants desired that a future AI instructor would incorporate this understanding to help keep the user informed and engaged during their workout. Participants also reported the importance of dialogue reflecting aspects of the AI's identity or personality, for example, including playful qualities as found in Ally.

Engaging conversation with the AI: Participants had very low expectations of the AI's ability to comprehend spoken language and therefore felt they would need to modify how they spoke to the AI. *"The issue with it feeling like it's a computer and it doesn't have that capability is that you are aware that you have to ask more closed questions. It's not as a natural way to speak when you're talking to something that you know is just using voice recognition software. [...] I feel like it would introduce an element of potential frustration without much gain"* (P3). Similarly, as P8 explains: *"When it's health [...] if they mis-hear that you've hurt one side or the other, or you've got any issue and they mis-hear that it could lead to an injury or something"*. However participants were willing to use voice commands for less critical operations*"If you're in the heat of the moment and you just want to get on to the next one, you don't really want to stop, walk over, [press the] button, and carry on."* (P7). Participant's expressed concerns about engaging in a more complex conversation with AI due to low expectations of the AI's language comprehension, which could potentially lead to more frustration and harm, rather than benefit. However, participants perceived simple control commands useful for avoiding disruption to the flow of the workout.

4.4 Adapting to the user

Participants often described their expectation of the AI to adapt to their individual needs. For example, long-term goals, injuries they may have, and ways in which they as individuals respond to motivational feedback over time. In this sub-theme, we unpack the role of AI adapting to the user's individuality, and the importance of tailoring workouts and feedback as a way to maintain user interest and motivation.

Adjusting workouts based on the user's individuality: Several participants commented that the apps lacked consideration of user's individual differences, which would be necessary to make effective and tailored workout programmes. One example of AI learning the user's capabilities is known as an 'on-boarding phase'. While a number of the apps had an on-boarding feature, P6 *"felt [the on-boarding was a] little bit sparse"*. Participants expected the on-boarding phase would be an opportunity for the AI instructor to get to know about the participant's background and motivation for working out. For example, P3 expressed concern *"at the moment, Ally doesn't know if I'm going to pick her up once a week or every day. How can there be any long-term planning there?"*. In this regard programs of work do not appear to be constructed with extended use in mind or assessment of the individual differences that must be accounted for to make effective training routines for each user. Likewise, participants expected the AI to be able to adjust the program of work to the user's goals, as P8 explains *"If I just tell the AI, these are my goals. I want to be stronger, I want to be able to hold cardio for 45 minutes or whatever. It can then fill [workouts] to a certain point or be able to build in exercises that allow you to do that"*. These findings illustrate that participants are willing to share long-term goals, current fitness, and many other individual differences to tailor their overall exercise experience instead of the current 'one size fits all' approach.

Several participants highlighted that injuries and disabilities were not accommodated by any of the apps. As P10 explains: *"if you weren't as strong on a certain side for a particular reason and you knew that you wouldn't be able to get stronger on that side, then actually if it learns that, then that would be quite useful"*. As injuries occur, participants expected the AI to remove exercises from their workout or for the AI to adapt to their limitations in real-time. Similarly, participants discussed how prior injuries might impact on their range of movement but had no mechanism to explain this to the AI *"[...] my wrists are not great [...] I had to do a one-arm plank and your wrist is 90 degrees to the ground, that really put loads of strain on there. I guess because I couldn't do the full 30 seconds, but with Kaia, it stops the countdown if you lose the pose, and so you have to get back up, and so it's like, "Back on it." I was like, "This is really hurting, this can't be good for me,""* (P1). Although all of our participants did not live with a disability, their expectation of AI systems to consider not only individual goals but to also take physiological factors such as injuries and disabilities raises concerns about the way AI systems are currently trained. As AI researchers, we should therefore continue to question how we design and train our AI systems to be better accustomed to the individual needs of the user to ensure more inclusive and personalised experiences.

Adjusting workouts over time: Participants expected the AI to act in a similar way to a human instructor in that it can detect their workout performance and dynamically adjust the intensity of the exercise to match the participant's ability over time. *"It would adapt to you and make things more difficult for you automatically without you having to choose because that's what a good trainer does."* (P10). Participants also discussed how the AI should adapt to them in real-time by increasing or decreasing exercise difficulty to maintain flow. As P9 explains *"when you're engaging with it [...] if the app can tell that you're struggling for a pose, they'll give you the easier option. If you choose to ignore it, it will just move on to the next. It doesn't hang around and that you let your heart rate go, keeps you engaged and motivated."*

Motivating users based upon previous experiences with the AI was also a recurring theme as participants expected the AI instructor to adapt the types of motivational statements to improve the user's performance. P3 suggested *"it could learn which things that it says change how motivated you are. If it can sense how motivated I am and it can learn, "When I use these types of prompts for [P3], then he seems to increase his intensity." [...] "He seems to be happier and more motivated when I say these things."*. Participants also wanted the AI to understand their prior performance for use in motivating them during their exercise, as P4 discusses *"it remembers types of exercises well enough that it knows what you were doing wrong or did well last time. It could say things like, "You did this well before. You can do better than that.""*. Similarly, P2 discussed *"if at the start of the exercise it would say something like, "Okay, your personal best for this is however many reps," [...] You got this less last time, let's try and break it, or maybe during the workout if you break, you get a little award, sound or a visual or something to say you broke your personal best."*.

The participants emphasis on AI remembering what the user has done before, further highlights the importance for AI to adapt over time and reflect on what the user has already done. This ongoing process for the AI should therefore build upon the initial on-boarding process to better support the user's on-going goals and abilities.

4.5 Working out with the instructor

Participants spent time discussing their experiences with the workout structure, connecting with the instructor, maintaining flow of workouts and their preferences for configuring and creating their own workouts plans. A personalised workout structure appears to be an important aspect for maintaining motivation for continued use.

Instructor-led experience: The six apps included in the study offered three main class formats to lead the user through an exercise session. First, the 'Follow-along' format led the participant through a continuous video of the same instructor or avatar performing exercises along-side the participant (Peloton, Fitness Ally). Second, the 'Hybrid Follow-along' used a series of short looping video clips for the duration of each exercise, often with different instructors shown in the video (Onyx, Zenia). Third, the 'Demo Video' that began by presenting an instructional video and then switched to display the user's pose captured by the front camera (VAY, Kaia).

Eight participants expressed a preference for the 'Follow-along' class format and often described that the experience felt more personable and akin to being led by an actual instructor. Similarly these participants found this mode to be more motivating as P11 explains" *"with Peloton, you were doing it live with the person, so that was really nice and it felt like it was quite personable."*, P1 added that with Fitness Ally *"there's something about following someone doing it. It was just much more motivational than seeing myself and trying to do what they had shown me"*. The same held true for engaging with Fitness Ally's avatar *"having the character working out along with you was quite good, and setting the pace for how you should work out"* (P2). Having an embodied and identifiable instructor to follow, allowed participants to understand exercise pacing and provided a constant visual demonstration of the exercise to learn from, unlike the changing instructor in the 'hybrid follow-along' and 'demo video' approaches.

