Chapter 2
Augmented Reality User's Experience: AI-Based Data Collection, Processing and Analysis

Check for updates

Yevgeniya Sulema, Andreas Pester, Bertrand Laforge, and Frederic Andres

Abstract This chapter presents an AI-based approach to the systemic collection of user experience data for further analysis. This is an important task because user feedback is essential in many use cases, such as serious games, tourist and museum applications, food recognition applications, and other software based on Augmented Reality (AR). For AR game-based learning environments user feedback can be provided as Multimodal Learning Analytics (MMLA) which has been emerging in the past years as it exploits the fusion of sensors and data mining techniques. A wide range of sensors have been used by MMLA experiments, ranging from those collecting students' motoric (relating to muscular movement) and physiological (heart, brain, skin, etc.) behaviour, to those capturing social (proximity), situational, and environmental (location, noise) contexts in which learners are placed. Recent research achievements in this area have resulted in several techniques for gathering user experience data, including eye-movement tracking, mood tracking, facial expression recognition, etc. As a result of user's activity monitoring during AR-based software use, it is possible to obtain temporal multimodal data that requires rectifying, fusion, and analysis. These procedures can be based on Artificial Intelligence, Fuzzy logic, algebraic systems of aggregates, and other approaches. This

Y. Sulema (✉)
Department of Computer Systems Software, Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine
e-mail: sulema@pzks.fpm.kpi.ua

A. Pester
Faculty of Informatics and Computer Science, The British University in Egypt, Cairo, Egypt
e-mail: andreas.pester@bue.edu.eg

B. Laforge
Laboratoire de Physique Nucléaire et des Hautes Energies (LPNHE), Sorbonne Université, Paris, France
e-mail: laforge@lpnhe.in2p3.fr

F. Andres
Digital Content and Media Sciences Research Division, National Institute of Informatics, Tokyo, Japan
e-mail: andres@nii.ac.jp

31

chapter covers theoretical and practical aspects of handling AR user's experience data, in particular, MMLA data. The chapter gives an overview of sensors, tools, and techniques for MMLA data gathering as well as presenting several approaches and methods for user experience data processing and analysis.

2.1 Introduction

Augmented Reality (AR) technology offers a new dimension to serious games, touristic and museum applications, food recognition applications, and other similar software. At the same time, ensuring the high quality of AR-based software requires the monitoring of the user's experience. For AR game-based learning environments user feedback can be provided as Multimodal Learning Analytics (MMLA) which has been emerging in the past years as it exploited the fusion of sensors and data mining techniques.

To collect MMLA data, we can use several devices, tools, and techniques which will complement the raw educational data coming from a digital educational application. For instance, one can use a camera whose image stream is analysed online by an AI algorithm to provide information about the body or face language of a learner. One can also use distinct types of bio-oriented sensors measuring blood-pressure, heart pulse, breath rhythm and importantly neuro-devices reporting about different signals inside the brain (some of them being only accessible on specific research platforms such as NeuroSpin (2022) but some devices can now also be accessible in a class). An important set of useful sensors are also available in the hands of the learners, namely their mobile phone and modern digital watches. Such connected devices are of great interest because they provide data collection options at large scale and low price.

Data presenting MMLA requires processing that can be based on Artificial Intelligence (AI) methods. AI can be used for real-time feedback to the learner, for aggregation of information about the learner's behaviour in-situ, provide summarised knowledge after the learning session to guide the learner in different learning situations, to update the digital learner model, and to give feedback to the developers of the AR systems. The AI methods can predict and classify learning situations as learner behaviour as well.

Thus, in this chapter, we present important aspects of AR-software user experience data collecting, processing, and analysis based on AI. The rest of the chapter is organised as follows. In Sect. 2.2, we analyse recent related work. In Sect. 2.3, we consider use cases and scenarios for serious AR-based games. In Sect. 2.4, we focus on systemic data collection. In Sect. 2.5, we present a range of capturing devices and tools that enable the monitoring of user experience and activity parameters. In Sect. 2.6, we discuss AI-enhanced data processing methods. In Sect. 2.7, we consider the creation of a complex model of an AR-software user. Finally, in Sect. 2.8, we summarise the presented approach.

2.2 Related Work

Recent research demonstrates special attention to the implementation of innovative approaches such as AI and digital twins in the educational area.

Cowling and Birt (2020) give an overview of the most common approaches for Mixed reality multimodal learning analytics.

Ochoa and Worsley (2016) analyse how and why audio and video can be used in non-technology-centred learning environments to collect signals (data) about the learners' actions and interactions. Several variants of multimodal learning analytics are discussed.

Di Mitri et al. (2018) conducted a literature survey of experiments using multi-modal data for multimodal learning analytics. They analysed input multimodal data and the related learning theories. Finally, they formulated a Multimodal Learning Analytics Model with three objectives:

- Enhancing the feedback in a learning context.
- Combining several machine learning algorithms with multimodal data.
- Aligning the used machine learning and learning science terminology.

