FEDWELL: Life-Long Federated User and Mental Modeling for Health and Well-being

Sabine Janzen[1,*], Prajvi Saxena[1,*], Cicy Agnes[1], Muhammad Ebad Ullah Khan[1], Amr Gomaa[1], Michael Feld[1], Andre Zenner[1], Pascal Lessel[1], Julian Wolter[1], Florian Daiber[1], Rafael Math[1], Niko Kleer[1], Tim Schwartz[1], Antonio Krueger[1,2] and Wolfgang Maass[1,2]

[1]*German Research Center for Artificial Intelligence (DFKI)*
[2]*Saarland University, Germany*

Abstract

Adaptive and personalized AI systems in healthcare rely on user-specific and contextual information to provide support. However, incomplete, unreliable, and outdated data prevents both patients experiencing illness, pain, or cognitive impairment, as well as therapists, in making proper and informed decisions. Patients specifically may not have the knowledge to comprehend complex medical information, or effectively communicate symptoms. AI-driven mental models and user models can bridge these cognitive gaps, ensuring personalized and effective patient care. The FedWell research project (09/2023–08/2026), funded by the Federal Ministry of Education and Research (BMBF), explores the integration of artificial mental models (AMMs) and user models from various sources into adaptive AI systems to assist patients in decision-making. The project focuses on two key applications: rehabilitation support after knee/hip surgery and treatment decision assistance for patients with cognitive impairments (e.g., multiple sclerosis, dementia). FedWell employs a combination of structured surveys, contextual data collection, and AI techniques to model patient behavior, attitudes, and intentions. A decision support system MENTALYTICS is developed from fine-tuned large language models (LLaMA-2, LLaMA-3, Mistral, Phi-3), that employs AMMs. By the end of the project, FedWell aims to deliver robust AMMs capable of representing patient beliefs and decision-making processes, ultimately guiding them toward treatment options that best fit their individual needs.

Keywords

Personalized healthcare, Adaptive AI-system, Artificial Mental Model, Decision support system, Rehabilitation

1. Introduction

In the field of prevention and rehabilitation, adaptive and personalized systems based on artificial intelligence (AI) methods can play a decisive role as supporters of patient-centered care [1]. In 2023, 1,886,876 patients with an average length of stay of 25.5 days were treated in a preventive care or rehabilitation facility in Germany. The majority of full inpatients (approx. 550,000) were treated for musculoskeletal diseases and injuries, e.g., osteoarthritis. The success of medical rehabilitation depends crucially on good cooperation between the patient, doctor and physiotherapist as well as the patient's willingness to actively participate in a structured, often painful program [2]. However, patients exhibit cognitive limitations in situations characterized by illness, pain and medical decisions. As a result, informed decision-making, understanding of complex medical issues and effective articulation of symptoms and individual concerns are not always possible [3, 4, 5, 6]. Up to 30% of patients in rehabilitation facilities can be characterized as linguistically impaired; i.e. they are unable to express themselves appropriately due to the aforementioned cognitive limitations, migration backgrounds or psychological problems [7, 8, 9, 10]. In addition, patients often experience conflicting goals during rehabilitation that are not immediately visible to doctors, physiotherapists and psychologists. On the one hand, there is a desire to complete rehabilitation quickly and successfully; on the other hand, there are conflicting motives, such as the perceived relief in rehab in case of excessive demands at home [11, 12]. One potential approach is the use of artificial mental models in therapeutic contexts and associated

*Corresponding author.
sabine.janzen@dfki.de (S. Janzen); prajvi.saxena@dfki.de (P. Saxena)

AI systems to improve personalized patient care and therapy outcomes [13, 14, 15]. FEDWELL is an ongoing research project funded by the German Federal Ministry of Education and Research (BMBF) (09/2023 - 08/2026) that investigates federated personalized user modeling (UM) and artificial mental models (AMMs) in adaptive AI systems to support patients for making informed decisions under risk conditions. Recognizing the cognitive impairments of patients in healthcare situation as described, we will design, develop, and evaluate AMM-powered AI systems tailored for both personal and professional applications, ensuring respectful treatment tailored to the user's behavior and meeting unique needs and challenges faced by the involved patients. This includes the evaluation of the adoption of such systems by users, i.e., patients, therapists and medical doctors, including the validation of their impact. The project consortium consists of partners from research and healthcare that take different roles in the project: end users from healthcare (rehabilitation hospital BG Klinikum Duisburg) and research & development (German Research Center for Artificial Intelligence (DFKI)). FEDWELL investigates the application of AMM-powered AI systems in two primary use cases: post-knee/hip surgery rehabilitation support and therapy decision support for patients with compromised decision-making abilities (e.g., multiple sclerosis, dementia).