All six participants who were given Fitness Ally were initially sceptical of the virtual avatar and were quick to dismiss it as a gimmick. However as P2 describes *"Fitness Ally kind of surprised me quite a lot, because from the initial trailers [I thought] it was going to be a bit crap, to be honest, because it had the cartoon character as your trainer and stuff. [...] I actually enjoyed that a lot more than I thought I would [...]. I find that having the character trainer works quite well"*. Similarly P3 stated *"it felt a bit like, even with Ally, who was not a human, I felt like I had more of a connection with her because she was guiding me through that session"*.

Several participants found the 'Hybrid Follow-along' class format (Onyx and Zenia) that used a set of short clips voiced over by a seemingly detached instructor instructor felt less personal, disjointed, and less engaging. *"[Zenia] seemed to be like the video was cut, like suddenly, the person would be standing up again, and you're thinking, "How are you standing up because you were just down there." Sometimes the person would change as well"* (P10). Being able to view the transitional states is especially important in yoga practices that often use continuous movements to transition from pose to pose. Similarly, P9 reported *"it didn't [...] leave any impact on me. It's not a consistent - unless you have the same people or the same interaction with the same thing, how are they even going to be able to pretend they know how well you're doing?"* (P9). In this 'Hybrid Follow-along' mode the sense of instructor identity was vastly eroded and felt less personal and less motivating. As P3 explains *"with Onyx it was just like a variety of random people with video clips of them do an exercise with the same voice over for all of it. It didn't engage me as much and make me want to keep coming back to it."* (P3).

In the case of the 'Demo Video' participants spoke about the lack of continuity and flow of the workout. This arises through the stop and start nature of watching the instructional video and then watching themselves copy the exercise movement. For example, several participants discussed the freeze-frame feature in Kaia that paused during the workout to show the user where they went wrong with their pose. *"It shows you yourself for 30 seconds. Shows you how it's* *supposed to be done for another 20 seconds [...] and then you go back to it. An exercise that was only meant to be about a minute [...] ended up being three or four minutes"* (P1). Another commented *"when you're in the flow of wanting to do a routine or an exercise, you want to get it done, dusted, over and done with kind of thing. You don't want to be faffing about after everything. "Right, okay, you didn't do this right,""* (P12). This sentiment provides insight into why the 'Follow Along' class format was preferred by the participants in that a real-life instructor might provide incremental feedback on your form throughout the workout or extensive feedback after a class without detracting from the overall flow of the class. P2 explains that *"[Kaia] pausing to correct you, I think took away from being in the moment with the exercise"*. Similarly both Vay and Kaia would turn off the front-facing camera and replay the instructor video if the user's pose was deemed to be incorrect which led to further frustrations with the workout flow. In this mode participants were also unable to use the instructor's body pose for reference whilst viewing themselves on screen.

Workout configuration: Participants expressed the importance of configuring their workout by duration, intensity level, workout types, exclusion criteria, and music during the workout. Many participants commented on the importance of music during workouts and having an ability to play their own music. *"Ally and Onyx, they both had their own soundtracks. Which was better but it would still be nice to have some form of function where you can maybe put your own onto it."* (P8), or having the AI to choose music to match the exercise tempo.

Two participant's wanted to be able to remove disliked exercises and *"rather than waiting for it to turn up in another activity, you could already know filter it off to be less of that. Just having a few more of these bits."* (P8). This requires a careful trade-off between personal choice and an instructor selecting exercises that emphasized the participant's weaker, and perhaps disliked, exercises. How the AI instructors selected workouts for participants varied across the apps in the study. Apps such as Onyx, Zenia, Peloton, and Fitness Ally provided workout classes that could be configured or filtered out, whereas Vay, and Kaia simply prescribed the next exercise. Participants also reported that the selection process was limited to configuring the immediate workout or overly prescriptive in what was available (P1) *"One end of the spectrum is Fitness Ally. It works like just choose the workout. Just, "What do you want to do today?" The other end of the spectrum is Kaia saying, "This is the one workout that we're providing for you, so do it." It'd be nice for somewhere maybe in the middle of say like, "I'm now getting into cycling so I want to do much more leg stuff," so recommend to me [...] leg workouts a bit more."*. The role of an an AI instructor was seen as being both an intelligent assistant that guides the user in creating individual workouts to the participant's preferences and also as the instructor that leads the overall workout program. Whilst the majority of participants trusted and expected the AI to build workouts, some participants expressed a need to have an option to build their own program of workouts. *"I'd like to have been able to create my own routine [...] that was specific for you."* (P12). P2 expressed the need to be in control of the workout program by modifying the AI-proposed workouts to suit their needs: *"It'd be nice to be able to build it yourself, but maybe it would say, "Based on your recent activity, or your lack of activity in a certain area, or what you're doing well or badly, we've come up with something to work well with what you're doing." [...] giving you the option to switch things out, or double up on certain things"*. It was clear that user preferences were not considered in the long-term, wider workout programme that would extend over weeks or months.

4.6 Summary

Our findings indicate that while initial experiences of AI fitness apps feel novel and exciting, the novelty eventually wears off due to missing features such as a lack of useful visualisations for metrics, or frustrations with inconsistent experiences caused by unreliability of current implementations of computer vision models. The findings also indicate that participants want AI systems to have more-in depth knowledge collection and curation about the individual user, which raises important considerations to how AI researchers train and create models that can adapt to the user's needs over time. Finally, our findings reflect on the important considerations that must be considered when designers and researchers design for an AI's 'identity' whether this is in the formality of speech, or the way it visually displays the AI character, for example through an avatar body. We consider these findings in our discussion and suggest a set of key design considerations for researchers to consider for future AI instructor implementations.

5 DISCUSSION

The goal of this research is to understand people's experiences of AI computer vision fitness instructor apps. We attempt to understand how existing technologies compare and contrast with our participant's expectations of these novel technologies. Based upon our findings, below we present five design considerations for designers in this space that relate to three key areas that were prevalent in our findings: *feedback and motivation*, *personalising the experience*, and *building a relationship with the AI*. Our design considerations extend beyond existing research [4, 11] and focus specifically on what participants expect and desire from an AI instructor experience.

5.1 Design Considerations

5.1.1 Feedback and Motivation: Real-time verbal and visual feedback were the key features differentiating the AI from the traditional non-AI fitness apps. Overall, the correct AI feedback was perceived favourably as it motivated, demonstrated the system's responsiveness to the user and provided a foundation to build trust between the user and the AI. More specifically, participants reported being highly motivated by a repetition counter, corrective visual overlays (Zenia and Kaia) and visual intensity measures (Ally and Onyx), as providing real-time feedback on participants' performance. By demonstrating it's responsive comprehension of the user's actions, the user could learn about AI system's abilities and start building trust. When designing with AI, ensuring that the user understands the AI system's capabilities is an important consideration [4]. Participants discussed the value of glanceable and easily understood real-time visual metrics to prevent distracting the user and their sense of flow during workout. For example, using a colour-coded intensity measure, instead of numerical scale that is difficult to interpret. Participants also desired the corrective visual overlay to explicitly show which body-part needs to change and how, rather than simply showing a complete skeleton overlay. Participants expressed how the post-workout metrics such as repetitions and accuracy scores, were the most useful metrics to monitor their progress. However, participants were unsure how the AI system calculated the accuracy scores and wished for more clarity. In addition, the users expect the AI to help them conclude why they have improved or declined over time. However, the expectation from participants for the AI to interpreting their progress should be approached with precaution, as evidence suggest that the user should be the one evaluating their progress to prevent adverse effects on user's motivation to use fitness technology [76].

