Communications of the Association for Information Systems

Volume 52 Paper in press

April 2023

Artificial Intelligence for Sign Language Translation – A Design Science Research Study

Gero Strobel
University Duisburg-Essen, Germany, gero.strobel@uni-due.de

Thorsten Schoormann
University of Hildesheim, Germany

Leonardo Banh
University Duisburg-Essen, Germany

Frederik Möller
TU Braunschweig/Fraunhofer ISST, Germany

Recommended Citation
Strobel, G., Schoormann, T., Banh, L., & Möller, F. (in press). Artificial Intelligence for Sign Language Translation – A Design Science Research Study. Communications of the Association for Information Systems, 52, pp-pp. Retrieved from https://aisel.aisnet.org/cais/vol52/iss1/33

Communications of the Association for Information Systems

Accepted Manuscript

Artificial Intelligence for Sign Language Translation – A Design Science Research Study

Gero Strobel
University Duisburg-Essen
Germany
gero.strobel@uni-due.de

Thorsten Schoormann
University of Hildesheim
Germany

Leonardo Banh
University Duisburg-Essen
Germany

Frederik Möller
TU Braunschweig/Fraunhofer ISST
Germany

Please cite this article as: Strobel, G., Schoormann, T., Banh, L., & Möller, F. (in press). Artificial intelligence for sign language translation – A design science research study. *Communications of the Association for Information Systems.*

Communications of the Association for Information Systems

Research Paper ISSN: 1529-3181

Artificial Intelligence for Sign Language Translation – A Design Science Research Study

Gero Strobel
University Duisburg-Essen
Germany
gero.strobel@uni-due.de

Thorsten Schoormann
University of Hildesheim
Germany

Leonardo Banh
University Duisburg-Essen
Germany

Frederik Möller
TU Braunschweig/Fraunhofer ISST
Germany

Abstract:

Although our digitalized society is able to foster social inclusion and integration, there are still numerous communities having unequal opportunities. This is also the case with deaf people. About 750,000 deaf people only in the European Union and over 4 million people in the United States face daily challenges in terms of communication and participation, such as in leisure activities but more importantly in emergencies too. To provide equal environments and allow people with hearing handicaps to communicate in their native language, this paper presents an AI-based sign language translator. We adopted a transformer neural network capable of analyzing over 500 data points from a person's gestures and face to translate sign language into text. We have designed a machine learning pipeline that enables the translator to evolve, build new datasets, and train sign language recognition models. As proof of concept, we instantiated a sign language interpreter for an emergency call with over 200 phrases. The overall goal is to support people with hearing inabilities by enabling them to participate in economic, social, political, and cultural life.

Keywords: Sign Language, Inclusion, Social Development, Machine Learning, Digital Innovation, Design Artifact.

1 Introduction

Imagine that you are not able to communicate, draw attention, or exchange thoughts with your (in)direct surroundings. What would you do in the case of an emergency? Do you have the patience to send a fax instead of making a quick call[1]? Deaf people face these situations in our society every day. Over 4 million people in the United States have limited abilities to communicate (Randolph et al., 2022), approximately 750,000 individuals in the European Union are deaf, and about 97% of the Deaf's social environments are unable to communicate with them on an equal basis (Council of Europe, n.d.). Worldwide, 97 million people would benefit from speech-language support (American Speech-Language-Hearing Association, n.d.). The inability to communicate with others is among the most challenging experiences of human beings (Brady et al., 2016) because communication helps to express needs, desires, and information to accomplish tasks (Dennis et al., 2008) and increase qualify of life (Randolph et al., 2022). As a result of not being able to communicate, people face numerous challenges in participating and accessing essential services provided by medical, public, business, and private institutions (Kritzinger et al., 2014; Mousley & Chaudoir, 2018; Skelton & Valentine, 2003; Taylor, 1999). Additionally, they are often stigmatized and discriminated against as most of our society perceives hearing loss as a dysfunction (Brennan, 2003; Mousley & Chaudoir, 2018). Society does not adequately address the issues of deaf communities (Aarts et al., 2021) and through their restricted ability to participate in conversations and their marginalization in professional and social environments, we can observe that deaf people tend to isolate themselves and create feelings of frustration (Randolph et al., 2022; Steinberg et al., 1999). The isolation is even reinforced by events (Jones, 2002; Mousley & Chaudoir, 2018), such as the COVID-19 pandemic and the need for social distancing (Park, 2020; Pietrabissa & Simpson, 2020; Recio-Barbero et al., 2020).

