Leveraging Emotion Recognition to Power Adaptability for More Effective Speech Therapies*

Miriana Calvano[1], Antonio Curci[1,2], Andrea Esposito[1], Rosa Lanzilotti[1], Antonio Piccinno[1] and Alfonso Pio Pretorino[1]

[1]Department of Computer Science, University of Bari Aldo Moro, Via E. Orabona 4, 70125, Bari, Italy
[2]Department of Computer Science, University of Pisa, Largo B. Pontecorvo 3, 56127, Pisa Italy

Abstract

Speech therapy is a medical area focused on diagnosing and treating speech impairments, which affect an individual's ability to communicate effectively and develop linguistic skills. While these difficulties can arise at any stage of life, they are most commonly observed in childhood. In this context, technology plays a crucial role in supporting therapists while also enhancing patient engagement, reducing boredom, and minimizing frustration during treatment. To address this challenge, the article explores the integration of AI-driven emotion recognition techniques in a web platform called *e-SpeechT* that supports the actors involved in speech therapy (i.e., therapist, caregiver and patient) when creating, managing and performing it. This research work proposes new functionalities that can be implemented to improve the effectiveness of the treatment while making the system more adaptable to patients' needs, skills and emotional states, fostering a seamless human-AI symbiosis. The main objective of e-SpeechT is to ensure a more sustainable usage and development of resources while providing an easier access to the treatment.

Keywords

Speech disorders, Emotion recognition, Symbiosis, User Engagement

1. Introduction

The use of Artificial Intelligence (AI) is increasingly spreading in every aspect of society. AI can be integrated into many fields, modifying and enhancing the interaction process, especially in critical fields like medicine, where it can be employed to aid diagnosing illnesses, and to monitor and assess therapy's progress [1]. Integrating technology in therapies allows therapists to continuously manage the treatments while allowing patients to perform exercises at home [2]. In addition, it avoids wasting resources in traveling to attend physical appointments at the physician's office, while benefiting from the advantages and guidance of technology and reducing the environmental impact [3, 4]. This research work focuses on speech therapy, which is a field of medicine that aims at treating impairments concerning linguistic abilities (e.g., speech, language, cognitive-communication) [5, 6]

e-Health has been changing how professionals and patients carry out their activities, enabling remote treatments, monitoring progress more easily, and minimizing stress levels [7, 8]. Integrating AI in this context can revolutionize e-health, elevating it to new heights. Regarding speech therapy, AI-based functionalities can be used for automatically correcting exercises or quicker diagnoses [9]. However, it is essential that AI systems comply with the legal and ethical requirements that delineate the standard to follow to safeguard human rights while promoting sustainability when creating AI systems [10]. More specifically, the AI Act represents the main reference point for building human-centric and compliant systems that fulfill these objectives.

This work revolves around a web-application, called *e-speechT*, which aims to support speech therapies by reducing the cognitive demand of tasks and improving its efficacy for professionals, patients—children

Joint Proceedings of IS-EUD 2025: 10th International Symposium on End-User Development, 16-18 June 2025, Munich, Germany.
miriana.calvano@uniba.it (M. Calvano); antonio.curci@uniba.it (A. Curci); andrea.esposito@uniba.it (A. Esposito); rosa.lanzilotti@uniba.it (R. Lanzilotti); antonio.piccinno@uniba.it (A. Piccinno); a.pretorino@studenti.uniba.it (A. P. Pretorino)
0000-0002-9507-9940 (M. Calvano); 0000-0001-6863-872X (A. Curci); 0000-0002-9536-3087 (A. Esposito); 0000-0002-2039-8162 (R. Lanzilotti); 0000-0003-1561-7073 (A. Piccinno); 0009-0000-8422-7808 (A. P. Pretorino)
© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

CEUR-WS.org/Vol-3978/short-s2-05.pdf

from 4 to 8 years old—, and caregivers [11]. Although the system was developed in previous research works, we propose a prototype of a new AI-based component in its architecture to overcome the challenges that concern catching children's attention while carrying out exercises for their treatment. The higher objective of e-SpeechT is contributing to a more sustainable development of healthcare, providing continuous and more rapid access to treatments even to individuals who cannot physically reach hospitals and therapists' offices.