Regarding real-time verbal feedback, participants found AI corrective verbal feedback useful, however there was a desire for the AI instructor to provide confirmation that they had achieved the desired body pose. Zenia provided a good example of where audible confirmational 'dings' and visual skeleton overlays ensured that the user could easily understand what was expected of them. Literature suggest that using positive feedback in combination with corrective feedback is more effective than corrective feedback alone [85]. Additionally, participants expressed the need for the AI to learn about the user and adapt its dialogue in relation to previous post-workout metrics, such as motivating the user by informing they had achieved new personal best. Participants' views regarding verbal, temporal and motivational verbal feedback was somewhat less positive, highlighting issues with the AI's inability to comprehend the context the participant was in. For example, participants highlighted the lack of temporality in the narrative dialogue, failing to inform the user where in a workout they were or how long they had left. In addition, participants reported how motivational feedback was given in inappropriate social contexts, as participants were praised during temporally inappropriate situations, such as when they were not moving. We refer to the social context in a similar way to [4] in that the AI should "match social norms" to ensure that the user feels that encouragement from the AI is warranted and appropriate. Designers should ensure that motivational and corrective dialogue is temporally appropriate and situated in the context of the user's efforts and progress but also account for previous experiences. In this respect, encouragement from the AI should be saved for the times when praise is warranted, and understand the user's current state.

Participants also reported their frustrations with a repetitive corrective feedback when the AI failed to detect the user or to acknowledge user's attempt to correct their form, due to computer vision and spatial limitations. Participants were frustrated by the lack of clarity regarding why the AI could not see them. Having to continuously adjust the camera to be in view or to change the surroundings through trial and error during a workout was distracting for the user. When detection issues occur the AI should be able to explain to the user why it is struggling to detect the user and what conditions in the environment are effecting the detection. The user should be made aware of the limitations of the computer vision algorithm prior to commencing the workout, rather than continuing in poor conditions. For example, the AI should tell the user "the lighting level is too low" or "the complexity of the background will make detection difficult, there's too many edges" so that the user can correct the error prior to launching into an exercise. Similarly during a workout it's important for the AI to respond to questions from the user about why it is struggling to "see" the user and respond with suggestions on how to improve the conditions. When designing with AI, "making clear what the system can do" and "how well the system can do what it can do" [4] is important for building trust with the user [11]. Those technologies that were less intuitive, eroded trust in the system's overall capabilities and were linked with increased likelihood of abandonment [77]. In addition, when talking about repetitive feedback, participants felt that if AI had to correct the user again, the AI should be able to address it in a different way, by providing alternative options or phrases.

Fitness technology that uses automated activity tracking has positive effect on user's engagement and continuous use [49]. The AI apps capable of monitoring and providing visual and verbal feedback, provides opportunity for the user to reflect and act on the suggestion. When the feedback is provided appropriately, it can promote user's sense of competence and autonomy, which are the key components to increase user's motivation and well-being. However the limitations of AI summarised above, risk compromising user's engagement with the AI fitness apps and their motivation for a continues use. Perceiving the world as a human means that the AI needs to understand both the physical and "situational" context of the user, ensuring that the AI is leveraging this understanding to show "contextually relevant information" [4]. Moving forward, the designers should **incorporate both temporally- and socially-relevant context of the user when engaging in dialogue with an AI fitness instructor. Similarly, the AI should be aware of *what* the user is doing and *how* to select appropriate audible or visual feedback for the user's current state**. When participants are furled up in a pose with their face down to the floor, the AI should understand that audible feedback is more appropriate than visual. In addition, to promote more trust with an AI, the designers should **develop a system that has capabilities to recognise poor conditions for optimal computer vision tracking and have ability to appropriately communicate this information to the user**.

5.1.2 Personalising the Experience: As mentioned in our findings, several of our participants noted a lack of features to tailor their 'personalised workout plans' [72]. Instead, only a limited set of questions were considered, such as 'is your goal to lose weight' or 'Better body tone'. The lack of customizability to tailor workouts to the users' goals calls attention to the AI apps following a one-solution-fits-all approach that remains a significant challenge in non-AI fitness apps [86]. Consolvo et al's. [22] work on technology that encourages physical activity, offered a series of guidelines to provide awareness to a user's activity level, social features, and awareness of the user's lifestyle. Similarly, our analysis describes how participants wanted an in-depth onboarding phase by considering a user's fitness state, motivation, time commitments, personal history [50], fitness routine and the type of exercises they like and dislike. This echoes a fitness instructor's [14] process where future implementations of AI fitness apps should take advantage of its pose-detection abilities to introduce users to a pre-assessment [71] feature to provide the system with insights into the user's current fitness alongside their goals and desires. Furthermore, our findings indicate that AI's current implementations could not adapt in real-time, such as detecting when the user struggles to perform specific exercises or are unwilling to do a movement they did not like. Users expected collaboration with their AI instructor where the routine would adapt based on the user's on-going relationship with the AI [11]. Therefore, this would help build a personalised dataset regarding what type of dialogue motivates the user, what exercises they struggling with, and how 'the AI' can develop a more tailored fitness program for the individual [4].

During our study, participants acknowledged the lack of consideration for when the user has an injury. In the apps current state, they offer no injury alternatives. Still, participants suggested 1) the AI instructor could give alternative exercises that focus on non-injured parts of the body or 2) the pose detection to ignore or discount the injured joint while they workout. In a similar vein, pose detection relying on estimating each joint of the human pose highlights concerns of how pose-detection features may currently contribute to the exclusion of particular groups such as those who are disabled [24]. With this in mind, **we should consider how pose-detection models may take a calibration phase where the system can determine the type of human pose model that would be appropriate for the individual or to be easily configured to which joints are required by the user**. Future work in this area may seek to design accessible open-source AI services that the user may overlay or integrate into their fitness apps or routines. Finally, creating open-access AI tools may re-frame what AI and fitness can be instead of the current commercial take of AI replacing the fitness instructor.

5.1.3 Building a Relationship with the AI:. With each of the AI fitness apps in our study we found that the experience of engaging with the instructor was limited to the immediate interactions contained within a single workout instance rather than as a continued and longer-term companionship that one might develop with a human instructor over time. This resulted in an ephemeral AI instructor relationship and lacked any memory of prior experiences with the user that could be leveraged to facilitate the more motivational and personalised experiences that we describe in our findings. However, participants expected greater guidance prior, during, and after engaging with the AI instructor to understand their long-term goals, frequency of use, and personal preferences for engaging in a wider fitness program. Embedded in this guidance would be the preferences of the user that the AI instructor could take into account when making these decisions. Importantly existing research has shown [4, 53] that improving AI intelligibility [11] and making clear why these decisions have been made and within what context are essential to building trust in the AI and sustaining an on-going relationship between user and AI instructor.

