Privacy-Preserving Edge AI for Rehabilitation Support: Lessons from a Field Study

Sabine Janzen[1], Prajvi Saxena[1], Muhammad Ebad U. Khan[2] und Wolfgang Maaß[1,2]

Abstract: AI-supported rehabilitation promises more individualized therapy by accounting for subjective factors such as perceived difficulty, expectations, motivation, fear of pain, and situational constraints. However, such systems must be evaluated not only for predictive validity, but also for privacy-preserving deployment feasibility. We report a one-day public field study of an Artificial Mental Model (AMM)-based rehabilitation support system with 220 valid study sessions. Participants completed a short profile, received an AI-based prediction of perceived exercise difficulty, performed a standardized sit-to-stand exercise, and reported their actual difficulty. The system was deployed locally on a constrained edge setup without cloud processing which enabled the study to be conducted without relying on cloud services for sensitive health data. Results show operationally robust no-cloud deployment, while predictive validity remained below pre-specified targets. Agreement responses provided a plausibility check, as participants were more likely to agree when predictions were close to their post-exercise ratings. The study highlights the need for context-sensitive calibration, human-centered evaluation, and low-latency privacy-preserving deployment.

Keywords: AI-supported rehabilitation, Artificial Mental Models, Edge AI, privacy-preserving AI, field study, human-centered evaluation

1 Introduction

AI-based rehabilitation support promises more individualized therapy by accounting for factors that are difficult to observe in routine care, such as perceived difficulty, expectations, motivation, fear of pain, prior experience, and situational constraints. Recent reviews similarly highlight the potential of AI for personalized rehabilitation, movement assessment, adaptive therapy support, and patient monitoring [At25; Ni25]. These factors influence how patients understand, perform, and evaluate rehabilitation exercises, and they are central to patient-centered AI in medicine [Kh23]. Building on mental model research, which describes how people interpret situations, anticipate consequences, and guide action [JGK17; Jo89; No87], prior work introduced Artificial Mental Models (AMMs) as computational meta-representations of patients' mental models for healthcare AI systems [JMS24; JSA+25]. In rehabilitation, AMMs aim to anticipate how patients may perceive exercises, for example in terms of expected effort, pain, or feasibility, and thereby support more individualized therapy decisions.

While previous work has evaluated AMM-based prediction in controlled settings, the present study is, to our knowledge, the first to assess this approach under real-world public field

[1] German Research Center for Artificial Intelligence (DFKI), sabine.janzen@dfki.de
[2] Saarland University,

conditions. Real-world evaluation of AMM-based rehabilitation support must address two challenges at once. First, perceived difficulty is inherently subjective and commonly assessed through rating scales such as Borg-type scales or numerical rating scales [Bo82; Mo18; St20; Wi17]. These measures are pragmatic, but sensitive to task design, interpretation, social context, and self-presentation; public field settings may further amplify these effects through distraction, unfamiliarity, challenge behavior, or interaction framing. Second, the data involved are health-related and sensitive. Therefore, privacy-preserving deployment must be feasible in practice. Local edge inference can avoid cloud processing and keep sensitive data close to the point of interaction, but must still satisfy requirements regarding latency, concurrency, reliability, usability, data minimization, privacy, and security [Al24; POS24; RAC24]. From a design science perspective, field evaluation is therefore needed to assess both predictive validity and deployment feasibility in the intended use context [He04; Pe07].

In this paper, we report a real-world field study of an AMM-based system for AI-supported rehabilitation at German Unity Day 2025 in Saarbruecken (Germany). Participants interacted with a locally deployed, privacy-preserving application via QR code using personal smartphones or study devices (tablets, smartphone). After informed consent and safety triage, they provided a short profile, performed a standardized sit-to-stand exercise (10 repetitions), received an AI-based prediction, and reported their perceived difficulty. The study evaluates two aspects: (1) the predictive validity of AMMs for perceived difficulty during a short standardized exercise, and (2) the feasibility of local edge deployment in a public, high-throughput setting without cloud processing. The results show that privacy-preserving local deployment was operationally feasible, while predictive validity remained below the pre-specified targets. The findings reveal a translation gap between AMM-based personalization and real-world rehabilitation support, particularly regarding calibration, task design, and user interpretation.