The step forward that this research proposes concerns employing techniques to detect children's emotions to adapt to their emotional state and change the system's behavior [12]. It is an AI-based feature that fosters a symbiotic relationship between children and technology without replacing professionals' expertise and supporting both parties. We consider the main characteristics of symbiosis to create a prototype of a new version of e-SpeechT, as described in Section 3 [13, 14]. The objective is to make e-SpeechT fall in the category of Symbiotic AI (SAI) systems, a specialization of Human-Centered AI (HCAI) [14], that highlights a bidirectional relationship between the two parties where the strengths of one compensate for the limitations of the other.

The article is structured as follows: Section 2 describes e-SpeechT functionalities and explores methods to assess emotions; Section 3 presents an overview of the new prototype, describing its motivations and current state; Section 4 draws conclusions and future works of this study.

2. Background

This section provides an overview of e-SpeechT, with a focus on its already-existing AI-driven capabilities. First, the platform's structure and architecture is described, detailing the roles of professionals, patients, and caregivers in the therapeutic process. In the second part, it explores different methods for assessing emotions that are being considered for the integration of the additional AI-based component.

2.1. eSpeechT

eSpeechT is a web application designed to support the treatment of speech disorders by assisting therapists in managing therapies, enabling patients to complete assigned exercises, and allowing caregivers to monitor treatment progress. The system has distinct areas, each tailored to the specific needs of the involved actor (i.e. therapist, caregiver and patient), each reported below.

Therapist They can create diagnoses, manage therapies, monitor and assess their patients. The system allows to create exercises based on three default categories defined a-priori with the aid of a group of professionals that work in this field (i.e., *Naming images*, *Minimum pair recognition*, and *Repetition of words*). These exercises can be packed together in a series or administered as standalone. The system features a functionality that allows therapists in automatically correcting exercises, tailoring feedback based on the severity of the child's impairment through Machine Learning (ML) techniques. The feature relies on a error tolerance threshold, set by the therapist, according to the severity of the patient's disorder (i.e., slight, moderate, and severe) [11, 4].

Patients They are the subjects of speech therapy and, in this case, they are children aged from 4 to 8. They are given exercises to complete in order to improve their condition and address their speech impairments [11]. As e-SpeechT deals with children, it employs *gamification* elements to make children feel more at ease and comfortable while carrying out therapies can increase engagement levels and distract them from the seriousness of the activity, leading to more positive outcomes [15].

Caregivers Since patients are not self-sufficient and must be guided through this process, caregivers play a crucial role in supporting children in performing the activities assigned by the therapist. They act as a middleman between the two parties, being able to both guide their caretakers while monitoring

their progress. They can also intervene in the User Interface (UI) of the patient's side of the system, customizing its appearance to make it more welcoming and adjusted to their preferences [11].

2.2. Techniques to assess children's emotions

Evaluating children's motivation and monitoring their emotions when interacting with technology can be a challenging activity. A rapid review of the literature was carried out to identify the main techniques that can be employed in the context of this research: self-reports, behavioral observations, physiological measurements, and the use of technological tools designed for emotion tracking [16]. An overview of such techniques is provided below with respect to the extent to which they can support the process of recognizing children's emotions.

- *Self-Report Measures*: these instruments are employed to assess children's motivation and emotional states when performing activities that require a cognitive effort. These tools often involve questionnaires or surveys where children are asked to reflect on their feelings and engagement levels. A crucial aspect to consider is the child's age and cognitive development state to ensure that they are not overwhelmed with the task [17].
- *Behavioral Observations*: direct observation of children's behavior can provide insights into their motivational and emotional states while performing their activities. The objects of the observation are task persistence, facial expressions, and body language; it can be conducted by recoding the child while performing the activity by individuals or through automated AI-based systems [18]. To observe and track patient's behaviors over time, longitudinal studies can be conducted during which therapists and/or caregiver can collect observations filling a diary.
- *Physiological Measurements*: physiological data, including heart rate, skin conductance, and facial muscle activity, can offer objective indicators of emotional arousal and motivation. Wearable devices are increasingly used in educational settings to collect this data, providing real-time insights into users' emotional experiences [19].

Implementing these methods requires careful consideration of ethical standards, especially concerning children's privacy and parents' consent. In the context of this research, combining multiple assessment approaches can provide a comprehensive understanding of children's motivational and emotional dynamics when carrying out activities to develop more engaging and effective educational tools.