During workouts participants strongly preferred the follow-along class format that used a consistent embodied AI instructor throughout the workout. Participants felt a more meaningful connection with the instructor (even in the case of Fitness Ally's avatar) and felt more motivated by having an instructor performing a workout with them. This shared exertion through the workout experience helped to reinforce that the AI was a part of the workout experience with the user. Having a constant visual embodiment of the AI instructor provided a similar experience of a human instructor physically demonstrating errors in pose using their own body. Participants could reference the instructor's pose to understand exercise pacing and form. Capturing the user's form using computer vision provided an opportunity to use this visual information post-workout. Participants wanted the AI to provide video summaries of their own mistakes so that they could compare and contrast their performance with the instructor's form as well as receive feedback on how to improve. We therefore propose that designers should **develop a sense of embodied and consistent identity that provides a visual reference to the user to help motivate them through follow-along exercises classes**.

Each AI instructor experience was also focused on relaying prior descriptive post-workout data rather than leveraging the AI's unique capability of data-driven predictive understanding of user performance to portray what the user might achieve in future with sustained use. Similarly, participants expected the AI to help them draw conclusions as to why these predictions or past-reflections might have occurred. In response to this we recommend that **designers of AI should provide more long-term human-like instruction that leverages continued understanding of the user to create a more personalised workout experience over time**.

Importantly, participants desired a more personalised and lasting relationship with the AI instructor that wasn't currently offered by existing technologies. However, the extent of this relationship is still unclear and researchers need to explore how it mirrors the AI instructor's human counterpart as this technology permeates into everyday life.

6 LIMITATIONS

Our paper provides preliminary insights into the experiences of using newly emerging AI fitness instructor technologies. Although participants were provided with three apps over a short eight-day period, we aimed to to compare and contrast their experiences of using these technologies. Therefore, our findings are a first look at the expectations and experiences of using these technologies and serve as initial guidelines for this emerging design space. As this area matures, future research should continue to explore the longitudinal experience of engaging with an AI fitness instructor over time.

7 CONCLUSION

This paper identifies the emerging trend of *on-device* computer vision AI fitness instructor apps and begins to explore people's experiences of using these technologies. We present findings from a user study with twelve participants who used five AI fitness instructor apps and identify five themes from our thematic analysis: limitations of computer vision, visual feedback, dialogue with the AI, and workout with the instructor. We draw upon our findings to present a series of design recommendations for designing AI fitness instructor apps and frame our discussion around three areas: *'Feedback and Motivation', 'Personalising the Experience', and 'Building a relationship with the AI'* to reflect on the opportunities for designers of these systems. In addition we contribute knowledge of user experiences and expectations about AI corrective visual and verbal feedback to the emerging domain. Reflecting on our findings and discussion it is clear that AI fitness instructor technologies are capable of providing an engaging and meaningful user experience however the existing implementations of these newly emerging technologies fall short of truly fulfilling the promises of next generation AI. We present this paper to the HCI community and hope to inspire the design of AI fitness instructor experiences in the future.

REFERENCES

[1] VAY AG. 2020. *VAY Fitness Coach*. VAY AG. Retrieved April 23, 2021 from https://vay.ai/

[2] Ahed Aladwan, Ryan M. Kelly, Steven Baker, and Eduardo Velloso. 2019. A Tale of Two Perspectives: A Conceptual Framework of User Expectations and Experiences of Instructional Fitness Apps. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press, New York, New York, USA, 1–15. https://doi.org/10.1145/3290605.3300624

[3] Ryan M. Alturki. 2016. A Systematic Review on What Features Should Be Supported By Fitness Apps and Wearables To Help Users Overcome Obesity. , 197–206 pages. https://doi.org/10.15623/ijret.2016.0509032

[4] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300233

[5] Apple. 2020. *Apple CoreML Machine Learning Toolkit*. Apple. Retrieved Sept 10, 2020 from https://developer.apple.com/machine-learning/

[6] Victoria Bellotti and Keith Edwards. 2001. Intelligibility and accountability: Human considerations in context-aware systems. *Human-Computer Interaction* 16, 2-4 (2001), 193–212. https://doi.org/10.1207/S15327051HCI16234_05

[7] Claudio Bettini, Oliver Brdiczka, Karen Henricksen, Jadwiga Indulska, Daniela Nicklas, Anand Ranganathan, and Daniele Riboni. 2010. A Survey of Context Modelling and Reasoning Techniques. *Pervasive Mob. Comput.* 6, 2 (April 2010), 161–180. https://doi.org/10.1016/j.pmcj.2009.06.002

[8] Claudio Bettini, Oliver Brdiczka, Karen Henricksen, Jadwiga Indulska, Daniela Nicklas, Anand Ranganathan, and Daniele Riboni. 2010. A survey of context modelling and reasoning techniques. *Pervasive and Mobile Computing* 6, 2 (2010), 161–180.

[9] Dominik Bial, Thorsten Appelmann, Enrico Rukzio, and Albrecht Schmidt. 2012. Improving cyclists training with tactile feedback on feet. In *International Conference on Haptic and Audio Interaction Design*. Springer, Berlin, Heidelberg, 41–50.

[10] Susanne Bødker. 2015. Third-wave HCI, 10 years later - Participation and sharing. *Interactions* 22, 5 (2015), 24–31. https://doi.org/10.1145/2804405

[11] Oliver Brdiczka. 2019. *Contextual AI: The Next Frontier of Artificial Intelligence*. Adobe. Retrieved Sept 10, 2020 from https://blog.adobe.com/en/2019/04/09/contextual-ai-the-next-frontier-of-artificial-intelligence.html

[12] Oliver Brdiczka. 2019. *Georgia Tech Talk: Contextual AI: The Next Frontier Towards Human-Centric Artificial Intelligence*. Adobe. Retrieved Sept 10, 2020 from https://smartech.gatech.edu/handle/1853/60924

[13] Yujun Cai, Liuhao Ge, Jun Liu, Jianfei Cai, Tat-Jen Cham, Junsong Yuan, and Nadia Magnenat Thalmann. 2019. Exploiting spatial-temporal relationships for 3d pose estimation via graph convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision*. IEEE, Seoul, Korea, 2272–2281.

[14] Francisco Campos, Vera Simões, and Susana Franco. 2015. Characterization and comparison of the participant's perception about the quality of the fitness group exercise instructor, considering the practiced activity. *Arena-Journal of Physical Activities* 4 (2015), 85–104.

[15] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2017. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE, Hawaii, USA, 7291–7299.

[16] Tara Capel, Johanna Frederike Schnittert, Stephen Snow, and Dhaval Vyas. 2015. Exploring Motivations of Young Adults to Participate in Physical Activities. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI EA '15)*. Association for Computing Machinery, New York, NY, USA, 1409–1414. https://doi.org/10.1145/2702613.2732800

[17] Joao Carreira and Andrew Zisserman. 2017. Quo vadis, action recognition? a new model and the kinetics dataset. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Hawaii, USA, 6299–6308.

[18] Yilun Chen, Zhicheng Wang, Yuxiang Peng, Zhiqiang Zhang, Gang Yu, and Jian Sun. 2018. Cascaded pyramid network for multi-person pose estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE, Utah, United States, 7103–7112.

[19] Yu Cheng, Bo Yang, Bo Wang, Wending Yan, and Robby T Tan. 2019. Occlusion-aware networks for 3d human pose estimation in video. In *Proceedings of the IEEE International Conference on Computer Vision*. IEEE, California, USA, 723–732.