The contribution of this paper is threefold: (1) empirical evidence from a public field evaluation of AI-based rehabilitation support with 220 study sessions; (2) an integrated analysis of predictive validity and privacy-preserving edge feasibility; and (3) design lessons for AI-based health support systems, highlighting the need for context-sensitive calibration, human-centered outcome measures, and deployment-aware evaluation beyond model accuracy alone.

2 Study Design and Deployment

We conducted a one-day public field study during a public science outreach event in Germany to evaluate an AMM-based system for AI-supported rehabilitation under real-world interaction and deployment conditions. The study was designed as a low-risk, high-throughput evaluation in a public setting rather than as a clinical trial. Its purpose was to assess whether the system can generate meaningful individualized predictions for a short

standardized exercise and whether privacy-preserving local deployment is feasible under realistic conditions.

2.1 Setting and Participants

Participants were recruited on site at a public exhibition booth. Eligible participants were adults aged 18 years or older who were able to provide informed consent and pass a brief safety triage. The triage was used to exclude participants for whom the exercise could be unsafe, for example due to acute pain, dizziness, medical exercise restrictions, or elevated fall risk. The final dataset contains 220 valid study sessions. Because shared study devices were used in addition to personal devices, device identifiers were treated as technical session information and not as unique participant identifiers. The study population was intentionally heterogeneous and not limited to rehabilitation patients. Self-reported disability status was treated as a broad participant-reported background variable rather than as a clinically verified diagnosis. In a public booth setting, this variable may capture heterogeneous conditions, including chronic or temporary limitations, perceived functional restrictions, or different interpretations of the questionnaire item. The study design allowed us to evaluate the system under public-use conditions with varied age groups, activity levels, health perceptions, and device usage patterns.

2.2 Study Procedure

The study followed a short interaction flow optimized for public participation. Participants accessed the application by scanning a QR code with their own smartphone or by using a study device, i.e., tablet or mobile phone. After reading the study information and providing consent, they completed a brief safety triage and a short profile questionnaire. The profile included demographic bands and self-reported information relevant for perceived exercise difficulty, such as activity level, sleep, stress, emotional state, health status, mobility, disability status, and prior surgery or rehabilitation experience. Participants then watched a video demonstrating the standardized exercise task before providing an initial self-assessment. The task consisted of ten sit-to-stand repetitions performed under supervision in a designated booth area with stable chairs and anti-slip mats. The AMM-based system generated an individualized prediction of perceived difficulty on a five-point numerical rating scale (1 = very easy, 5 = very difficult). Participants subsequently performed the exercise under supervision and reported their actual perceived difficulty after completion. Task completion was recorded as a binary outcome by the supervising staff. The AMM prediction was generated by the system after participants completed their exercises, and finally, participants indicated whether they agreed with the model prediction. This enabled the analysis of both predictive validity and user agreement as complementary outcomes.

Tab. 1: Local edge deployment setup.

Component	Implementation
Client access	QR-based access via personal smartphones or study devices (3 Microsoft Surface Tablets, 1 Samsung Galaxy Z foldable mobile phone)
Network	Isolated local router without internet access
Application	Locally hosted multilingual web interface (German, English, French) (Streamlit)
Inference host	MacBook Pro 13-inch (2020), 2 GHz quad-core Intel Core i5, Intel Iris Plus Graphics 1536 MB, 16 GB LPDDR4X memory, Python, llama.cpp
Model	Fine-tuned LLaMA 3.1 8B, deployed in 4-bit quantized form
Request handling	Multi-threaded inference queue for concurrent clients
Data handling	Local storage only; no cloud services or external APIs
Privacy measures	Generated session IDs; no names, email addresses, IP addresses, or device fingerprints
Prediction target	Perceived exercise difficulty on a five-point scale (1 = very easy, 5 = very difficult)

2.3 Outcome Measures

The first evaluation objective concerned predictive validity. The primary prediction target was perceived exercise difficulty on a five-point numerical rating scale (1 = very easy, 5 = very difficult). We evaluated prediction quality using exact accuracy, accuracy within ± 1 scale point, and mean absolute error (MAE). In addition, we analyzed task completion as a safety-relevant binary outcome. Since task non-completion was rare, completion prediction was evaluated descriptively rather than treated as a robust classification result.