3. Proposal of the Emotion Recognition Feature

Following the Human-Centered Design (HCD) approach [20], an interview carried out with two speech therapists. They both had been involved in previous user studies concerning e-SpeechT and, when providing their thoughts regarding their experience with the system, they raised the need real-time adaptation of the therapy to the children's emotional state in order to keep them engaged. These interviews acted as a springboard for the next development steps for e-SpeechT, suggesting that the feature proposed by the therapists could significantly improve its medical effectiveness. The proposal consists of a new component, powered by an AI model, that can fulfill the purpose of recognizing the children's emotions. This section illustrates the prototypes that were created, focusing on the User Interfaces (UIs) that belong to this functionality, fostering its adaptability while enabling therapists to monitor the level of attention during the treatment, helping them refine their interventions.

3.1. AI-Act Driven Design

The characteristics that we consider for the creation of new components are: *Transparency*, *Fairness*, *Automation Level*, and *Protection* [13, 21, 22]. More specifically, Transparency, ensures AI operations are understandable through explainability and interpretability while Fairness promotes unbiased and equitable AI behaviors. The Automation Level principle aims at balancing human control with automation to keep humans always on- or in-the-loop. At the same time, Protection, safeguards users' privacy,

safety, and security. SAI reinforces *Trustworthiness* and *Sustainability* as characteristics of HCAI since the first fosters reliable and ethical interactions, and the second minimizes environmental impact and promoting long-term societal benefits [13, 23].

In the case of e-SpeechT, the goal is to recognize emotions to adapt to children's behavior, eliminating unnecessary sessions and avoiding burnout for both parties. The final result that it is intended to obtain consists of a system that embodies the four principles safeguarding humans in all of their dimensions. It is important to underline that the principles are not only embodied in the user interfaces, but in the whole interaction process.

3.2. Emotion Monitoring and Analysis

During the creation of the therapy, medical professionals can enable emotion recognition functionality for each exercise of the session they are creating. In case it is enabled, the system will record the session, which can be performed both at home or in person, using the device's camera and/or biometric sensors to gather data to provide to the AI model. At the end of the therapy a single-question questionnaire appears to directly ask the child how they feel. The answer is provided through a simleyometer [24, 9] in order to let participants reply directly regardless of their age. Upon the session completion, therapists can reproduce the session with an overlay displaying the detected emotional states at various points as shown in Figure 1. In this proposal, a graphical representation of the emotions detected during the session is provided: the y-axis represents different emotional states (e.g., happiness, frustration, concentration); the x-axis represents time in minutes. At the same time, user studies are necessary to identify a more suitable technique to illustrate the emotions detected, since providing access to this information can have a strong impact on users' perceptions [25].

Figure 1: User interface prototype for the emotions' monitoring during the treatment.

The button "Why this emotion?" provides an explanation to therapists about why specific emotions were detected, enabling them to eventually correct the system's classification through the "Correct" button, as described in the following section.

It is important to underline that, in order to protect users from potential harm and prevent discriminatory behaviors, the access to data will be restricted to individuals involved in the project and possible bias will be reduced as much as possible [26].

3.3. Explainability and Human Intervention

Providing explanations concerning the detected emotions is crucial for the *Transparency* principle. This can include visual explanations, such as heatmaps or tracking matrices, in order highlight specific frames where certain emotions were identified.

In accordance with the *Automation Level* principle, therapists can review detected emotions and validate predictions, as shown in Figure 2. It is intended to implement an *Interactive Machine Learning* mechanism to re-train the AI model based on the expert's validation in order to improve its accuracy over time [27].

During in-person therapy sessions, therapists can gather insights about a child's emotions and later use this information to provide additional valuable feedback to the system. This fosters a collaborative approach where human expertise and AI capabilities complement each other, ultimately enhancing the effectiveness of therapy.