In this paper, we present an overview of the FEDWELL project, including its objectives, work packages and expected outcomes. We will discuss the role of information systems in FEDWELL, especially focusing on the current state of work and first results in form of a decision support system in the domain of rehabilitation.

2. Project Objectives

Objective of the FEDWELL project is to design and develop adaptive, personalized AI systems that enhance decision-making by patients, therapists and medical doctors by leveraging UM and AMMs. FEDWELL seeks to address the cognitive limitations and communication challenges faced by patients — particularly those recovering from orthopedic surgeries or suffering from cognitive impairments — by constructing AI systems that can understand, anticipate, and adapt to users' mental states, behaviors, and individual preferences. By integrating real-time feedback and contextual patient data into UM and AMMs, FEDWELL aims to enable continuous personalization and update of the models. With focus on two use cases — post knee/hip surgery rehabilitation and therapy decision assistance for cognitively impaired patients — the project aspires to deliver AI-powered systems that foster patient autonomy and support more informed treatment choices.

The term mental model stems from cognitive science. Mental models are cognitive frameworks that people use to understand and navigate their environment, i.e., the world that surrounds them [16, 17, 18, 19]. A patient's mental model reflects their assumptions about the target system, i.e., the environment they interact with, e.g., assumptions about their therapy and rehabilitation progress [20, 17, 21]. Since the patient's mental model is implicit and therefore unknown, the approach of AMMs is to create a conceptual model of the patient's mental model that anticipates it in form of a meta-representation. Cognitive mental models are conceptually similar to world models in AI [22]. These refer to the internal representation of a system's environment that it uses to understand, predict and interact with the world around it. Essentially, a world model allows an AI system to simulate possible outcomes of its actions, anticipate changes and adapt its behavior accordingly. Both mental models and world models serve the purpose of internally simulating the outside world to enable adaptive and goal-oriented behavior. However, mental models are formed through a combination of congenital mechanisms, learning and social influences, while world models are explicitly designed and trained through data and algorithms. AMMs can be understood as world models that anticipate unknown mental models of patients. The use of AMM in rehabilitation can promote the effectiveness of therapies, support improved decision-making processes and help to identify and correct knowledge gaps and misconceptions of patients [18]. Existing research emphasizes the need to accurately capture and understand mental models, especially in therapy and rehabilitation scenarios [23, 24, 20, 25], which are essential but challenging after knee injury or surgery. Even minor restrictions have a significant impact on mobility. After injuries such as cruciate ligament tears or meniscus damage, patients must gradually regain strength, mobility and stability [26]. One of the biggest challenges is the slow progress. Frustration arises when freedom from pain or full mobility is not achieved immediately. The right balance between strain and rest is crucial: too much strain can interfere with healing, too much rest leads to stiffness or muscle weakness [27]. Pain can also reduce motivation, causing patients to avoid important exercises. Psychological factors, such as fear of re-injury, also play a role [28]. An AMM for a patient undergoing knee rehabilitation can capture how she perceives her injury, the recovery process and her environment. The model combines physical, psychological and environmental factors. Physically, it includes the condition of the knee, pain levels and movement restrictions. It recognizes relationships between movement and pain, e.g., that stretching reduces stiffness. Expectations of healing time are also taken into account. Psychologically, the AMM integrates beliefs about the ability to recover, influenced by previous experiences and confidence in the rehabilitation process. Optimistic patients see progress, while discouraged people perceive the therapy as slowed down. Environmental factors such as family support or work commitments also have an influence. The AMM maps the patient's beliefs, behavior and progress. It can therefore predict how they will react to exercises and what fears they have. This enables therapists to develop personalized strategies to increase motivation and optimize rehabilitation success.