Marcel (2019) provides a field study about different mobile Augmented Reality applications and learning objects for higher education on the mobile AR platform HP Reveal.

Spikol et al. (2018) provide an experimental study on how to find the most notable features for project-based learning set-ups using supervised machine learning for multimodal learning analytics.

Guo et al. (2022) give an overview of actual trends in the use of AI to enhance the metaverse project. These trends are not only related to Metaverse but have an impact on the area of computer vision and natural language processing, as the sensor architecture as well.

Tao and Xu (2022) analyse the application of digital twin technology in different areas including aviation and education. The authors come to the conclusion that digital twin technology has potential for the field of education. To prove this, they present an example of constructing a digital twin for education based on HTC Vive.

Berisha-Gawlowski et al. (2021) considers the potential of the digital twin from two perspectives: (1) application of digital twin technology to control and manage a cyber-physical system, and (2) representation of humans in the virtual world. Considering the application of digital twin technology in the field of education, the authors raise questions related to both technological and methodological aspects of using digital twins for educational purposes. To illustrate their vision, the authors refer to the example of team learning.

Rudra (2022) analyses the possibility of digital twin technology becoming the next step in immersive learning. The author provides ideas on enhancing learning in medicine and chemistry by employing digital twin technology in combination with AR.

The analysis of these and other recent papers allows us to draw the conclusion that these technologies have a high potential for AR-based educational game applications as well.

2.3 Use Cases for Educational AR Games

Given the current constraints on energy consumption, while targeting the largest possible usage of an educational game, it is important to design AR games with the minimum amount of technology that allows it to meet its educative objectives. A question is then to identify cases where AR is worth implementing in an educative game. On the other hand, educational games including AR are an effective way to collect data for MMLA.

The first obvious use case is when a game is using the camera of a mobile phone or a computer to enrich the scene with live information overlaid with the real image taken by the player. In terms of education, this can be a modality to identify in the scene some objects of interest to play with or to bring complementary information to the learner. Depending on the amount of such information that a learner may need, can for instance gain clues about the autonomy of the learner to achieve a specific task. Let us imagine for instance a game that teaches security measures inside a laboratory or a firm's workshop: "how many critical situations can be identified by the learner in full autonomy?" and "how much help should be brought to detect some missed cases?" are questions that are of high relevance to assessing the competence of the learner.

The second case of interest is provided by a game aiming at discovering the full history of a painting and how this history was learned. In such a game, the learner could try to inspect the real painting with his camera while having some extra tools to perform dedicated analysis (it could be a magnifying glass showing the x-ray picture of a part of the painting or a simulated spectrum that is acquired using a proton or neutron beam to inspect the composition of the painting, or it could be a tool that reveals the modifications brought by the conservators over time, etc.). Such complementary information is of importance in presenting art pieces in a very new way.

The third case of importance is given by the interesting options that are provided by AR in terms of discovering levels of reality. Our immediate environment is indeed not what it seems to be as seen by our human senses. In a class or a museum of sciences, it is remarkably interesting to have the ability to change one's scale of observation, for instance, to go to the molecular level by providing a simulation at a scale chosen by the learner. This can be particularly useful to teach physics or chemistry. Consider, for example, how difficult it is to discover what a magnet is in terms of atomic properties and how AR can support this learning task.

2.4 Systemic Data Collection

Systemic data collection in the context of digital education aims at collecting nominative data that will be used by learning analytics tools to provide immediate or delayed personalised feedback to learners while also providing large-scale data samples to support meaningful research and to train complex AI algorithms. Research can be devoted to demonstrating interesting features of specific pedagogical approaches such as AR-based game training but can also deal with the difficult problem of connecting the dots between a model of the learner, a curriculum, and the effective learning achieved as the result of this curriculum.

Large-scale data collection means going beyond a single educational institution scale and requires a specific design that can be generalised to many educational structures. The need for nominative data handling at the level of each institution is best achieved using a standard data format such as Learning Record Stores (LRS) using the xAPI ecosystem (Šimić et al. 2019; Rustici Software 2022). On the other hand, the usage of data from different institutions for research purposes is also needed and requires interoperable data formats and tools. This second need also requires installing a Data Processing Unit (DPU) that simultaneously has direct access to the local data and can communicate with other DPUs. With such a design, it is possible to analyse data locally but also request data processing at other sites. The privacy of data is of course questioned by such a large-scale analysis. This question can be solved by adding a software layer (API) that guarantees that processing outputs at any DPU can only be transmitted to other DPUs if they are aggregated results from the processing of a collection of individual information. An obvious example would be the output of a numerator and a denominator at each DPU level to reconstruct a percentage at the requesting DPU level. A more elaborated one could be the transmission of a mean gradient calculated over a set of learners at a given site transmitted to the requesting DPU to train a Machine Learning algorithm. Such a system can be viewed as an edge processing scheme, and it is then easy to expand to a larger scale. The CPU power will be used where the data is, and only aggregated results would be transmitted over the network. This guarantees the privacy of the data at each site and is potentially more efficient in terms of storage and network traffic than extracting anonymous data from each site to be centralised on a research cluster. It should be noticed that such a design is in line with current efforts made in Europe (Gaia-X Hub Germany 2022; Gaia-X European Association for Data and Cloud AISBL 2021; Prometheus 2022) to develop edge computing, interoperability of data, and sovereign cloud computing.