These facts point to fundamental issues for the individuals and their surroundings but also for the broader society and its overall intention to promote social development—i.e., improve well-being and remove barriers to provide equal opportunities for every individual and allow them to reach their full potential (United Nations, 1995). As digital technologies offer opportunities to create new value (e.g., Buck et al., 2022; Drechsler et al., 2020), the Information Systems (IS) discipline has a great potential to positively contribute to social concerns (Kranz et al., 2022; Olbrich et al., 2015; Schoormann & Kutzner, 2020; Wang et al., 2015). Following this, we can observe that research and practice have attempted to develop (digital) solutions for deaf people (e.g., Davydov & Lozynska, 2017; Harini et al., 2020; Kahlon & Singh, 2021; López-Ludeña et al., 2013). These attempts follow different strategies, from rather manual support to purely digital environments. Among them are solutions that are designed for the integration of human interpreters, for example, when using video conferencing tools. These options were mostly created during the COVID-19 lockdown and are implemented through features, such as the pinning of another window in the video conferencing screen for a language interpreter (Zoom, n.d.). Following this, many conference services provide live 'speech to text' translation (Acosta-Vargas et al., 2021; Hersh et al., 2020). Similar approaches can be found in the public sector. The states of the Federal Republic of Germany, for example, are developing *Nora*, an application that is intended to enable people with speech and hearing disabilities to make emergency calls (Ministry of the Interior of the State of NRW, 2022). This app provides an icon-based menu and text-based chat for communication. However, although we can observe progress towards more inclusive and sustainable approaches, numerous hurdles remain, including the dependency between deaf individuals and interpreters/translators, the unintended tendency to push the Deaf into the listener role instead of an active participator, and the limited accessibility of services via icons but not actual sign language. This is problematic because an unequal level of social engagement might be achieved, failing to fulfill the needs of the target user group. Furthermore, people are often prevented from communicating in their native languages (e.g., Mäkipää et al., 2022) as well as prompted to interpret icons and texts, which can be challenging as some individuals have a lower level of written competence due to their deafness (Dostal & Wolbers, 2014; European Union of the Deaf, 2021).

Against this backdrop, we draw on the promising opportunity from the IS community to advance inclusive and sustainable social development (e.g., AbuJarour et al., 2019; Andrade & Doolin, 2016; Watson et al., 2021) and aim to support a specific community, namely deaf people. Therefore, we ask:

> **How to design a system that supports equal communication and participation with deaf people?**

[1] Emergency Fax - https://t1p.de/ssa9c

In pursuing to answer the question, this paper reports on results from a design science research (DSR) project aiming to design and implement a novel artifact in the form of a sign language translator. This translator is intended to support deaf people, especially within digital environments but also in physical situations by means of complementary digital applications, such as from a smartphone. Given that digital technologies can promote equality and inclusion (Schoormann et al., 2023; Vinuesa et al., 2020), we drew on artificial intelligence (AI), machine learning in particular. Our solution covers (A) a novel machine learning pipeline that can be adopted and transferred to other contexts and (B) a software-based translator that instantiates the pipeline. Our translator is capable of recognizing sign language based on over 500 data points from gestures and converting them into textual representations. With our results, we support people with hearing handicaps and contribute to the field of social development by enabling people to participate in social and cultural life (e.g., Schoormann & Kutzner, 2020; Wilson & Secker, 2015). In doing this, it also leverages recent (governmental) endeavors for equal access to information, communication, and other services, particularly for hearing handicaps (e.g., European Union of the Deaf, 2021). We complement the available research on sign translation (e.g., Camgoz et al., 2020b; Yin & Read, 2020) as well as augmentative and alternative communication systems (e.g., Randolph et al., 2022) by presenting a holistic and functional software prototype and its underlying machine learning-based pipeline.

The paper is structured as follows. After the Introduction, we present the basic theoretical underpinnings and briefly outline the concept of sign language, its structure and usage. We also review existing literature in the area of ML-based sign language recognition and identify relevant gaps. Building on the presented methodological approach, we present our results in the form of the sign language interpreter and the ML pipeline including the demonstration of the result artifacts in the context of an emergency call. Finally, we elaborate on the implications and future directions for research as well as conclude with the paper.