The project has four work packages (aside of the ongoing project management that applies at all time). WP1 (Federated User Modeling Concepts and Platform) is about developing an architecture and platform for federated UM. Federated UM refer to systems where user identities are managed across multiple independent systems or nodes, rather than being stored and processed by a single service. It will serve as basis for the use cases of the project, although the specified architecture is generalizable and applicable beyond the scope of the domains of interest in the project. Results of this phase will be introduced in section 3.1. WP2 (User and Mental Model Design) intends to specify and deploy AMMs. Here, a research design based on a Design Science approach was specified for the investigation of AMM in AI systems in the healthcare sector [29]. The research design comprises four iterative phases: Elicitation, Individualization, Action and Transfer. Aim of the elicitation phase was to generate a non-discriminatory and bias-free, domain-specific basic model of an AMM in the field of knee rehabilitation. The AMM was trained with LLaMA-2 (7B), LLaMA-3 (8B, 70B), GPT-4.o-mini, Phi-3 and Mistral (7B) in a two-stage approach combining systematic data scraping and an empirical user study (n=116). The evaluation of the AMM included classification tasks, such as predicting whether a patient can perform an exercise and predicting her expected pain or effort. Further evaluation of the AMM focused on minimizing bias, ensuring demographic fairness in predictions, and using confidence probabilities to increase the reliability of predictions. Results of this phase will be introduced in section 3.2. Based on the results of WP1 and WP2, in WP3 (Multimodal Interface Presentation), a multimodal interface for end users, means patients, therapists and medical doctors will be developed. Finally, the objective of WP4 (Feedback-based Adaptation) is to measure the user adoption of the specified, adaptive AI-system supported by AMMs and UM.

3. Current Status and Intermediate Results

3.1. Federated User Model Platform Architecture (WP1)

Existing literature highlights systems similar to our Federated User Model Platform, such as MYRROR [30], MyrrorBot [31], and HELENA [32]. MYRROR is a holistic UM system that merges data from social networks, smartphones, and FitBit devices using NLP and machine learning to infer user characteristics. However, it lacks incremental model adaptation and does not distinguish between a user's AMM and the system-sided user model. MyrrorBot acts as a Personal Digital Assistant for personalized online services; but relies on heuristics for UM, limiting adaptive machine learning-based approaches. HELENA focuses on lifelong health UM, but does not account for outdated information. The proposed architecture aims to address these limitations by enabling adaptable, continually learning components that interact through RESTful APIs to update information and models Figure 1. The

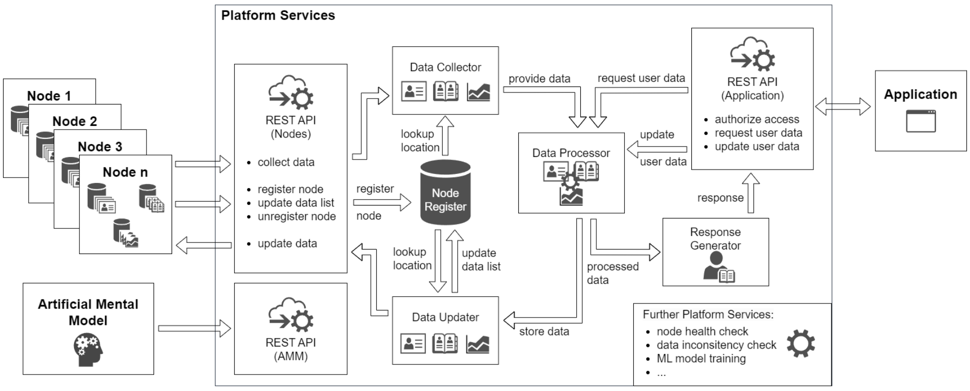

Figure 1: Architecture overview of the federated user model platform. Each node is connected to different sensor or data sources with its own privacy and security protection in a federated manner.

architecture consists of several distributed nodes and shared platform services. Nodes are connectors to various data sources (e.g., dialogue system, user's smartphone, devices for measuring physiological parameters, social media accounts, etc.), which continuously collect data about different users, annotate them (e.g., patient's name, location, time, confidentiality, etc.), and store them in their local storage. Out of this input (symbolic data, sub-symbolic data, ML models), the platform will generate patient-specific UM, that can be queried by an adaptive system in order to generate personalized output. While data is collected and updated continuously, several platform services (e.g., data inconsistency checks, re-training of ML models, etc.) need to be executed regularly to enable up-to-date query results. Each node stores the data of the connected sources locally while providing separated repositories for raw data (e.g., generated by a sensor), ontological data, and ML models. The platform services can address individual data sets of each connected node through a REST API by using a URL to read and write their values.