It is important to mention that the xAPI standard is unfortunately so flexible that further standardisation needs to be done to collect data to be used in an interoperable way. Going beyond the "actor + verb + complement" specification of xAPI is a necessity to allow global analyses of heterogeneous curricula in a coherent way while still allowing a detailed and specific analysis of a particular curriculum or educational tool. The xAPI ecosystem provides semantic versioning of statement

vocabulary that allows Clients and LRSs to remain interoperable as the specification changes but there is much more to define in the context of many specific educative activities, especially for serious games and AR activities.

In the context of the Ikigai (2022) serious gaming platform (Fig. 2.1) developed in France by a consortium of higher education institutions and handled by a non-profit organisation known as Ikigai-Games for Citizens, a design for such specifications has been elaborated defining a hierarchy of xAPI statements that goes from global to specific concerns. Consequently, each game has to declare for instance a "start of game" and "end of game" that allows counting the number of games that have been played by learners regardless of the type of games. Another level of mandatory information requires the game to report specific activities that can be linked to a category of games, for instance, the registration of each shoot in a first-person shooter game. This specification also allows game-specific xAPI statements. On purpose, this hierarchy allows researchers to develop analyses with different granularities in a coherent way either aggregating data from different games to provide for instance global statistics about the overall usage of games in the educational activities of learners or choosing a lower granularity to provide statistical information of the impact or usage of a specific game type on student performance up to game specific information. This multiscale approach can be extended to other types of digital educative activities and requires a community effort to specify for instance the generic information that is needed to study AR systems such as an AR serious game. Learning analytics-based platforms should then have access to a hierarchy of AR system related xAPI statements when defining the type of data, they provide through AR activities. The same information should be also available to data analysts. Moreover, one needs a mechanism that guarantees that the collected information collected at two different periods can be analysed simultaneously. This feature will be optimally implemented through a versioning system of the AR xAPI information hierarchy available through a dedicated server used by all platforms. With such a system any learning analytics-based platform will be able to define the type of data that it provides in a way that will also be immediately available to data analysts through the same server. Interoperability of this metadata will be further simplified by a possible unique request to this centralised server providing the description of the available information in a dataset using a specific version of the hierarchy of xAPI statements.

In terms of infrastructures, such a large-scale system, designed to provide personalised guidance to students based on their individual data, also need to be thought of in association with a complementary pedagogical resource management system. This latter provides a mechanism to find content to recommend to the learners. Given the ongoing evolution towards a competence approach at the higher education level, a remarkably interesting framework is given by the Memorae platform developed at Université Technologique of Compiègne in France (Atrash et al. 2015). This platform implements the idea that a curriculum can be described as a tree or a graph of related competencies by a teacher. Each node is related to a competence while a graph connection models the relative dependence between two individual competencies. This map can then be made available to a learning community as an interactive tool allowing one to enter at the level of each node into a forum dedicated to its related
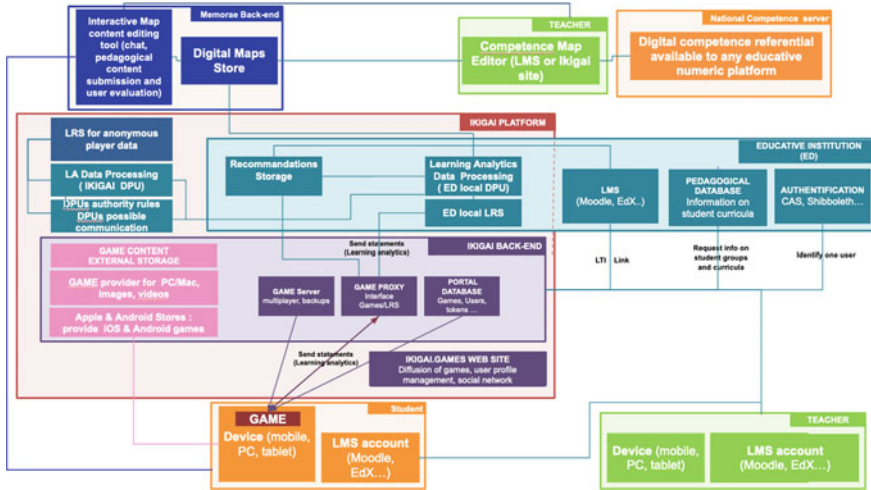
Fig. 2.1 Ikigai platform architecture

competencies. It further enables recommending content to work this competence and to evaluate the recommended contents (a PDF-file, a book, a film, a game, a simulation, etc., related to that competence). Beyond live or asynchronous exchanges between learners, this interactive map is then an educative content database that can be further used as a data structure to recommend content to be used by a student in the context of pedagogical advice provided by an AI-based recommendation system. The same map can be used to project single player data to identify its level of mastering of the various competencies defined in the map and then used to make those recommendations.