The second evaluation objective concerned deployment feasibility. We assessed whether the system could be operated locally in a public, high-throughput setting without cloud processing. Feasibility indicators included service availability, concurrent client access, system crashes, latency, and practical usability of the QR-based access flow. These indicators were selected because privacy-preserving AI support in rehabilitation must be deployable not only in controlled laboratory environments, but also in settings with heterogeneous devices, intermittent user flow, and limited technical supervision.

2.4 Local Edge Deployment

The system was deployed as a booth-hosted local edge setup (cf. Table 1). Participants accessed a locally hosted web application via QR code using either their own smartphones or study devices provided at the booth. Client devices acted as thin clients, while inference and data storage were performed on a dedicated edge laptop within an isolated local network without internet access. Thus, the study setup did not rely on cloud services or external APIs during operation. The AMM-based prediction component used a locally deployed,

fine-tuned LLaMA 3.1 8B model in 4-bit quantized form. The model received participants' profile variables together with their pre-exercise self-assessment of expected difficulty as one input among the broader set of profile features, rather than as the primary predictive basis. The underlying LLaMA 3.1 8B model was fine-tuned following the multi-stage training process described [JSA+25]: (1) selection of candidate pre-trained models, (2) evaluation of candidate model performance in combination with the given data, and (3) fine-tuning of the selected model, here using 4-bit quantization for the deployed edge setup. Inference was performed on a MacBook Pro 13-inch (2020) with a 2 GHz quad-core Intel Core i5 processor, Intel Iris Plus Graphics, and 16 GB LPDDR4X memory. The setup therefore represented a constrained CPU-oriented local inference scenario without a dedicated accelerator. All interaction data, model outputs, and technical telemetry were stored locally, using generated session identifiers and avoiding directly identifying information.

2.5 Ethics and Consent

The study was reviewed and approved by the institutional ethics committee under approval number MFS–25/25 (Mentalytics Field Study - Unity Day 2025). The approval required that all interactions with participants and all study procedures be conducted with care, sensitivity, and respect for participants' specific needs. All participants received study information and provided informed consent before taking part. Participation was voluntary, and participants completed a brief safety triage before performing the exercise.

3 Results

The final dataset contains 220 valid study sessions (cf. Table 2). Most sessions were completed in German, with additional sessions in English and French. The sample covered a broad age range from 18 years to 65 years and above, and included participants with different employment statuses, activity levels, and self-reported health backgrounds. As described above, the sample should be interpreted as a heterogeneous public field sample rather than as a clinical rehabilitation cohort. The reuse of study devices resulted in repeated technical device identifiers. Therefore, device identifiers were not used to infer unique participants. All analyses are reported at the level of valid study sessions.

3.1 Predictive Validity

The AMM-based system predicted perceived exercise difficulty on a five-point numerical rating scale (1 = very easy, 5 = very difficult). The results provide a differentiated picture of predictive validity under public field conditions. Exact accuracy was 39.5%, while 69.1% of predictions were within ± 1 scale point of the post-exercise rating. The mean absolute

Tab. 2: Sample characteristics of the field study.

Characteristic	**Distribution**
Valid study sessions	220
Language	German: 183; English: 33; French: 4
Gender	Male: 112; Female: 107; Prefer not to say: 1
Age bands	18–24: 49; 25–34: 54; 35–44: 29; 45–54: 27; 55–64: 34; 65+: 27
Self-reported disability status	No: 164; Yes: 56
Employment status	Employed: 138; Student: 51; Retired: 24; Unemployed: 5; Homemaker/Caregiver: 2
Task completion	Completed correctly: 214; Not completed correctly: 6
Agreement with model prediction	Agree: 117; Disagree: 103

Tab. 3: Predictive validity for perceived exercise difficulty.