3.2. Decision Support System MENTALYTICS (WP2)

As mentioned before, objective of WP2 was the design of AMMs. Therefore, a research design based on a Design Science approach was implemented that consisted of four phases. Aim of the first phase - elicitation was the generation of a non-discriminatory and bias-free, domain-specific basis model of an AMM in the field of knee rehabilitation. As such a model requires large amounts of data in order to identify patterns, correlations and variations in the inputs and to develop generalization capability, this phase included indirect observation of patients to create a large data set by means of a two-stage approach consisting of systematic data scraping and an empirical user study (n=116). The data scraping strategy was implemented for indirect observation, targeting conversations between patients, physiotherapists, and doctors, along with curated articles on relevant physiotherapy topics. This yielded 67,000 conversations and 7,000 articles, with 65% discussing knee/leg physiotherapy, 15% specific methods, and 21% news topics. After preprocessing and filtering using Named Entity Recognition and clustering [33], a focused dataset (n=4364) was derived, comprising 2,321 conversations and 2,043 articles categorized into four clusters: therapies, clinical picture, progression, and diagnostics. The second part of the indirect observation consisted of conducting an empirical study with 116 participants. The aim of the study was to compare the effort expected and actually perceived by the participants during sports exercises with their personality traits, medical history and psychosocial factors. Participants completed a digital questionnaire and performed video-guided exercises (e.g., squats, calf raises) before rating their perceived effort and pain, allowing us to compare expected and actual exertion. Data were collected using Likert, interval, and nominal scales covering demographics, psychosocial and physical activity factors, medical history, personality traits (TIPI) [34], and pain ratings (NRS) [35]. Notably, 46.6% had prior surgeries, with only 27% fully adhering to rehabilitation plans. A 40% gap was observed between expected and perceived exertion. For the training of the basis model with

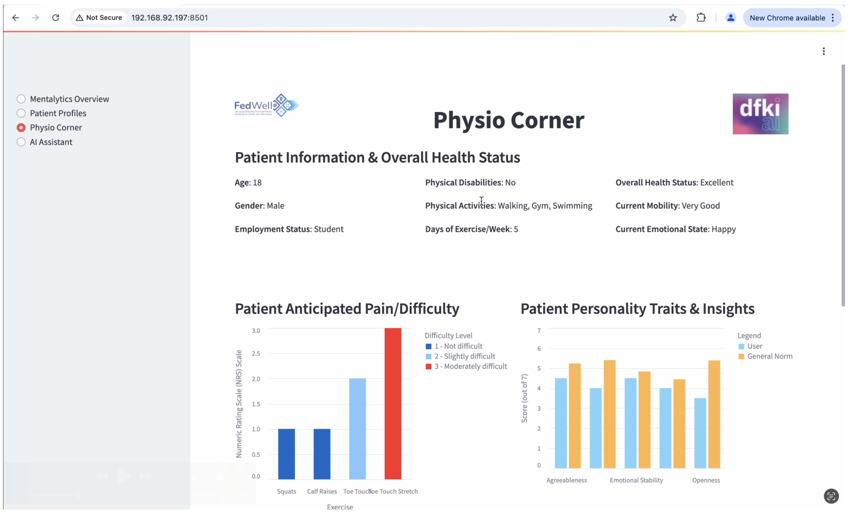

Figure 2: Interface of MENTALYTICS decision support system for therapists showing real-time predictions of anticipated pain scores, personality traits and key insights for patients.

the indirect observation data, a multi-stage process consisting of (1) selection of pre-trained machine learning models, (2) evaluation of the performance of the selected models in combination with the given data, and (3) fine-tuning of the models was carried out. Large, pre-trained large language models (LLMs) were selected for training: LLaMA-2 (7B) [36], LLaMA-3 (8B, 70B) [37], GPT-4.o-mini [38], Phi-3 [39], Mistral (7B) [40]. The evaluation of the AMM included classification tasks, such as predicting whether a patient could perform an exercise and predicting her expected pain or exertion. Further evaluation of the AMM focused on minimizing bias, ensuring demographic fairness in the predictions, and using confidence probabilities to increase the reliability of the predictions [41, 42, 43, 44]. The resulting basis AMM model is represented by the pre-trained LLaMA-3 (8B) model after fine-tuning, which was adjusted for bias and hallucinations and achieved 94% accuracy in predicting pain scores for specific patients. Thus, it outperformed larger models such as LLaMA-3 (70B), which showed significant performance degradation due to overfitting and reduced generalization. The results show that LLMs can serve as effective basis models for AMMs. The resulting AMM basis model was integrated into the decision support system - MENTALYTICS - a tool for medical doctors and therapists in knee rehabilitation that provides predictions of patients' expectations regarding the expected pain and effort of physical exercises in their rehabilitation[1]. Additionally, the system features a conversational AI assistant, enabling doctors and physiotherapist to gain further insights and explainations through interactive discussions.