2.5 Monitoring Parameters Capturing Hardware and Tools

Multimodal analysis has demonstrated effectiveness in studying and modelling several human–human and human–computer interactions. In this section, we review the role of parameter capturing hardware and multimodal tools in the service of studying complex learning analytics environments.

One important set of parameters represents the affective factors (e.g., motivation, stress, or flow) and can be measured in diverse ways (e.g., clicks, postings, messages, views, writes and likes). Several studies (Kumar et al. 2022; Siddharth and Sejnowski 2022) demonstrated that there is a correlation between certain types of sensor data, so called biomarkers (e.g., heart rate), and higher-level states of persons that are relevant for learning (e.g., emotion, including anxiety). In the domain of IoT, a digital biomarker is defined as digitized data collected from learners via IoT devices (Nam et al. 2019). Learners' biomarkers are collected with three types

of affordable sensors: an eye-tracker, an electroencephalogram, and a camera that monitors variations in different modalities, e.g., speaking, gesturing, gazing, typing (Lazar et al. 2017). The use of sensor technology enables the collection of learners' physiological and behavioural data. These data are the tracking of all the attitudes, interests, and motivations of the learners toward the educational experience, summarised as affective learning (Bamidis 2017).

To monitor the student's activity during a learning session, we can also use devices that capture and track the movement of the students' eyes: features including pupil size, saccade, fixations, velocity, blink, pupil position, fixation, electrooculogram, and gaze point (Eye Square 2022; Sangu 2020; Wang et al. 2018). Collected eye-tracking data serves to identify and analyse patterns of visual attention of individuals during learning experiences. There are four types of eye trackers:

- Tower mounted type (INITION London 2022; University of Edinburgh 2022)
- Screen-based type (Grossman et al. 2019; Biopac Systems Inc. 2022)
- Head mounted type (Sugano and Bulling 2015; Cognolato et al. 2018; Melnyk et al. 2022; Franchak and Chen 2022)
- Mobile type (Callemein et al. 2019; Müller et al. 2019; Liu et al. 2019).

Innovative approaches to include mood flow in learning experiences have been using non-intrusive mood-related sensors (Nashed et al. 2021). In general, Multimodal Data Processing uses several channels and spaces for monitoring. Di Mitri et al. (2018) distinguish an input space (different sensor signals) and a hypothesis space (latent attributes). The difference is in observability: input signals can be observed directly, and latent parameters can be concluded based on different inference models. The parameters monitored in these channels are mostly person and application centralised. It takes less into consideration the requirements of large-scale data collection described in Sect. 2.4.

One question is also, where the data are processed: in edge computing directly, on the mobile device, or centralised in the cloud. This needs standardisation. The standardisation of data processing should be in accordance with the standardisation, granulation, and semantic versioning of the systemic data collection.

For AR applications a smaller set of parameters can be used, including direct spatial data, motoric and behavioural data of the user of the learning game, and data from the interoperability of gamer AR user groups.

The implementation of the feature fusion strategy of deep neural networks for multimodal activity recognition faces the challenge of integrating features of different scales for better performance, as it has been shown in (Dai et al. 2021). The authors of (Münzner et al. 2017) classified the feature fusion strategy in four categories, initially applied in CNN-based architectures:

- Feature Fusion
- Sensor-based Fusion
- Axis-based Fusion
- Shared-filter Fusion.

Research enhancement has been focussing on the generalisation of feature fusion to any deep learning architectures (Schweizer et al. 2021; Siegfried and Odobez 2022; Sugano and Bulling 2015).

2.6 AI-Enhanced Data Processing Methods

For the Data Processing of the MMLA data, different methods can be used: Machine and deep learning classification methods, fuzzy logic predictions of linguistic variables, but also a special algebra of aggregates, developed for the handling of multimodal data in general.

The methods of processing data can be characterised by the algorithms used, the location of processing (edge or cloud-computing), the explainability of the results, and the used output.

From an algorithmic point of view, the methods differ in the question of whether they are data-driven (machine and deep learning) or rule-driven (fuzzy logic, algebra). In the first case the collected, cleaned, and fused data are used to train and test a machine learning model (for the user or learner or for the relation game input and learning outcome). Different methods had been used, for example, regression models, Bayes naive learning, support vector machines with linear and Gaussian kernels, and distinct types of artificial neural networks. In the last two years primarily transformer algorithms for language-based models have been used. The problem of transformers is that they need a huge amount of memory and computational resources and cannot be used directly by small or medium scale platforms or devices. In the future methods, which can take into consideration the multiple connections between the multimodal input data, like graph neural networks can be used. If multimodal data are considered as multivariate sequential data, methods used for time series are also an interesting approach.