Metric	**Result**	**Target**	**Status**
Exact accuracy	39.5%	–	–
Accuracy within ±1 scale point	69.1%	≥85%	Not met
Mean absolute error (MAE)	1.027	≤0.60	Not met
Correctly completed sessions only	69.2% within ±1; MAE = 1.033	≥85%; ≤0.60	Not met

Tab. 4: Distribution of model predictions and actual post-exercise difficulty ratings.

Score	**Meaning**	**Model prediction**	**Actual rating**
1	Very easy	59	148
2	Easy / low difficulty	59	48
3	Moderate difficulty	76	13
4	Difficult	13	4
5	Very difficult	13	7

error was 1.027 scale points. Compared with the pre-specified targets of at least 85% within ±1 and an MAE of at most 0.60, predictive performance remained below the intended threshold.

The error pattern was not random. Rather, the model showed a conservative tendency to predict higher difficulty than participants reported after the exercise. Actual post-exercise ratings were strongly skewed toward the lower end of the scale: 148 sessions were rated as 1, 48 as 2, 13 as 3, 4 as 4, and 7 as 5. In contrast, the model produced fewer very-low predictions and more medium-level predictions. This indicates that the low-risk exercise produced a rating distribution in which most participants experienced the task as easy, while the model remained more cautious in its assessment. Task completion was recorded as a binary supervisor-assessed outcome. Overall, 214 of 220 sessions were completed correctly. The high completion rate confirms the low-risk character of the exercise protocol, but also limits the evaluation of failure prediction. Using a model score threshold of ≥ 4 to predict

Tab. 5: Model prediction vs. pre-exercise self-assessment, constant baseline, and (for the subgroup who changed their assessment) expectation vs. actual outcome.

Comparison	**Exact**	**Within** ±1	**MAE**
Model vs. actual (all, n=220)	39.5%	69.1%	1.027
Self-assessment vs. actual (all, n=220)	60.0%	87.3%	0.545
Constant baseline vs. actual (all, n=220)	67.3%	89.1%	0.518
Model vs. expectation (changed subgroup, n=88)	76.1%	93.2%	0.307
Model vs. actual (changed subgroup, n=88)	5.7%	59.1%	1.443

Tab. 6: Prediction error by user agreement.

Agreement with model	**Sessions**	**MAE**	**Within** ±1
Agree	117	0.359	94.9%
Disagree	103	1.786	39.8%

non-completion resulted in a balanced accuracy of 0.696 and a failure recall of 0.500; using a stricter threshold of ≥ 5 resulted in a balanced accuracy of 0.641 and a failure recall of 0.333. These results should therefore be interpreted descriptively rather than as robust evidence for safety-relevant classification performance.

For reference, we also compared the model against two simple baselines already available from the collected data: participants pre-exercise self-assessment and a constant baseline predicting the modal actual rating. Both outperformed the model (cf. Table 5). A targeted analysis of the 88 participants (40%) whose self-reported difficulty changed after the exercise showed that the model's prediction matched their pre-exercise expectation in 93.2% of cases (±1, MAE = 0.307) but their actual post-exercise rating in only 59.1% of cases (MAE = 1.443), indicating the model tracked stated expectation more closely than experienced outcome.

3.2 User Agreement

Agreement responses served as a complementary plausibility indicator. Overall, 117 sessions indicated agreement with the model prediction, while 103 indicated disagreement. As expected, agreement was associated with numerical proximity between the prediction and the post-exercise rating. In sessions where participants agreed with the model, the MAE was 0.359 and 94.9% of predictions were within ±1 scale point. In sessions where participants disagreed, the MAE was 1.786 and 39.8% of predictions were within ±1 scale point. This result should not be interpreted as independent validation of the model. Rather, agreement indicates whether participants perceived the AI assessment as plausible in relation to their own task experience. For rehabilitation support, such perceived plausibility is relevant because AI-generated recommendations must be understandable and acceptable to patients and healthcare professionals before they can inform therapy decisions.

Tab. 7: Edge feasibility results.