4. Relevance of Project for CAISE

The research project FEDWELL is relevant to the International Conference on Advanced Information Systems Engineering (CAiSE) due to its innovative approach to integrate advanced UM techniques and AMM into the design of personalized information systems in the healthcare domain. By focusing on federated UM and AMMs, FEDWELL contributes to the growing body of research that explores context-aware, adaptive, and user-centric information systems. FEDWELL addresses complex challenges in medical rehabilitation where user requirements are diverse, decision-making is impaired, and therapeutic processes are sensitive to individual behavior and motivation. The project exemplifies how advanced engineering of information systems can bridge the gap between technical sophistication and human-centered design, especially when supporting vulnerable user groups under high-risk, emotionally charged, and ethically complex conditions. Furthermore, the project's commitment to federated learning paradigms and privacy-preserving data processing reflects critical advancements in system architecture

[1]Screencast of MENTALYTICS: https://youtu.be/c6eKn5Yk16A?feature=shared

and distributed AI, aligning with CAiSE's interest in secure, trustworthy, and decentralized information systems. By designing AMM-powered systems based on a Design Science Approach [45] and validating their impact in real-world rehabilitation contexts, FEDWELL contributes valuable insights into system evaluation methodologies, user adoption, and socio-technical integration.

5. Conclusion and Future Work

The FEDWELL project addresses a critical gap in rehabilitation by developing adaptive, AI-driven systems that account for patients' cognitive limitations, linguistic impairments, and conflicting therapeutic goals. By leveraging federated personalized user modeling and artificial mental models, FEDWELL aims to support more informed, individualized, and ethically aware decision-making processes in medical contexts. Initial applications focus on post-surgical rehabilitation and therapy support for patients with compromised decision-making capacity, with active involvement from both clinical and research stakeholders. In future work, we will further refine AMM-supported decision support systems, expand their applicability to broader patient groups in clinical settings, and rigorously evaluate their impact on therapy outcomes, user experience, and interdisciplinary cooperation in healthcare settings. This will also include the integration of further heterogeneous data sources, e.g., sensor data, patient-reported outcomes, and clinical observations, to improve the model's robustness and responsiveness. Emphasis will also be placed on further developing explainability and transparency into system design to foster trust and acceptance among patients and professionals.

Acknowledgments

This work was partially funded by the German Federal Ministry of Education and Research (BMBF) under the contract 01IW23004.

Declaration on Generative AI

During the preparation of this work, the author(s) used GPT-4 and Grammarly in order to: Grammar and spelling check. In addition, large language models such as LLaMA-2, LLaMA-3, Mistral, and Phi-3 were employed for fine-tuning and experimentation purposes within the scope of this study. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] R. Polacsek-Ernst, F. Benda, J. Bretschneider, C. Pagel, Digitales mindset im gesundheitswesen–chancen und herausforderungen digitaler technologien, in: Technologien und Technologiemanagement im Gesundheitswesen: Potenziale nutzen, Lösungen entwickeln, Ziele erreichen, Springer, 2024, pp. 89–111.

[2] C. Valle, Rehabilitation in der knieendoprothetik, Die Orthopädie 53 (2024) 817–818.

[3] T. C. Dildine, C. M. Amir, J. Parsons, L. Y. Atlas, How pain-related facial expressions are evaluated in relation to gender, race, and emotion, Affective Science 4 (2023) 350–369.

[4] T. C. Dildine, E. A. Necka, L. Y. Atlas, Confidence in subjective pain is predicted by reaction time during decision making, Scientific reports 10 (2020) 21373.

[5] A. N. Meyer, T. D. Giardina, L. Khawaja, H. Singh, Patient and clinician experiences of uncertainty in the diagnostic process: current understanding and future directions, Patient Education and Counseling 104 (2021) 2606–2615.

[6] A. Timm, T. Schmidt-Wilcke, S. Blenk, B. Studer, Altered social decision making in patients with chronic pain, Psychological Medicine 53 (2023) 2466–2475.