To ensure consistency of data processing, the algebraic system of aggregates (ASA) can be used in combination with the fuzzy logic approach (Sulema and Kerre 2020). The ASA offers a way of aggregating data sets that can be used for the synchronisation and consolidation of data describing the same object of research (a learner in our case) if the data streams are received from multiple sources (sensors, devices, tools, etc.). It is especially important in the case of temporal multimodal data processing. An aggregate in the ASA is defined as a mathematical object that has the following distinguishing features:

- An aggregate consists of data tuples that are ordered; the order of tuples is important for the result of operations on aggregates (logical operations and ordering operations).
- A tuple in the aggregate can consist of values of the same type, values of several types, or tuples of values.

These features enable multiple variants of time-wise data processing in an effortless way.

Also, important is the location of data processing. Most of the AI applications are processed on a powerful computer with a GPU processor with enough RAM or directly on the cloud by different machine learning service providers. All Large Language Models are cloud applications. But AR applications are running on a smartphone or tablet and so far in a lot of cases the data should be processed directly on the device, otherwise, the latency time would be too long. Therefore, the algorithms should be dressed in such a way that they fit the device hardware. Some smartphones are already able to handle larger models, but AI on the edge is still an extensive research option. Sometimes a CNN accelerator is implemented directly in the visual recognition sensor unit, such as ShiDianNao, RedEye, and other applications (see in Guo et al. 2022). Another approach is to compress the models or to work with edge inference machines. A third way to solve the computing bottleneck is the development of in-memory computing, SRAM-based, DRAM-based, and novel memory technologies. Applying AI to more complex AR will require solving the problem of how to increase the capacity to run larger-scale Deep Learning models directly on a device.

A major challenge in Machine and Deep Learning and for AI applications to AR is the explainability of the results. On the one hand, the MMLA results are used to model the learners' behaviour, on the other side, with the help of modern generative methods like stable diffusion, it will be possible to create AR objects directly during the learner's communication with the learning object and the learning community. This can have an impact on the learner's behaviour, which we should be able to understand. The used deep learning methods are seen as a black box and the results so far require explanation. For classical machine learning, with the methods SHAP and LIME, explainability can be supported (Gashi et al. 2022), but for deep learning this is still a question under research.

The output of the data processing in MMLA can be represented as a classification of the learner/user/feedback type, detection of abnormal behaviour, but can also include sorting and compressing of the features of the input data. If Graph Neural Networks are used, a new node (virtual object) or edge (for the learner unbelievably valuable information), can be predicted, but also learning phases can be sorted by score.

With these methods, the mental and emotional situation of the learners can be learned and used for learning analytics and as input to the user's digital twin.

2.7 AR-Game User's Digital Twin

An educational AR-game user's digital twin is a continuously updating model that reflects individual features of the game player: a level of involvement in the study process, learning strategy and behaviour, etc. This model can be used for providing the learner with recommendations on the individual study trajectory for achieving

better progress in learning. A collection of digital twins can be used for gathering generalised anonymized statistics about the AR-game quality, typical strategies, and learners' behaviour to help the game developers in game improvement.

The user's digital twin is represented through data obtained on the level of systemic data collection discussed in Sect. 2.4 of this chapter.

Beyond the data collection infrastructure itself, the hierarchy of information must be flexible enough to handle heterogeneous types of information. This can be achieved by defining a list of devices that provide the information, with a description of their capabilities as metadata, and add them as actors in the reporting of the actual scene of the games: "the thermometer reports that the face temperature of the learner is 38 °C", "the camera has been turned to infrared mode", etc. should be possible statements in the data collection. This type of information is complementary to the usual learner action or achievement information discussed previously and reported as a statement involving the learner as an actor. Extending the nature of available information to be processed by AI algorithms is a key point of improvement of current learner experience data collection since they should bring much more contextual information that allows a better understanding of the conditions under which an educative objective is met or not. All is not always a question of digesting formal information. Known facts about how a learner achieves memorising is very dependent on the psychological state of his/her brain, suggesting the use of neuroscience devices in the learner experience reporting.

To construct the user's digital twin based on data to be received from different devices and tools, we can use a formal specification to represent the important level of the user's model. This specification can be considered as a semantic model of the data describing the user as an object of research. Figure 2.2 shows an example of the semantic model for the user's digital twin creation.
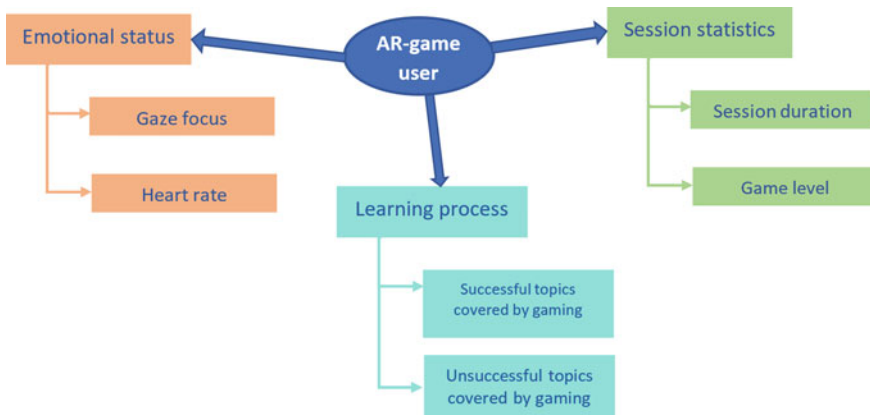

Fig. 2.2 AR-game user's semantic model

To develop the digital twin, we need to implement this model as a software tool that receives data streams according to the defined semantics, fuses them based on the synchronisation rules, and provides AI-based analysis and processing of aggregated data. The architecture of such software is shown in Fig. 2.3.