Metric	Result	Target	Status
Service availability	Continuous availability during field operation	≥99% uptime	Met
Concurrent access	7–8 clients at peak	≥5 clients	Met
System crashes	0	0	Met
Latency	Approx. 30–40s typical latency; higher under peak load	P50 ≤1s; P90 ≤2s	Not met
Cloud independence	Local inference and local storage only	No cloud processing	Met

3.3 Edge Feasibility

As supporting infrastructure for the field evaluation, the local edge deployment achieved its core operational goals regarding availability, concurrent access, continuous operation, and cloud independence. While the deployment target was sub-second median latency and a 90th percentile below two seconds, observed inference latency was approximately 30–40 seconds under typical load and increased under peak concurrency. Thus, the deployment results support the feasibility of privacy-preserving local operation, while identifying latency as the main optimization target.

4 Discussion

A central finding of the study is the systematic tendency toward conservative over-prediction. The model often predicted higher difficulty than participants reported after performing the exercise. In a rehabilitation context, conservative predictions may be preferable to unsafe underestimation, since underestimating difficulty could increase the risk of overload, frustration, or adverse events. However, persistent over-prediction may also reduce perceived usefulness, lead to unnecessary caution, or weaken trust if users repeatedly experience tasks as easier than predicted. Since perceived difficulty and exertion are subjective constructs commonly captured through rating scales, prediction errors must be interpreted in relation to task design, scale interpretation, and user context [Bo82; Mo18; Wi17].

This calibration challenge is consistent with issues reported for other deployed clinical prediction models, where systematic bias required explicit recalibration before safe use [Li24], and reinforces the broader argument that AI-based health support must be evaluated against credible baselines and evidence standards rather than accuracy thresholds alone [KHB25]. The resulting design lessons calibrating predictions against credible references, prioritizing human-centered outcome measures, and evaluating deployment feasibility alongside model accuracy substantially confirm established human–AI interaction guidelines [Am19]. These guidelines were developed independently of rehabilitation-specific contexts; the present

study empirically tests them in a real-world health deployment, and the design-science framework underlying the AMM concept itself [JMS24] similarly emphasizes iterative evaluation of an artifact in its intended use context [He04; Pe07] as a precondition for design knowledge which this field study directly addresses.

Future AMM-based systems therefore need calibration strategies that distinguish between safety-oriented caution and excessive overestimation. The public field setting likely contributed to this calibration gap. The standardized exercise was intentionally low-risk and suitable for broad public participation, resulting in a strongly skewed rating distribution toward low perceived difficulty. Moreover, public demonstration settings can create interaction effects that differ from rehabilitation practice. Participants may be motivated to outperform the AI prediction, present themselves as capable, or interpret the task as a challenge rather than as a therapeutic exercise. These effects are not merely sources of noise; they reveal how AI assessments can be socially interpreted and contested in real interaction contexts. From a design science perspective, such observations are valuable because they show how an artifact behaves in its intended or prospective use context and what design knowledge can be derived beyond technical functionality alone [He04; Pe07].

Agreement should therefore not be interpreted as independent validation of the model. Rather, it indicates whether participants perceived the AI assessment as plausible in relation to their own task experience. This is relevant for rehabilitation support because acceptance of AI-based recommendations depends not only on technical accuracy, but also on perceived plausibility, trust, and how users negotiate AI advice in context [CE22; Si23]. This is consistent with the mental model perspective, according to which people interpret situations and system outputs through internal representations that guide understanding and action [JGK17; Jo89; No87].

For the local edge deployment for the study feasibility results further show that privacy-preserving local inference is a viable deployment direction, but model latency remains a major design constraint. The use of a fine-tuned 8B model on a CPU-oriented consumer laptop without a dedicated accelerator enabled no-cloud operation, but resulted in response times that are not yet suitable for time-sensitive rehabilitation interaction. This reflects a broader challenge in edge AI for healthcare. Local processing can support privacy and data minimization, but must be balanced against constraints in computation, latency, reliability, usability, and security [Al24; POS24; RAC24]. Future implementations should therefore explore smaller specialized models, distillation, hardware acceleration, and interaction designs that make waiting time acceptable or move inference to moments where immediate feedback is not required.