[7] D. Rost, et al., Verena dreißig (2005). interkulturelle kommunikation im krankenhaus. eine studie zur interaktion zwischen klinikpersonal und patienten mit migrationshintergrund, in: Forum Qualitative Sozialforschung, volume 8, 2007.
[8] M. P. S. Oberholz, General practitioner care for people with a migration background in Germany, Dissertation, Rheinische Friedrich-Wilhelms-University Bonn, 2024.
[9] S. Tagay, Kommunikation–andere länder, andere sitten–patienten mit migrationshintergrund verstehen, Aktuelle Urologie 47 (2016) 128–130.
[10] S. Ullrich, D. Briel, Y. Nesterko, A. Hiemisch, E. Brähler, H. Glaesmer, Verständigung mit patienten und eltern mit migrationshintergrund in der stationären allgemeinpädiatrischen versorgung, Das Gesundheitswesen 78 (2016) 209–214.
[11] I. Buchholz, T. Kohlmann, Deskriptoren und prädiktoren von reha-zielen von patienten in der medizinischen rehabilitation, PPmP-Psychotherapie· Psychosomatik· Medizinische Psychologie 64 (2014) 364–372.
[12] T. Merten, Beschwerdenvalidierung in der Begutachtung, Klinik und Rehabilitation, Hogrefe Verlag GmbH & Company KG, 2023.
[13] L. A. C. L. Bollos, Y. Zhao, G. P. Soriano, T. Tanioka, H. Otsuka, R. Locsin, Technologies, physician's caring competency, and patient centered care: a systematic review, The Journal of Medical Investigation 70 (2023) 307–316.
[14] R. Khera, A. J. Butte, M. Berkwits, Y. Hswen, A. Flanagin, H. Park, G. Curfman, K. Bibbins-Domingo, Ai in medicine—jama's focus on clinical outcomes, patient-centered care, quality, and equity, Jama 330 (2023) 818–820.
[15] T. T. Wan, H. S. Wan, Predictive analytics with a transdisciplinary framework in promoting patient-centric care of polychronic conditions: Trends, challenges, and solutions, AI 4 (2023) 482–490.
[16] P. N. Johnson-Laird, The history of mental models, in: Psychology of reasoning, Psychology Press, 2004, pp. 189–222.
[17] P. N. Johnson-Laird, G. P. Goodwin, S. S. Khemlani, Mental models and reasoning, in: International handbook of thinking and reasoning, Routledge, 2017, pp. 346–365.
[18] N. A. Jones, H. Ross, T. Lynam, P. Perez, A. Leitch, Mental models: an interdisciplinary synthesis of theory and methods, Ecology and society 16 (2011).
[19] D. A. Norman, Some observations on mental models, in: Mental models, Psychology Press, 2014, pp. 7–14.
[20] A. Hermans, S. Muhammad, J. Treur, You feel so familiar, you feel so different: A controlled adaptive network model for attachment patterns as adaptive mental models, in: Mental Models and Their Dynamics, Adaptation, and Control: A Self-Modeling Network Modeling Approach, Springer, 2022, pp. 321–346.
[21] K. LaMere, S. Mäntyniemi, J. Vanhatalo, P. Haapasaari, Making the most of mental models: Advancing the methodology for mental model elicitation and documentation with expert stakeholders, Environmental Modelling & Software 124 (2020) 104589.
[22] D. Ha, J. Schmidhuber, Recurrent world models facilitate policy evolution, Advances in neural information processing systems 31 (2018).
[23] T. Barber, K. Crick, L. Toon, J. Tate, K. Kelm, K. Novak, R. O. Yeung, P. Tandon, D. C. Sadowski, S. Veldhuyzen van Zanten, et al., Gastroscopy for dyspepsia: Understanding primary care and gastroenterologist mental models of practice: A cognitive task analysis approach, Journal of the Canadian Association of Gastroenterology 6 (2023) 234–243.
[24] M. Gabbas, K. Kim, Gamified user interface design for dysphagia rehabilitation based on common mental models (2022).
[25] A. D. Naik, Collaborative decision-making: Identifying and aligning care with the health priorities of older adults, in: Geriatric Medicine: A Person Centered Evidence Based Approach, Springer, 2023, pp. 1–21.
[26] E. Klein, K. Stamer, K. Beitzel, A. B. Imhoff, Knie: Rehabilitation, in: Rehabilitation in der orthopädischen Chirurgie: OP-Verfahren im Überblick-Physiotherapie-Sporttherapie, Springer,