This architecture implements the following logic of using the AR-game user's digital twin. During an AR-gaming session, the user's parameters defined by the AR-game user's semantic model are being monitored using a set of devices and tools. These data streams come to the Data Fusion Module where they are synchronised and aggregated according to the data synchronisation rules. The data aggregate is stored in the cloud Data Storage. Using the AR-game user's individual data, the AI-based Analytics Module, through the Visualisation Module, provides the learner with analytics on study progress. It also provides the Individual Learning Recommendation System with the data to be used for helping the learner in forming an individual learning trajectory aimed at achieving higher quality of learning.

A teacher can use the system to get generalised information about learners. The queries for consolidated data analytics are processed by the Query Processing Module, which gets data from the cloud Data Storage and sends the requests for data processing to the AI-based Analytics Module. The resulting analytics is demonstrated by the Visualisation Module. These analytics help the teacher to improve the methodology of the AR-game. Besides, the AI-based Analytics Module provides feedback to the AR-game developer through the Game Development Feedback mechanism, enabling the advancement of the game development.

Thus, the general framework of AR-game user's digital twin software implementation and exploitation includes:

- The game development and advancement use case.
- The game efficiency analysis use case.
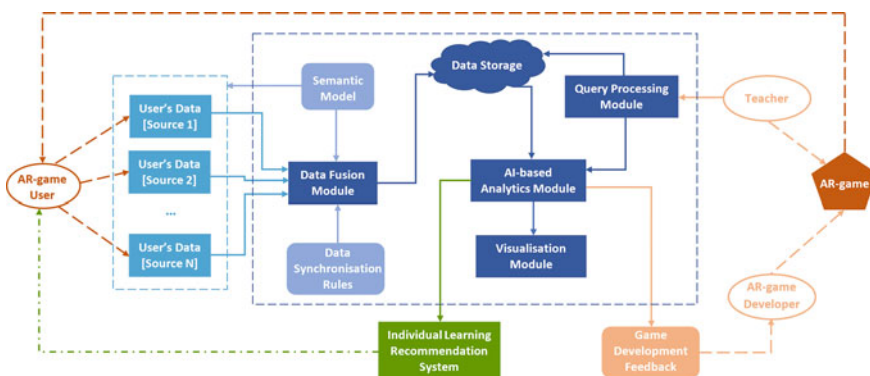- The studying process enhancement use case.

Fig. 2.3 AR-game user's digital twin software architecture in the general context

In the game development context, the AR-game user's digital twin can be used for obtaining analytics for a better understanding of the necessary advancement of the game: its logic, AR elements, controls, etc. In the game efficiency and learning strategies context, the digital twin can be used for the improvement of the teaching methodology. In the study context, the digital twin can be used for individual learning trajectory elaboration and recommending to the learner.

2.8 Conclusion

The systemic approach developed to collect data efficiently and at large scale in the domain of educative games reported here can be extended to AR games and other digital learning tools in a straightforward way and provides a minimal mechanism to handle educative recommendations for all the disciplines built by the learning communities themselves. By including heterogeneous types of data inside a global hierarchy of content, this system will allow the educative community to build complex models of learners and of their related curricula. Such a systemic approach will break the current limits that prevents extensive usage of AI in education. This system will further enhance the usage of data because it allows both a nominative usage of the data at the scale of a local institution and an anonymised access to these data for large scale research in education.

Multi-modal sensory data captures cognitive, motivational, and emotional behaviour of the learners during the learning processes. All those perceptions enable us to define the learners' affective learning.

AI enhances AR application for games and education in three aspects: (1) smart user feedback prediction and classification, (2) online learning (training) of the AI model based on the systemic data collection and (3) and predictions for similar user and use-cases. In the future, AI for AR research should concentrate on the AR object creation process based on modern technologies like text/speech to image, image to image diffusion, NLP for steering the scenes, and knowledge graphs for interoperational metadata exchange and processing.

The digital twin software is a useful tool that is aimed at the improvement of three important aspects: (1) AR-game design at the systemic level, (2) game exploitation teaching methodology, and (3) game-based studying. Future research in this context can be focused on the elaboration of the standardisation of AR-game user's digital twin software and accumulation of AR-based game exploitation best practices.