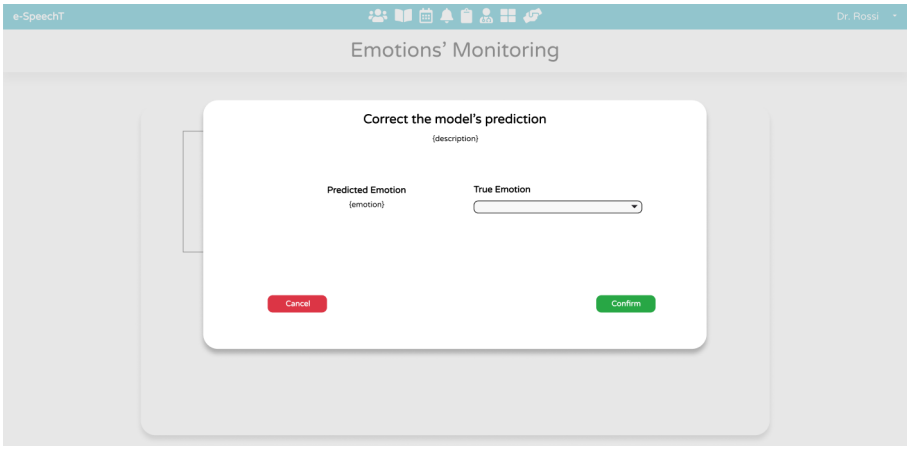

Figure 2: User interface prototype for the emotions' correction performed by the therapist.

We highlight that the caregiver of the patient will be mandatorily asked the permission to activate this functionality.

3.4. Treatment Personalization

The integration of emotion detection and analyzing the emotional states of children during exercise sessions, e-SpeechT can provide deeper insights into a child's engagement and emotional responses, allowing for more tailored therapeutic interventions. In addition, the adaptability of the system plays an important role when performing the exercises: if signs of fatigue, distraction, or frequent mistakes are detected, it can dynamically adjust the difficulty, offering simpler tasks or interactive games to help maintain engagement and motivation. By monitoring and analyzing children's emotional states, the system empowers therapists to adjust interventions based on real-time insights optimizing therapeutic outcomes. The long-term goal is to support a sustainable scientific progress in the context of speech therapy that integrates seamlessly into existing healthcare infrastructures to support long-term, ethical, and cost-effective speech therapy solutions.

4. Conclusions

The integration of AI-driven solutions in e-Health can bring strong advantages to the effectiveness of therapies and treatments, improving individuals' lives. At the same time, such systems must be accurately developed, considering human needs, preferences, and cognitive models. The proposal presented in this study concerns e-SpeechT, a web-application to support speech therapy. The goal is to continuously monitor and analyze the emotional state of children during therapy sessions to let the system adapt to their behavior and skills. The creation of more personalized therapy plans based on real-time feedback can improve the long-term impact of the system increasing the efficiency of treatment while ensuring the responsible use of resources. Thus, the possibility of providing remote support to professionals and patients can be enhanced by the adaptation of the e-SpeechT to the ever-evolving needs of children safeguarding the environment surrounding them.

Future works concern the actual development of the AI model and the implementation of the UIs. It is intended to build a multi-modal model that processes logs as tabular data and snapshots of the child's facial expressions while carrying out the exercises. By analyzing both structured evaluations and therapist observations, the system can assess a child's attention and emotional state, contributing to a more tailored therapeutic approach. To assess the validity and efficacy of the proposed solutions user studies will be conducted.

Acknowledgments

The research of Miriana Calvano, Antonio Curci, Rosa Lanzilotti, and Antonio Piccinno is supported by the co-funding of the European Union - Next Generation EU: NRRP Initiative, Mission 4, Component 2, Investment 1.3 – Partnerships extended to universities, research centers, companies, and research D.D. MUR n. 341 del 15.03.2022 – Next Generation EU (PE0000013 – "Future Artificial Intelligence Research – FAIR" - CUP: H97G22000210007).

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] E. L. Grigorenko, D. L. Compton, L. S. Fuchs, R. K. Wagner, E. G. Willcutt, J. M. Fletcher, Understanding, educating, and supporting children with specific learning disabilities: 50 years of science and practice, American Psychologist 75 (2020) 37–51. doi:10.1037/amp0000452.

[2] D. A. Ramalingam, D. A. Karunamurthy, D. T. Amalraj Victoire, B. Pavithra, Impact of Artificial Intelligence on Healthcare: A Review of Current Applications and Future Possibilities, Quing: International Journal of Innovative Research in Science and Engineering 2 (2023) 37–49. URL: https://quingpublications.com/journals/ijirse/2023/2/2/se230522005.pdf. doi:10.54368/qijirse.2.2.0005.

[3] V. Barletta, M. Calvano, A. Curci, A. Piccinno, A Protocol to Assess Usability and Feasibility of e-SpeechT, a Web-based System Supporting Speech Therapies:, in: Proceedings of the 16th International Joint Conference on Biomedical Engineering Systems and Technologies, SCITEPRESS - Science and Technology Publications, Lisbon, Portugal, 2023, pp. 546–553. URL: https://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0011893300003414. doi:10.5220/0011893300003414.