Overall, the study highlights a translation gap between AI-based personalization concepts and real-world rehabilitation support. This gap is not only technical, but also methodological and interactional. Future evaluations should include more clinically representative participant groups, adaptive task difficulty, stronger ground truth measures, and longitudinal settings in which AMMs can learn from repeated patient feedback.

5 Limitations and Implications

This study has several limitations. First, it was conducted in a public exhibition setting rather than in a clinical rehabilitation environment. The heterogeneous sample supports insights into public-use feasibility and field behavior, but it should not be interpreted as a representative rehabilitation patient cohort or as evidence of clinical effectiveness. Second, self-reported disability status was collected as a broad background variable and not clinically verified. The relatively high share of sessions indicating disability status may therefore reflect heterogeneous conditions, including chronic disability, temporary impairment, perceived functional limitations, or different interpretations of the item.
Third, the standardized exercise was selected to be safe, short, and suitable for broad public participation, which resulted in a strongly skewed rating distribution toward low perceived difficulty and only six cases of incorrect task completion. Fourth, the main outcome relied on self-reported perceived difficulty, which is sensitive to scale interpretation, self-presentation, interaction framing, and situational context. Future studies should therefore combine self-reports with therapist ratings, movement quality assessment, physiological indicators, repeated measures, and clinical exercise protocols.

These limitations lead to several implications for the design and evaluation of AI-supported rehabilitation systems. Predictive validity should be evaluated together with user-centered indicators such as perceived plausibility, acceptance, and understandability. In this study, agreement responses mainly served as a plausibility check; participants were more likely to agree when predictions were close to their post-exercise ratings. Future studies should therefore assess perceived plausibility more directly, for example through structured usability, trust, and acceptance measures. The edge deployment results also depend on the specific technical setup. The constrained local configuration demonstrated that no-cloud operation is feasible under public field conditions, but the observed latency limits direct transfer to time-sensitive rehabilitation interaction. Finally, future studies should move toward clinically representative settings, adaptive exercise protocols, and longitudinal feedback mechanisms. Such designs would allow AMMs to learn from repeated patient interaction and to support rehabilitation decisions across stationary, outpatient, and home-related contexts.

6 Conclusion

This paper presented a field study of an Artificial Mental Model (AMM)-based system for AI-supported rehabilitation. The study evaluated both predictive validity for perceived exercise difficulty and the feasibility of privacy-preserving local edge deployment in a public, high-throughput setting. The results show that predictive personalization remains sensitive to task design, rating distributions, and user interpretation, while the underlying local no-cloud deployment used here to enable privacy-preserving handling of sensitive health data remained operationally feasible even under constrained hardware conditions. Overall, the study highlights a translation gap between AMM-based personalization concepts and

real-world rehabilitation support. Future work should therefore focus on context-sensitive calibration, clinically representative evaluation, low-latency privacy-preserving deployment, and integration pathways that make AMM-based support actionable in rehabilitation practice.

Acknowledgments

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

Literaturverzeichnis

[Al24] Alzu'Bi, A. et al.: A Review of Privacy and Security of Edge Computing in Smart Healthcare Systems: Issues, Challenges, and Research Directions. Tsinghua Science and Technology 29 (4), S. 1152–1180, 2024, https://doi.org/10.26599/TST.2023.9010080.

[Am19] Amershi, S. et al.: Guidelines for human-AI interaction. In: Proceedings of the 2019 chi conference on human factors in computing systems. S. 1–13, 2019.

[At25] Attoh-Mensah, E. et al.: Artificial Intelligence in Personalized Rehabilitation: Current Applications and a SWOT Analysis. Frontiers in Digital Health 7, S. 1606088, 2025, https://doi.org/10.3389/fdgth.2025.1606088.

[Bo82] Borg, G. A. V.: Psychophysical Bases of Perceived Exertion. Medicine & Science in Sports & Exercise 14 (5), S. 377–381, 1982.

[CE22] Choudhury, A.; Elkefi, S.: Acceptance, Initial Trust Formation, and Human Biases in Artificial Intelligence: Focus on Clinicians. Frontiers in Digital Health 4, S. 966174, 2022, https://doi.org/10.3389/fdgth.2022.966174.