[20] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, Vincent Wade, and Benjamin R. Cowan. 2019. What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300705

[21] Victoria Clarke, Virginia Braun, and Nikki Hayfield. 2015. Thematic analysis. *Qualitative psychology: A practical guide to research methods* (2015), 222–248.

[22] Sunny Consolvo, Katherine Everitt, Ian Smith, and James A Landay. 2006. Design requirements for technologies that encourage physical activity. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. Association for Computing Machinery, Montréal, Canada, 457–466.

[23] Jess Cording. 2020. *Covid-19-transforming-fitness-industry*. Forbes. Retrieved Sept 17,2020 from https://www.forbes.com/sites/jesscording/2020/07/13/covid-19-transforming-fitness-industry/

[24] Rúbia Eliza de Oliveira Schultz Ascari, Luciano Silva, and Roberto Pereira. 2021. Computer vision applied to improve interaction and communication of people with motor disabilities: A systematic mapping. *Technology and Disability* 33, 1 (2021), 1–18.

[25] Anind K. Dey, Gregory D. Abowd, and Daniel Salber. 2001. A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human-Computer Interaction* 16, 2-4 (2001), 97–166. https://doi.org/10.1207/S15327051HCI16234_02

[26] Rowan Diamond and Erin Byrd. 2020. Standing up for health–improving mental wellbeing during COVID-19 isolation by reducing sedentary behaviour. *Journal of affective disorders* 277 (2020), 232–234.

[27] Moritz Einfalt and Rainer Lienhart. 2020. Decoupling video and human motion: towards practical event detection in athlete recordings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, Virtual, 892–893.

[28] Moritz Einfalt, Dan Zecha, and Rainer Lienhart. 2018. Activity-conditioned continuous human pose estimation for performance analysis of athletes using the example of swimming. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, Utah, United States, 446–455.

[29] Kat Eschner. 2020. *COVID-19 has changed how people exercise, but that doesn't mean gyms are going away.* Fortune. Retrieved Sept 17, 2020 from https://fortune.com/2020/06/11/coronavirus-gyms-workouts-fitness-apps-reopening/

[30] Chloe Fan, Jodi Forlizzi, and Anind Dey. 2012. Considerations for Technology That Support Physical Activity by Older Adults. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility* (Boulder, Colorado, USA) *(ASSETS '12)*. Association for Computing Machinery, New York, NY, USA, 33–40. https://doi.org/10.1145/2384916.2384923

[31] Forbes. 2020. *New AI Helps People Exercise At Home During Pandemic.* Forbes. Retrieved Sept 16, 2020 from https://www.forbes.com/sites/kevinanderton/2020/05/15/new-ai-helps-people-exercise-at-home-during-pandemic-infographic

[32] Thomas Fritz, Elaine M. Huang, Gail C. Murphy, and Thomas Zimmermann. 2014. Persuasive Technology in the Real World: A Study of Long-Term Use of Activity Sensing Devices for Fitness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 487–496. https://doi.org/10.1145/2556288.2557383

[33] Andrew Garbett, David Chatting, Gerard Wilkinson, Clement Lee, and Ahmed Kharrufa. 2018. *ThinkActive: Designing for Pseudonymous Activity Tracking in the Classroom.* Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173581

[34] Kathrin Gerling, Mo Ray, Vero Vanden Abeele, and Adam B. Evans. 2020. Critical Reflections on Technology to Support Physical Activity among Older Adults: An Exploration of Leading HCI Venues. *ACM Trans. Access. Comput.* 13, 1, Article 1 (April 2020), 23 pages. https://doi.org/10.1145/3374660

[35] Google. 2020. *Google ML Kit.* Google. Retrieved Sept 10, 2020 from https://developers.google.com/ml-kit

[36] Google. 2020. *Google: People + AI Research (PAIR) AI Design Guidelines.* Google. Retrieved Sept 10, 2020 from https://pair.withgoogle.com/

[37] Jonna Häkkilä, Ashley Colley, Virve Inget, Mira Alhonsuo, and Juho Rantakari. 2015. Exploring digital service concepts for healthy lifestyles. In *Design, user experience, and usability: Design discourse.* Springer, Switzerland, 470–480.

[38] Jonna Häkkilä and Jani Mäntyjärvi. 2006. Developing design guidelines for context-aware mobile applications. In *Proceedings of the 3rd international conference on Mobile technology, applications & systems.* Association for Computing Machinery, Bangkok, Thailand, 24–es.

[39] Christina N. Harrington, Lauren Wilcox, Kay Connelly, Wendy Rogers, and Jon Sanford. 2018. Designing Health and Fitness Apps with Older Adults: Examining the Value of Experience-Based Co-Design. In *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare* (New York, NY, USA) *(PervasiveHealth '18)*. Association for Computing Machinery, New York, NY, USA, 15–24. https://doi.org/10.1145/3240925.3240929

[40] M Hassenzahl. 2013. *User Experience and Experience Design - SAP User Experience Community.* SAP. http://experience.sap.com/basics/post-62/

[41] Kaia Health. 2020. *Kaia Personal Trainer.* Kaia Health. Retrieved April 23, 2021 from https://www.kaiahealth.com/

[42] Lynn Katherine Herrmann and Jinsook Kim. 2017. The fitness of apps: a theory-based examination of mobile fitness app usage over 5 months. *mHealth* 3, December 2016 (2017), 2–2. https://doi.org/10.21037/mhealth.2017.01.03

[43] Hayley Holmes, Jamie Wood, Sue Jenkins, Peta Winship, Dianne Lunt, Susan Bostock, and Kylie Hill. 2013. Xbox Kinect represents high intensity exercise for adults with cystic fibrosis. *Journal of Cystic Fibrosis* 12, 6 (2013), 604–608. https://doi.org/10.1016/j.jcf.2013.05.005

[44] IBM. 2019. *IBM Design for AI.* IBM. Retrieved Sept 15, 2020 from https://www.ibm.com/design/ai/

[45] Onyx Inc. 2020. *Onyx: Home Workout.* Onyx Inc. Retrieved April 23, 2021 from https://www.onyx.fit/

[46] Zenia Inc. 2020. *Zenia App: Yoga for Beginners.* Zenia Inc. Retrieved Sept 2, 2020 from https://zenia.app/

[47] Peloton Interactive. 2020. *Peloton: At Home Fitness.* Peloton Interactive. Retrieved April 23, 2021 from https://www.onepeloton.com/app

[48] Vrushabh Jadhav, Gaurav Jain, Onkar Khandekar, and Bansod Smita. 2018. Virtual Fitness Trainer with Spontaneous Feedback using a line of motion sensing input device Kinect Xbox 360. *International Research Journal of Engineering and Technology* 5, 10 (2018), 1627–1630.

[49] Iman Khaghani-Far, Svetlana Nikitina, Marcos Baez, Ekaterina A Taran, and Fabio Casati. 2016. Fitness applications for home-based training. *IEEE Pervasive Computing* 15, 4 (2016), 56–65.