[4] M. Calvano, A. Curci, A. Pagano, A. Piccinno, Speech Therapy Supported by AI and Smart Assistants, in: A. Jedlitschka, A. Janes, V. Lenarduzzi, X. Li (Eds.), Proceedings of the 24th International Conference on Product-Focused Software Process Improvement, Springer Nature Switzerland, Cham, 2024, pp. 97–104. URL: https://doi.org/10.1007/978-3-031-49269-3_10. doi:10.1007/978-3-031-49269-3_10.

[5] R. DePompei, Speech-Language Therapy, Springer New York, New York, NY, 2011, pp. 2343–2344. URL: https://doi.org/10.1007/978-0-387-79948-3_925. doi:10.1007/978-0-387-79948-3_925.

[6] L. D. Shriberg, J. Kwiatkowski, H. L. Mabie, Estimates of the prevalence of motor speech disorders in children with idiopathic speech delay, Clinical Linguistics & Phonetics 33 (2019) 679–706. URL: https://www.tandfonline.com/doi/full/10.1080/02699206.2019.1595731. doi:10.1080/02699206.2019.1595731.

[7] C. McKean, S. Bloch, The application of technology in speech and language therapy, International Journal of Language & Communication Disorders 54 (2019) 157–158. URL: https://onlinelibrary.wiley.com/doi/10.1111/1460-6984.12464. doi:10.1111/1460-6984.12464.

[8] R. L. Bashshur, G. W. Shannon, B. R. Smith, D. C. Alverson, N. Antoniotti, W. G. Barsan, N. Bashshur, E. M. Brown, M. J. Coye, C. R. Doarn, S. Ferguson, J. Grigsby, E. A. Krupinski, J. C. Kvedar,

J. Linkous, R. C. Merrell, T. Nesbitt, R. Poropatich, K. S. Rheuban, J. H. Sanders, A. R. Watson, R. S. Weinstein, P. Yellowlees, The empirical foundations of telemedicine interventions for chronic disease management, Telemedicine and e-Health 20 (2014) 769–800. URL: http://dx.doi.org/10.1089/tmj.2014.9981. doi:10.1089/tmj.2014.9981.

[9] V. Barletta, M. Calvano, A. Curci, A. Pagano, A. Piccinno, Evaluation of "Speech System" and "Skill": An Interaction Paradigm for Speech Therapy:, in: Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies, SCITEPRESS - Science and Technology Publications, Rome, Italy, 2024, pp. 568–576. URL: https://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0012416700003657. doi:10.5220/0012416700003657.

[10] A. Aseeva, Liable and Sustainable by Design: A Toolbox for a Regulatory Compliant and Sustainable Tech, Sustainability 16 (2023) 228. URL: https://www.mdpi.com/2071-1050/16/1/228. doi:10.3390/su16010228.

[11] V. Barletta, M. Calvano, A. Curci, A. Piccinno, A new interactive paradigm for speech therapy, in: J. Abdelnour Nocera, M. Kristín Lárusdóttir, H. Petrie, A. Piccinno, M. Winckler (Eds.), Human-computer interaction – INTERACT 2023, Springer Nature Switzerland, Cham, 2023, pp. 380–385. URL: https://link.springer.com/chapter/10.1007/978-3-031-42293-5_39. doi:10.1007/978-3-031-42293-5_39.

[12] W. Wang, G. Athanasopoulos, G. Patsis, V. Enescu, H. Sahli, Real-Time Emotion Recognition from Natural Bodily Expressions in Child-Robot Interaction, Springer International Publishing, 2015, p. 424–435. URL: https://link.springer.com/chapter/10.1007/978-3-319-16199-0_30. doi:10.1007/978-3-319-16199-0_30.

[13] High Level Expert Group on Artificial Intelligence, Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment, 2020.

[14] G. Desolda, A. Esposito, R. Lanzilotti, A. Piccinno, M. F. Costabile, From human-centered to symbiotic artificial intelligence: a focus on medical applications, Multimedia Tools and Applications (2024). URL: https://link.springer.com/10.1007/s11042-024-20414-5. doi:10.1007/s11042-024-20414-5.