[He04] Hevner, A. R. et al.: Design Science in Information Systems Research. MIS Quarterly 28 (1), S. 75–105, 2004.

[JGK17] Johnson-Laird, P. N.; Goodwin, G. P.; Khemlani, S. S.: Mental Models and Reasoning. In (Ball, L. J.; Thompson, V. A., Hrsg.): International Handbook of Thinking and Reasoning. Routledge, S. 346–365, 2017.

[JMS24] Janzen, S.; Maaß, W.; Saxena, P.: Investigation of Artificial Mental Models for Healthcare AI Systems. In: 19th International Conference on Design Science Research in Information Systems and Technology. International Conference on Design Science Research in Information Systems and Technology (DESRIST-2024), June 3-5, Trollhättan, Sweden. University West, Springer, Högskolan Väst 461 86 Trollhättan, 2024.

[Jo89] Johnson-Laird, P. N.: Mental Models. In (Posner, M. I., Hrsg.): Foundations of Cognitive Science. MIT Press, Cambridge, MA, S. 469–499, 1989.

[JSA+25] Janzen, S.; Saxena, P.; Agnes, C. et al.: KI in der Rehabilitation – Anwendung künstlicher mentaler Modelle für eine personalisierte Medizin. Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz 68, S. 889–897, 2025.

[Kh23] Khera, R. et al.: AI in Medicine—JAMA's Focus on Clinical Outcomes, Patient-Centered Care, Quality, and Equity. JAMA 330 (9), S. 818–820, 2023.

[KHB25] Kouzy, R.; Hong, J. C.; Bitterman, D. S.: One shot at trust: building credible evidence for medical artificial intelligence. The Lancet Digital Health 7 (7), S. 100883, 2025, https://doi.org/10.1016/j.landig.2025.100883.

[Li24] Liou, L. et al.: Assessing calibration and bias of a deployed machine learning malnutrition prediction model within a large healthcare system. npj Digital Medicine 7 (1), S. 149, 2024, https://doi.org/10.1038/s41746-024-01141-5.

[Mo18] Morishita, S. et al.: Relationship between the Rating of Perceived Exertion Scale and the Load Intensity of Resistance Training. Strength and Conditioning Journal 40 (2), S. 94–109, 2018.

[Ni25] Nicora, G. et al.: Systematic Review of AI/ML Applications in Multi-Domain Robotic Rehabilitation: Trends, Gaps, and Future Directions. Journal of NeuroEngineering and Rehabilitation 22 (79), 2025, https://doi.org/10.1186/s12984-025-01605-z.

[No87] Norman, D. A.: Some Observations on Mental Models. In (Baecker, R. M.; Buxton, W. A. S., Hrsg.): Human-Computer Interaction: A Multidisciplinary Approach. Morgan Kaufmann, San Francisco, CA, S. 241–244, 1987.

[Pe07] Peffers, K. et al.: A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems 24 (3), S. 45–77, 2007.

[POS24] Pereira, C. V. F.; Oliveira, E. M. d.; Souza, A. D. d.: Machine Learning Applied to Edge Computing and Wearable Devices for Healthcare: Systematic Mapping of the Literature. Sensors 24 (19), 2024.

[RAC24] Rancea, A.; Anghel, I.; Cioara, T.: Edge Computing in Healthcare: Innovations, Opportunities, and Challenges. Future Internet 16 (9), 2024.

[Si23] Sivaraman, V. et al.: Ignore, Trust, or Negotiate: Understanding Clinician Acceptance of AI-Based Treatment Recommendations in Health Care. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. CHI '23, Association for Computing Machinery, New York, NY, USA, 2023, https://doi.org/10.1145/3544548.3581075.

[St20] Stuckenschneider, T. et al.: Rating of Perceived Exertion – A Valid Method for Monitoring Mild to Moderate Exercise Intensity in Individuals with Subjective and Mild Cognitive Impairment? European Journal of Sport Science 20 (2), S. 261–268, 2020, https://doi.org/10.1080/17461391.2019.1629632.

[Wi17] Williams, N.: The Borg Rating of Perceived Exertion (RPE) Scale. Occupational Medicine 67 (5), S. 404–405, 2017.