[50] Amanda Lazar, Christian Koehler, Theresa Jean Tanenbaum, and David H. Nguyen. 2015. Why We Use and Abandon Smart Devices. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 635–646. https://doi.org/10.1145/2750858.2804288

[51] Valentina F. Leuciuc. 2015. Determining the fitness level of the students through the physical activity index questionnaire. *The Annals of "Dunarea de Jos" University of Galati. Fascicle XV, Physical Education and Sport Management* 2 (2015), 24–26.

[52] Qingyang Li, Clara Caldeira, Daniel A. Epstein, and Yunan Chen. 2020. Supporting Caring among Intergenerational Family Members through Family Fitness Tracking. In *Proceedings of the 14th EAI International Conference on Pervasive Computing Technologies for Healthcare* (Atlanta, GA, USA) *(PervasiveHealth '20)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3421937.3422018

[53] Q. Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: Informing Design Practices for Explainable AI User Experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376590

[54] Rainer Lienhart, Moritz Einfalt, and Dan Zecha. 2018. Mining automatically estimated poses from video recordings of top athletes. *International Journal of Computer Science in Sport* 17, 2 (2018), 94–112.

[55] Jun Liu, Gang Wang, Ping Hu, Ling-Yu Duan, and Alex C Kot. 2017. Global context-aware attention lstm networks for 3d action recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1647–1656.

[56] Nintendo Co. Ltd. 2006. *Wii Sports*. Retrieved Feb 11, 2021 from https://nintendo.com

[57] Nintendo Co. Ltd. 2007. *Wii Fit*. Retrieved Feb 11, 2021 from https://nintendo.com

[58] Meethu Malu and Leah Findlater. 2016. Toward accessible health and fitness tracking for people with mobility impairments. In *Proceedings of the 10th EAI International Conference on Pervasive Computing Technologies for Healthcare*. 170–177.

[59] Christine F Martindale, Markus Wirth, Stefan Schneegass, Markus Zrenner, Benjamin H Groh, Peter Blank, Dominik Schuldhaus, Thomas Kautz, and Bjoern M Eskofier. 2016. Workshop on wearables for sports. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. 877–880.

[60] Eleonora Mencarini, Amon Rapp, Lia Tirabeni, and Massimo Zancanaro. 2019. Designing wearable systems for sports: a review of trends and opportunities in human–computer interaction. *IEEE Transactions on Human-Machine Systems* 49, 4 (2019), 314–325.

[61] Florian Michahelles and Bernt Schiele. 2005. Sensing and monitoring professional skiers. *IEEE Pervasive Computing* 4, 3 (2005), 40–45.

[62] Anouk Middelweerd, Julia S Mollee, C Natalie van der Wal, Johannes Brug, and Saskia J Te Velde. 2014. Apps to promote physical activity among adults: a review and content analysis. *International journal of behavioral nutrition and physical activity* 11, 1 (2014), 1–9.

[63] Maria D. Molina and Jessica G. Myrick. 2020. The 'how' and 'why' of fitness app use: investigating user motivations to gain insights into the nexus of technology and fitness. *Sport in Society* (2020). https://doi.org/10.1080/17430437.2020.1744570

[64] Florian Floyd Mueller, Joe Marshall, Rohit Ashok Khot, Stina Nylander, and Jakob Tholander. 2015. Understanding Sports-HCI by going Jogging at CHI. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. 869–872.

[65] Adrià Muntaner-Mas, Antonio Martinez-Nicolas, Carl J. Lavie, Steven N. Blair, Robert Ross, Ross Arena, and Francisco B. Ortega. 2019. A Systematic Review of Fitness Apps and Their Potential Clinical and Sports Utility for Objective and Remote Assessment of Cardiorespiratory Fitness. *Sports Medicine* 49, 4 (2019), 587–600. https://doi.org/10.1007/s40279-019-01084-y

[66] Twenty Billion Neurons. 2020. *Fitness Ally: AI Workouts Coach*. Retrieved April 23, 2021 from https://fitnessallyapp.com/

[67] Joseph W Newbold, Nadia Bianchi-Berthouze, and Nicolas E Gold. 2017. Musical expectancy in squat sonification for people who struggle with physical activity. Georgia Institute of Technology.

[68] Sharmilee M. Nyenhuis, Justin Greiwe, Joanna S. Zeiger, Anil Nanda, and Andrew Cooke. 2020. Exercise and Fitness in the Age of Social Distancing During the COVID-19 Pandemic. *Journal of Allergy and Clinical Immunology: In Practice* 8, 7 (2020), 2152–2155. https://doi.org/10.1016/j.jaip.2020.04.039

[69] Mengyue Pan, Sagar Salvi, and Erin Brady. 2018. Designing auditory feedback from wearable weightlifting devices. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–6.

[70] Tomas Pfister, James Charles, and Andrew Zisserman. 2015. Flowing convnets for human pose estimation in videos. In *Proceedings of the IEEE International Conference on Computer Vision*. 1913–1921.

[71] Sharon R Phillips, Risto Marttinen, and Kevin Mercier. 2017. Fitness assessment: Recommendations for an enjoyable student experience. *Strategies* 30, 5 (2017), 19–24.

[72] Matthew Rand, Paul Norman, and Elizabeth Goyder. 2020. A systematic review of interventions to increase attendance at health and fitness venues: identifying key behaviour change techniques. *BMC Public Health* 20, 1 (2020), 1–13.

[73] Kyle Rector, Cynthia L Bennett, and Julie A Kientz. 2013. Eyes-free yoga: an exergame using depth cameras for blind & low vision exercise. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–8.

[74] Kyle Rector, Lauren Milne, Richard E. Ladner, Batya Friedman, and Julie A. Kientz. 2015. Exploring the Opportunities and Challenges with Exercise Technologies for People Who Are Blind or Low-Vision. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility* (Lisbon, Portugal) *(ASSETS '15)*. Association for Computing Machinery, New York, NY, USA, 203–214. https://doi.org/10.1145/2700648.2809846

[75] Microsoft Research. 2019. *Guidelines for Human-AI Interaction*. Retrieved Sept 15, 2020 from https://www.microsoft.com/en-us/research/project/guidelines-for-human-ai-interaction/

[76] Robert Rockmann. 2019. Don't Hurt Me. . . No More? An Empirical Study on the Positive and Adverse Motivational Effects in Fitness Apps. (2019).

[77] Michael A Rupp, Jessica R Michaelis, Daniel S McConnell, and Janan A Smither. 2018. The role of individual differences on perceptions of wearable fitness device trust, usability, and motivational impact. *Applied ergonomics* 70 (2018), 77–87.

[78] John Shepherd. 2020. Recovery: can subjective monitoring be more effective than objective methods? Retrieved August 17, 2020 from https://www.peakendurancesport.com/endurance-training/recovery-strategies/recovery-can-subjective-monitoring-effective-objective-methods/

[79] Aneesha Singh, Stefano Piana, Davide Pollarolo, Gualtiero Volpe, Giovanna Varni, Ana Tajadura-Jimenez, Amanda CdeC Williams, Antonio Camurri, and Nadia Bianchi-Berthouze. 2016. Go-with-the-flow: tracking, analysis and sonification of movement and breathing to build confidence in activity despite chronic pain. *Human–Computer Interaction* 31, 3-4 (2016), 335–383.