References

Atrash A, Abel MH, Moulin C (2015) Notes and annotations as information resources in a social networking platform. Comput Human Behav Elsevier 51:1261–1267

Bamidis PD (2017) Affective Learning: Principles, Technologies, Practice. In: Frasson C, Kostopoulos G (eds) Brain function assessment in learning. BFAL 2017 (Lecture notes in computer science), vol 10512. Springer, Cham. https://doi.org/10.1007/978-3-319-67615-9_1

Berisha-Gawlowski A, Caruso C, Harteis C (2021) The concept of a digital twin and its potential for learning organizations. In: Ifenthaler D, Hofhues S, Egloffstein M, Helbig C (eds) Digital transformation of learning organizations. Springer, Cham. https://doi.org/10.1007/978-3-030-55878-9_6

Biopac Systems Inc (2022) Robust eye tracking in a compact module. https://www.biopac.com/product/eye-tracking-eyetech/. Accessed 27 Oct 2022

Callemein T, Van Beeck K, Brône G, Goedemé T (2019) Automated analysis of eye-tracker-based human-human interaction studies. In: Kim K, Baek N (eds) ICISA 2018 (Lecture notes in electrical engineering), vol 514. Springer, Singapore. https://doi.org/10.1007/978-981-13-1056-0_50

Cognolato M, Atzori M, Müller H (2018) Head-mounted eye gaze tracking devices: an overview of modern devices and recent advances. J Rehabilit Assistive Technol Eng 5. https://doi.org/10.1177/2055668318773991

Cowling MA, Birt JR (2020) Mixed reality multimodal learning analytics. Encycl. Educ. Innov. 2020. https://doi.org/10.1007/978-981-13-2262-4

Dai Y, Gieseke F, Oehmcke S, Wu Y, Barnard K (2021) Attentional feature fusion. In: 2021 IEEE WACV2021, pp 3559–3568. https://doi.org/10.1109/WACV48630.2021.00360

Di Mitri D, Schneider J, Specht M, Drachsler H (2018) From signals to knowledge: a conceptual model for multimodal learning analytics. J Comput Assist Learn 34:338–349. https://doi.org/10.1111/jcal.12288

Eye Square (2022) Head mounted eye tracking https://www.eye-square.com/en/headmounted-eye-tracking/. Accessed 27 Oct 2022

Franchak JM, Chen Y (2022) Beyond screen time: Using head-mounted eye tracking to study natural behavior, Editor(s): Rick O. Gilmore, Jeffrey. J Lockman, Adv Child Develop Behav JAI 62:61–91. https://doi.org/10.1016/bs.acdb.2021.11.001

Gaia-X European Association for Data and Cloud AISBL (2021) Gaia-X architecture document. https://docs.gaia-x.eu/technical-committee/architecture-document/latest/. Accessed 27 Oct 2022

Gaia-X Hub Germany (2022) What is Gaia-X? https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html. Accessed 27 Oct 2022

Gashi M, Vuković M, Jekic N, Thalmann S, Holzinger A, Jean-Quartier C, Jeanquartier F (2022) State-of-the-art explainability methods with focus on visual analytics showcased by Glioma classification. BioMedInformatics 2:139–158. https://doi.org/10.3390/biomedinformatics2010009

Grossman RB, Zane E, Mertens J et al. (2019) Facetime versus screentime: Gaze patterns to live and video social stimuli in adolescents with ASD. Sci Rep 9:12643. https://doi.org/10.1038/s41598-019-49039-7

Guo Y, Yu T, Wu J et al. (2022) Artificial intelligence for metaverse: a framework. CAAI Artif Intell Res 1(1):54–67. https://doi.org/10.26599/AIR.2022.9150004

Ikigai (2022) Games for Citizens. https://www.gfc.ikigai.games/?lang=en. Accessed 27 Oct 2022

INITION London (2022) Motion capture & tracking sensomotoric instruments iView X Hi-speed. https://www.inition.co.uk/product/sensomotoric-instruments-iview-x-hi-speed/. Accessed 27 Oct 2022

Kumar A, Sharma K, Sharma A (2022) MEmoR: a multimodal emotion recognition using affective biomarkers for smart prediction of emotional health for people analytics in smart industries. Image vis Comput 123:104483. https://doi.org/10.1016/j.imavis.2022.104483

Lazar J, Feng JH, Hochheiser H (2017) Research methods in human-computer interaction. Cambridge, MA: Morgan Kaufmann. ISBN: 978-0-12-805390-4

Liu ZX, Liu Y, Gao X (2019) Using mobile eye tracking to evaluate the satisfaction with service office. In: Marcus A, Wang W (eds) Design, user experience, and usability. Practice and case studies. HCII 2019, (Lecture notes in computer science). Springer, Cham, p 11586. https://doi.org/10.1007/978-3-030-23535-2_14

Marcel F (2019) Mobile augmented reality learning objects in higher education. Res Learn Technol 27. https://doi.org/10.25304/rlt.v27.2133

Melnyk R, Chen Y, Holler T et al (2022) Utilizing head-mounted eye trackers to analyze patterns and decision-making strategies of 3D virtual modelling platform (IRIS™) during preoperative planning for renal cancer surgeries. World J Urol 40:651–658. https://doi.org/10.1007/s00345-021-03906-z