[15] G. Desolda, R. Lanzilotti, A. Piccinno, V. Rossano, A System to Support Children in Speech Therapies at Home, in: CHItaly 2021: 14th Biannual Conference of the Italian SIGCHI Chapter, ACM, Bolzano Italy, 2021, pp. 1–5. URL: https://dl.acm.org/doi/10.1145/3464385.3464745. doi:10.1145/3464385.3464745.

[16] B. D. Jones, Motivating and Engaging Students Using Educational Technologies, Springer International Publishing, 2020, p. 9–35. URL: http://dx.doi.org/10.1007/978-3-030-36119-8_2. doi:10.1007/978-3-030-36119-8_2.

[17] E. Iacono, C. Becchimanzi, A. Brischetto, Emotional design: Affective evaluation methods to assess the emotional response of 6-11 years children., in: S. Fukuda (Ed.), Affective and Pleasurable Design, volume 41, AHFE International, 2022. doi:10.54941/ahfe1001785.

[18] A. Beynon, D. Hendry, C. Lund Rasmussen, A. L. Rohl, R. Eynon, G. Thomas, S. Stearne, A. Campbell, C. Harris, J. Zabatiero, L. Straker, Measurement method options to investigate digital screen technology use by children and adolescents: A narrative review, Children 11 (2024) 754. doi:10.3390/children11070754.

[19] S. Ba, X. Hu, Measuring emotions in education using wearable devices: A systematic review, Computers & Education 200 (2023) 104797. URL: http://dx.doi.org/10.1016/j.compedu.2023.104797. doi:10.1016/j.compedu.2023.104797.

[20] ISO/TC 159/SC 4 Ergonomics of human-system interaction, Ergonomics of Human-System Interaction — Part 11: Usability: Definitions and Concepts, Standard ISO 9241-11:2018, International Organization for Standardization (ISO), 2018. URL: https://www.iso.org/standard/63500.html.

[21] G. Pavlidis, Unlocking the black box: Analysing the EU artificial intelligence act's framework for explainability in AI, Law, Innovation and Technology 16 (2024) 293–308. doi:10.1080/17579961.2024.2313795.

[22] J. Covelo De Abreu, The "Artificial Intelligence Act" Proposal on European e-Justice Domains Through the Lens of User-Focused, User-Friendly and Effective Judicial Protection Principles, in:

H. Sousa Antunes, P. M. Freitas, A. L. Oliveira, C. Martins Pereira, E. Vaz De Sequeira, L. Barreto Xavier (Eds.), Multidisciplinary Perspectives on Artificial Intelligence and the Law, volume 58, Springer International Publishing, Cham, 2024, pp. 397–414. doi:10.1007/978-3-031-41264-6_21.

[23] U. Pagallo, J. Ciani Sciolla, M. Durante, The environmental challenges of AI in EU law: Lessons learned from the Artificial Intelligence Act (AIA) with its drawbacks, Transforming Government: People, Process and Policy 16 (2022) 359–376. doi:10.1108/TG-07-2021-0121.

[24] A. Bell, Designing and testing questionnaires for children, Journal of Research in Nursing 12 (2007) 461–469. URL: http://dx.doi.org/10.1177/1744987107079616. doi:10.1177/1744987107079616.

[25] A. Esposito, G. Desolda, R. Lanzilotti, The fine line between automation and augmentation in website usability evaluation, Scientific Reports 14 (2024) 10129. URL: https://www.nature.com/articles/s41598-024-59616-0. doi:10.1038/s41598-024-59616-0.

[26] M. Saltarella, G. Desolda, R. Lanzilotti, V. S. Barletta, Translating Privacy Design Principles Into Human-Centered Software Lifecycle: A Literature Review, International Journal of Human–Computer Interaction 40 (2024) 4465–4483. URL: https://www.tandfonline.com/doi/full/10.1080/10447318.2023.2219964. doi:10.1080/10447318.2023.2219964.

[27] D. Arendt, C. Komurlu, L. M. Blaha, Chissl: A human-machine collaboration space for unsupervised learning, in: D. D. Schmorrow, C. M. Fidopiastis (Eds.), Augmented Cognition. Neurocognition and Machine Learning, Springer International Publishing, Cham, 2017, pp. 429–448. URL: https://link.springer.com/chapter/10.1007/978-3-319-58628-1_33. doi:10.1007/978-3-319-58628-1_33.