[80] Eline Suzanne Smit and Nadine Bol. 2020. From self-reliers to expert-dependents: identifying classes based on health-related need for autonomy and need for external control among mobile users. *Media Psychology* 23, 3 (May 2020), 391–414. https://doi.org/10.1080/15213269.2019.1604235

[81] Katta Spiel, Fares Kayali, Louise Horvath, Michael Penkler, Sabine Harrer, Miguel Sicart, and Jessica Hammer. 2018. Fitter, happier, more productive? The normative ontology of fitness trackers. In *Extended abstracts of the 2018 CHI conference on human factors in computing systems*. 1–10.

[82] Raluca-Alexandra Stoica. 2018. "Keep Going!" Understanding the Implications of Coaching through Fitness Apps to Support Physical Training. (2018). https://uclic.ucl.ac.uk/content/2-study/4-current-taught-course/1-distinction-projects/12-18/stoica{_}ralucaalexandra{_}2018.pdf

[83] Microsoft Game Studios. 2010. *Xbox 360: Kinect Sports*. Retrieved Feb 11, 2021 from https://xbox.com

[84] Salvatore Tedesco, John Barton, and Brendan O'Flynn. 2017. A review of activity trackers for senior citizens: Research perspectives, commercial landscape and the role of the insurance industry. *Sensors* 17, 6 (2017), 1277.

[85] George Tzetzis, Evandros Votsis, and Thomas Kourtessis. 2008. The effect of different corrective feedback methods on the outcome and self confidence of young athletes. *Journal of sports science & medicine* 7, 3 (2008), 371.

[86] Gabriela Villalobos-Zúñiga and Mauro Cherubini. 2020. Apps That Motivate: a Taxonomy of App Features Based on Self-Determination Theory. *International Journal of Human-Computer Studies* 140 (2020), 102449.

[87] Kimberly Voon, Ilan Silberstein, Aditya Eranki, Michael Phillips, Fiona M. Wood, and Dale W. Edgar. 2016. Xbox Kinect™ based rehabilitation as a feasible adjunct for minor upper limb burns rehabilitation: A pilot RCT. *Burns* 42, 8 (2016), 1797–1804. https://doi.org/10.1016/j.burns.2016.06.007

[88] Paweł W. Woundefinedniak, Przemysław Piotr Kucharski, Maartje M.A. de Graaf, and Jasmin Niess. 2020. *Exploring Understandable Algorithms to Suggest Fitness Tracker Goals That Foster Commitment.* Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3419249.3420131

[89] Wei Xu. 2019. Toward Human-Centered AI: A Perspective from Human-Computer Interaction. *Interactions* 26, 4 (June 2019), 42–46. https://doi.org/10.1145/3328485

[90] Wei Yang, Shuang Li, Wanli Ouyang, Hongsheng Li, and Xiaogang Wang. 2017. Learning feature pyramids for human pose estimation. In *proceedings of the IEEE international conference on computer vision*. 1281–1290.

[91] Krystyna Zatoń and Stefan Szczepan. 2014. The impact of immediate verbal feedback on the improvement of swimming technique. *Journal of Human Kinetics* 41, 1 (2014), 143–154.

[92] Xiaoxiao Zhang and Xiaoge Xu. 2020. Continuous use of fitness apps and shaping factors among college students: A mixed-method investigation. *International Journal of Nursing Sciences* (2020).

[93] Zhengyou Zhang. 2012. Microsoft kinect sensor and its effect. *IEEE multimedia* 19, 2 (2012), 4–10.

A APPENDICES

A.1 Interview Questions

We used a semi-structured approach to interviewing our participants using the following questions.

A.1.1 Introduction.

- Can you tell me about your experience of using the three apps (list them to remind the names)?
- Can you describe the format that each app used to teach the class?
- Which class format did find easiest to understand the instructions from the instructor?

A.1.2 Visual Feedback.

- What visual feedback did each app offer?
- Can you tell me about when you made use of the visual overlay?
- Can you tell me about a time when the AI didn't work?
- Do you think it's important to see the instructor on screen?
- Is it important to see yourself on screen whilst you work out?
- Would you rather see visual feedback continuously or only just when you get something wrong.
- Did any of the apps include any real-time metrics on screen (reps, intensity graph, calories, pace in seconds)?
- For exercises you completed do you think having detailed instructions would be useful or is video demo enough?

A.1.3 Audible Feedback.

- Can you think of a time when the instructor gave you some instructional feedback?
- How effective was that feedback?
- Did you feel that the feedback was appropriate (what it was saying or how often it was giving you feedback?

A.1.4 Motivation.

- How did the instructor motivate you through the class?
- What kinds of things did the instructor say or show you to help motivate you through the class?

A.1.5 Reviewing Data.

- Did you review your data after you had worked out, if not, is this something you could see yourself doing if you used it for an extended period of time?
- When browsing your workout results what did you want to be able to see and do with that information?

A.1.6 Personalisation and Adapting Over Time.

- After you've used these apps for a period of time how do you think the AI would adapt to you over time?
- When the AI get's something wrong, when would you want to correct it (in the moment or at the end)?
- How would you like to be able to correct it (voice, draw on screen, show it a video etc)?
- What aspects of the AI would you like to change or adapt after your workout?

A.1.7 Configuration.

- Can you remember what you were able to change and configure in these apps?
- Which features are important for you to configure when working out with the AI instructor and why?

A.1.8 Adaptability.

- Do you think it's important for the AI to understand where you are when you're working out?
- Do you think it's important for the AI to know who is in the room with you?
- Would you expect the AI to adapt to this situation, if so, how?

A.1.9 Intelligibility.

- Did the AI ever show or explain to you how it knows what it can see and how it understands your movements?
- Would there be anything that you would want to ask the AI when it made a decision?

A.1.10 Dialogue.

- Can you recall what types of things that the AI fitness instructor spoke to you about?
- Would you ever consider talking to your AI fitness instructor in the future?
- What types of things might you say to an AI fitness instructor?
- How often would you expect the instructor to talk to you during a workout?

A.1.11 Open Design Questions.

- If you could design your own AI instructor, what features would be most important? (considering class format, visual and auditory feedback, what you would see on screen, and how you would configure the app)
- Do you think that any of these apps would adapt to your workout over time, if so how?
- What aspects of the fitness instructor would you want to configure?
- How would you want to interact with the post-workout data?

A.1.12 Closing Questions.

- If you had to pick one of these apps to keep on your phone which one would you continue to use?
- Do you think that an app like this would be incorporated into your workout?