Müller P, Buschek D, Huang MX, Bulling A (2019) Reducing calibration drift in mobile eye trackers by exploiting mobile phone usage. In: Proceedings of the 11th ACM symposium on eye tracking research and applications (ETRA'19). NY, USA, pp 1–9. Article 9. https://doi.org/10.1145/3314111.3319918

Münzner S, Schmidt P, Reiss A, Hanselmann M, Stiefelhagen R, Dürichen R (2017) CNN-based sensor fusion techniques for multimodal human activity recognition. In: Proceedings of the 2017 ACM international symposium on wearable computers (ISWC'17). NY, USA, pp 158–165. https://doi.org/10.1145/3123021.3123046

Nam KH, Kim DH, Choi BK, Han IH (2019) Internet of things, digital biomarker, and artificial intelligence in Spine: current and future perspectives. Neurospine 16(4):705–711. https://doi.org/10.14245/ns.1938388.194

Nashed NN, Lahoud C, Abel M-H, Andres F, Blancan B (2021) Mood detection ontology integration with teacher context. In: 20th IEEE conference ICMLA2021, pp 1710–1715. https://doi.org/10.1109/ICMLA52953.2021.00272

NeuroSpin (2022) https://joliot.cea.fr/drf/joliot/en/Pages/research_entities/NeuroSpin.aspx. Accessed 27 Oct 2022

Ochoa X, Worsley M (2016) Editorial: augmenting learning analytics with multimodal sensory data. J Learn Anal 3(2):213–219. https://doi.org/10.18608/jla.2016.32.10

Prometheus-X (2022) Data space education and skills (DASES) within the GAIA-X initiative. https://prometheus-x.org/?locale=en. Accessed 27 Oct 2022

Rudra S (2022) What are digital twins and how can higher ed use them? EdTech online magazine. https://edtechmagazine.com/higher/article/2022/03/what-are-digital-twins-and-how-can-higher-ed-use-them. Accessed 27 Oct 2022

Rustici Software (2022) xAPI solved and explained https://xapi.com/. Accessed 27 Oct 2022

Sangu S, Shimokawa T, Tanaka S (2020) Ultracompact eye and pupil tracking device using VCSEL arrays and position sensitive detector. In: Proceedings SPIE 11310, optical architectures for displays and sensing in augmented, virtual, and mixed reality, p 113101F. https://doi.org/10.1117/12.2542593

Schweizer T, Wyss T, Gilgen-Ammann R (2021) Detecting soldiers' fatigue using eye-tracking glasses: practical field applications and research opportunities. Mil Med usab509. https://doi.org/10.1093/milmed/usab509

Siddharth T-PJ, Sejnowski TJ (2022) Utilizing deep learning towards multi-modal bio-sensing and vision-based affective computing. In: IEEE Trans Affect Comput 13(1):96–107. https://doi.org/10.1109/TAFFC.2019.2916015

Siegfried R, Odobez J-M (2022) Robust unsupervised Gaze calibration using conversation and manipulation attention priors. ACM Trans Multimedia Comput Commun Appl 18(1):27. Article 20. https://doi.org/10.1145/3472622

Šimić G, Maksimović Z, Jevremović A (2019) xAPI – New eLearning standard for LMS–simulations integration. In: Konjović Z, Zdravković M, Trajanović M (ed) ICIST 2019 proceedings, pp 33–36

Spikol D, Ruffaldi E, Dabisias G, Cukurova M (2018) Supervised machine learning in multimodal learning analytics for estimating success in project-based learning. J Comput Assist Learn 34:366–377. https://doi.org/10.1111/jcal.12263

Sugano S, Bulling A (2015) Self-calibrating head-mounted eye trackers using egocentric visual saliency. In: Proceedings of the 28th annual ACM symposium on user interface software & technology (UIST '15). NY, USA, pp 363–372. https://doi.org/10.1145/2807442.2807445

Sulema Y, Kerre E (2020) Multimodal data representation and processing based on algebraic system of aggregates. In: Chakraverty S (ed) Book mathematical methods in interdisciplinary sciences. Wiley, USA, p 464. https://doi.org/10.1002/9781119585640

Tao Z, Xu G (2022) Digital twin technology in the field of education—take the management of the HTC vive as an example. In: Resilience and future of smart learning. ICSLE 2022 (Lecture notes in educational technology). Springer, Singapore. https://doi.org/10.1007/978-981-19-5967-7_6

University of Edinburgh (2022) SR research EyeLink 1000 eye tracker, upgraded to 2KHz, with Tower Mount (monocular). https://www.ed.ac.uk/ppls/psychology/research/facilities/eye-tracking-analysis-laboratories. Accessed 27 Oct 2022

Wang H, Pi J, Qin T, Shen S, Shi BE (2018) SLAM-based localization of 3D gaze using a mobile eye tracker. In: Proceedings of the 2018 ACM symposium on eye tracking research and applications (ETRA '18). NY, USA, pp 1–5 Article 65. https://doi.org/10.1145/3204493.3204584