2024, pp. 147–182.
[27] B. Kladny, Rehabilitation nach knieendoprothetischer versorgung., Der Orthopäde 50 (2021).
[28] M. Kunnen, R. A. Dionigi, C. Litchfield, A. Moreland, Psychological barriers negotiated by athletes returning to soccer (football) after anterior cruciate ligament reconstructive surgery, Annals of Leisure Research 26 (2023) 545–566.
[29] S. Janzen, W. Maaß, P. Saxena, Investigation of artificial mental models for healthcare ai systems, in: 19th International Conference on Design Science Research in Information Systems and Technology (DESRIST-2024), University West, Springer, Högskolan Väst 461 86 Trollhättan, 2024.
[30] C. Musto, M. Polignano, G. Semeraro, M. de Gemmis, P. Lops, Myrror: a platform for holistic user modeling: Merging data from social networks, smartphones and wearable devices, User Modeling and User-Adapted Interaction 30 (2020) 477–511.
[31] C. Musto, F. Narducci, M. Polignano, M. De Gemmis, P. Lops, G. Semeraro, Myrrorbot: A digital assistant based on holistic user models for personalized access to online services, ACM Transactions on Information Systems (TOIS) 39 (2021) 1–34.
[32] M. Polignano, P. Lops, M. de Gemmis, G. Semeraro, Helena: An intelligent digital assistant based on a lifelong health user model, Information processing & management 60 (2023) 103124.
[33] A. Fantechi, S. Gnesi, S. Livi, L. Semini, A spacy-based tool for extracting variability from nl requirements, in: Proceedings of the 25th ACM International Systems and Software Product Line Conference-Volume B, 2021, pp. 32–35.
[34] S. D. Gosling, P. J. Rentfrow, W. B. Swann Jr, A very brief measure of the big-five personality domains, Journal of Research in personality 37 (2003) 504–528.
[35] S. M. Nugent, T. I. Lovejoy, S. Shull, S. K. Dobscha, B. J. Morasco, Associations of pain numeric rating scale scores collected during usual care with research administered patient reported pain outcomes, Pain Medicine 22 (2021) 2235–2241.
[36] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale, et al., Llama 2: Open foundation and fine-tuned chat models, arXiv preprint arXiv:2307.09288 (2023).
[37] A. Grattafiori, A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle, A. Letman, A. Mathur, A. Schelten, A. Vaughan, et al., The llama 3 herd of models, arXiv preprint arXiv:2407.21783 (2024).
[38] A. Hurst, A. Lerer, A. P. Goucher, A. Perelman, A. Ramesh, A. Clark, A. Ostrow, A. Welihinda, A. Hayes, A. Radford, et al., Gpt-4o system card, arXiv preprint arXiv:2410.21276 (2024).
[39] M. Abdin, J. Aneja, H. Awadalla, A. Awadallah, A. A. Awan, N. Bach, A. Bahree, A. Bakhtiari, J. Bao, H. Behl, et al., Phi-3 technical report: A highly capable language model locally on your phone, arXiv preprint arXiv:2404.14219 (2024).
[40] G. Jung, M. A. Hiltunen, K. R. Joshi, R. D. Schlichting, C. Pu, Mistral: Dynamically managing power, performance, and adaptation cost in cloud infrastructures, in: 2010 IEEE 30th International Conference on Distributed Computing Systems, IEEE, 2010, pp. 62–73.
[41] Z. Lin, S. Trivedi, J. Sun, Contextualized sequence likelihood: Enhanced confidence scores for natural language generation, in: Y. Al-Onaizan, M. Bansal, Y.-N. Chen (Eds.), Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Miami, Florida, USA, 2024, pp. 10351–10368.
[42] S. Desai, G. Durrett, Calibration of pre-trained transformers, in: B. Webber, T. Cohn, Y. He, Y. Liu (Eds.), Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Online, 2020, pp. 295–302.
[43] T. Calders, F. Kamiran, M. Pechenizkiy, Building classifiers with independency constraints, in: 2009 IEEE International Conference on Data Mining Workshops, 2009, pp. 13–18.
[44] M. Yuan, V. Kumar, M. A. Ahmad, A. Teredesai, Assessing fairness in classification parity of machine learning models in healthcare, 2021. arXiv:2102.03717.
[45] K. Peffers, T. Tuunanen, B. Niehaves, Design science research genres: introduction to the special issue on exemplars and criteria for applicable design science research, 2018.