Table 2. AI Fitness Instructor Application Features

	Fitness Ally (v1.7)	**Kaia (v1.6)**	**ONYX (v1.7.32)**	**VAY (v1.2.0)**	**Zenia (v3.8)**	**Peloton (v14.4.0)**
Release date	May '20	Jan '19	April '20	Jan '19	Jan '20	Jun '18
Phone orientation	Portrait	Landscape	Portrait	Portrait	Landscape	Landscape
On-boarding:						
Goal settings	✓		✓	✓	✓	
Demographic settings		✓	✓	✓		✓
Workout Types:						
HIIT	✓		✓			✓
Body-weight		✓	✓	✓		✓
Yoga					✓	✓
Many other						✓
Class Format:						
Follow-along	✓					✓
Hybrid follow-along			✓		✓	
Demo video		✓		✓		
Calibration method:						
User's outline			✓	✓		
User's skeleton		✓			✓	
View on a screen:						
Avatar only	✓					
Instructor only						✓
Instructor & user's reflection		✓		✓	✓	
Instructor & user's outline			✓			
Metrics:						
Repetition count	✓	✓	✓	✓		
Workout Intensity indicator	✓		✓			
Visual corrective feedback		✓			✓	
Verbal corrective feedback	✓	✓	✓	✓		
Confirmatory feedback					✓	
Motivational feedback	✓		✓	✓		✓
Instructor's voice	Human	AI	Human	Human	Human	Human
Post-workout metrics:						
Practice time			✓	✓	✓	
Accuracy/progress score	✓		✓	✓	✓	
Repetitions	✓		✓			
Achievement badges	✓		✓		✓	✓
Wearable integration						✓
Long term progress tracking	✓		✓	✓	✓	

- What would you like to be able to do with this app that you can't currently do?
- Would you recommend it to your friends and family?
- How much would you pay for this app per month?
- Is there any other feedback you would like to discuss that we haven't covered yet?

Table 3. Overview of findings

Theme	Sub-theme	Total Number of Participants	Participants ID
Detection Issues		**11**	
	AI failed to recognise participants correct pose	10	1, 2, 5, 6, 7, 8, 9, 10, 11, 12
	AI failed to correctly count repetitions	5	1, 3, 5, 10, 12
Spatial Limitations		**8**	
	Participants needed a lot of space	7	1, 2, 3, 4, 6, 10, 12
	The screen was too small to see visual feedback	7	1, 2, 3, 4, 5, 6, 10
	Phone orientation affected pose tracking	3	1, 3, 5
Seeing themselves on the screen		**12**	
	Preferred instructor only	4	1, 3, 4, 8
	Preferred seeing instructor and themselves continuously	7	1, 4, 5, 6, 7, 9, 10
	Preferred seeing themselves intermittently	6	2, 5, 6, 10, 11, 12
	Participants wished to configure what they see on the screen	2	1, 2
Skeleton and Body-outline		**10**	
	Participants found skeleton and body-outline useful	7	1, 3, 4, 6, 10, 11, 12
	Participants preferred to see feedback only for incorrect body part	2	4, 10
	Participants found ONYX outline without feedback useless (Out of 6pp)	4	2, 3, 8, 9
Real-time glanceable metrics		**10**	
	Participants found glanceable metrics useful	9	2, 3, 4, 6, 7, 8, 9, 10, 12
	Participants found Intensity measure useful	7	1, 2, 3, 7, 8, 9, 10
	Participants preferred intensity measure of Fitness Ally over ONYX (Out of 4pp)	3	2, 3, 8
Post-workout feedback		**10**	
	Performance scores made no sense	3	2, 4, 8
	Repetitions and pose accuracy most useful to understand progress and to improve overtime	8	1, 2, 3, 4, 6, 8, 10, 12
	Exercise balance check feature	1	4
	Side by side video feature	3	3, 6, 7
	Picture comparison over time feature	1	11
Corrective feedback		**12**	
	Corrective verbal feedback useful in general	6	1, 2, 3, 8, 10, 12
	Corrective verbal feedback useful when Participant cannot see the screen	4	1, 2, 8, 9
	Confirmatory feedback is useful	5	2, 7, 8, 10, 12
	Repetitive phrases are frustrating	9	1, 2, 3, 5, 6, 7, 9, 10, 11
	Corrective the user too often is undesirable	5	1, 2, 4, 7, 8
Conversation qualities		**12**	
	AI feedback over-complimentary	10	1, 2, 4, 5, 7, 8, 9, 10, 11, 12
	AI feedback at a wrong time and context	5	3, 5, 6, 7, 9
	AI has no notion of time	1	5
	AI's playful personality perceived favourably	4	1, 2, 3, 9
Engaging conversation with AI		**9**	
	Participants had low expectations of AI's speech comprehension	7	2, 3, 4, 7, 8, 10, 12
	Voice commands perceived favourably	7	1, 2, 4, 6, 7, 8, 12
Adjusting workouts based on the user's individuality		**10**	
	Participants wished for better on-boarding feature	5	3, 5, 6, 8, 9
	Participants wished for workouts to be based on user's goals	7	1, 2, 3, 4, 6, 8, 11
	Participants wished AI taken into consideration injuries and disabilities	8	1, 2, 4, 5, 6, 8, 9, 10
Adjusting workouts over time		**10**	
	Participants wished for the AI to modify workouts over time	9	1, 2, 3, 4, 6, 8, 9, 10, 12
	Participants wished for the AI to modify exercises in real-time	5	3, 6, 8, 9, 10
	Participants wished that the AI used verbal feedback based on user's past performance	9	2, 4, 5, 6, 8, 9, 10, 11, 12
Instructor-led experience		**8**	
	Participants preferred follow-along class format	8	1, 2, 3, 8, 9, 10, 11, 12
	Participants found hybrid-follow along class format less engaging	4	2, 3, 9, 10
	Participants found that demo class format interrupted workout flow	4	1, 2, 7, 12
	Participants liked avatar instructor (Out of 6pp)	4	1, 2, 3, 8
Workout configuration		**12**	
	Participants wished more freedom to configure workout features in general	9	1, 2, 3, 4, 6, 7, 8, 9, 11
	Participants found music an important aspect of workout	5	4, 8, 9, 10, 11
	Participants wished to remove disliked exercises	2	8, 9
	Participants trusted AI to recommend them a workout	7	1, 2, 4, 5, 8, 10, 11
	Participants wished for an option to create their own workout	3	2, 7, 12
	Participants wished for an option to modify AI recommended workouts	2	2, 4

Table 4. Definitions

Key terms	**Definitions**
Follow-along	A workout format that uses a continuous video of the same instructor or avatar performing exercises and guiding the participant throughout exercises and rest times.
Hybrid follow-along	A workout format that uses a series of short looping video clips for the duration of each exercise, often with different instructors shown in the video. Both the video clips and the user's mirrored image are shown on the screen.
Demo video	A workout format that begins with presenting a short instructional video and then switches to display only the user's pose captured by the front camera.
Calibration method (User's outline)	At the start of the AI app workout, the users are asked to place the phone in a specific position in order to make sure optimal conditions are met for computer vision tracking. The user is required to fit the length and width of their body into a body shape outline.
Calibration method (User's skeleton)	At the start of the AI app workout, the users are asked to place the phone in specific position in order to make sure optimal conditions are met for computer vision tracking. The user is provided with a skeleton overlay that lights up green in correct calibration.
Visual corrective feedback	Visual feedback that users receive to monitor and/or correct their pose (i.e. skeleton overlay that lights up red if the pose is incorrect or stays green to indicate the pose is correct)
Verbal corrective feedback	Verbal feedback that users receive from AI to correct their pose (e.g. "lift your hips up to keep it in line with your spine")
Confirmatory feedback	Verbal feedback that users receive from AI to confirm they have corrected their form. (e.g. "That's much better")
Motivational feedback	Verbal feedback that users receive from AI to motivate them (e.g. "Keep going!", "You're doing great!")