2 Research Background

2.1 Social Development and Inclusion through Sign Language

The aforementioned challenges for individuals with hearing loss point to fundamental issues for the person concerned and their surroundings but also for the broader society and its overall intention to promote social development. Among other objectives, social development seeks to improve well-being and remove barriers to provide equal opportunities for every individual and allow her/him to reach their full potential (Vygotskij, 1981). To contribute to this, the Division for Inclusive Social of the United Nations Department of Economic and Social Affairs launched the United Nations Social Development Network (United Nations, n.d.). As part of the agreement of the World Summit for Social Development, social integration is specified as one of the three key issues of social development. Social integration refers to challenges from "*the pluralistic nature of most societies [which inhibits] equal access to all resources*" (United Nations, 1995). To overcome this, we need to "*end all de jure and de facto discrimination against persons with disabilities [through the] elimination of physical and social barriers with the aim of creating a society accessible for all*" (United Nations, 1995).

Advancing social developments is also of increasing interest to contribute to our grand challenges (Leal Filho et al., 2019), such as those presented by the 17 Sustainable Development Goals (SDGs) that seek to preserve peace and prosperity for people and the planet (United Nations, 2015). Having an inclusive world is a prerequisite for providing people with the same opportunities and is relevant in several areas, including access to quality education (SDG 4), healthcare (SDG 3), infrastructure (SDG 9), and safety (SDG 16), to governmental services and political participation (e.g., Barbosa et al., 2001; Corbett & Mellouli, 2017), as well as to general business offerings. Concerning the creation of an inclusive world, the IS discipline has particularly focused on how to make use of digital artifacts. Thereby, different streams of research have emerged, such as *e-inclusion* (Weerakkody et al., 2012) and *digital inclusion* (NDIA, 2019), which aim to improve quality of life via communication and information technology (Schoormann & Kutzner, 2020).

The ability to communicate is among the most fundamental ones to engage within a social community and make use of its offerings and services (e.g., Brady et al., 2016). It allows, for instance, to express feelings, needs, desires, and opinions (e.g., Randolph et al., 2022), to create attention in case of emergencies, as well as to participate in political initiatives. To have equal communication opportunities for all, people with hearing handicaps and their languages need to be considered appropriately. While digital tools for in-time translation are well-accepted to support communication between different spoken languages, there are only limited digital approaches to translate from spoken words (e.g., in English, Spanish, and German) to sign language, hindering inclusion (Núñez-Marcos et al., 2023; Patel et al., 2020; San-Segundo et al., 2008).

Sign language, similar to spoken language, belongs to the family of natural complex languages (Stokoe, 2005; Sutton-Spence & Woll, 1998). Internationally, there are over 100 different sign languages with corresponding grammar and regional dialects (Zeshan, 2006). In addition to national-based languages, there is also an international sign language that enables communication across national boundaries (Karpov et al., 2016). The central difference between spoken and signed language is the perspective of perception: Whereas spoken language draws on auditory modalities, sign language uses visual modalities to communicate (Stokoe, 2005). Thereby, it builds on a set of main building blocks, including hand position, hand movement, place of execution, and facial expressions. In addition, hand signs are supported by movements of the head or upper body. In consequence, it is important to consider the combination(s) of the individual building blocks of a sign (Boyes-Braem & Sutton-Spence, 2001). The complexity of recognizing numerous signs and combinations of blocks can be illustrated using the example of the words "*son*" and "*village*" from German sign language, which have the same hand position and movement but a different place of execution. This ultimately leads to a different meaning for the sign.

It should be noted that only about 30% of all gestures are iconographic (i.e., pictorial) and therefore recognizable by the communication partner without any additional knowledge (Klima & Bellugi, 2010; Marshall & Morgan, 2015; Ortega, 2017). The other two-thirds must be learned beforehand (Ortega, 2017). Klima and Bellugi (2010) assumed that about 30%-50% of the vocabulary of an adult deaf person is iconographic. Iconicity can be divided into three levels and ranges: *transparent signs* comprising signs recognizable by persons without knowledge of sign language, such as the sign for eating (one hand to the mouth); *semi-transparent signs*, which are not directly obvious, but after clarification, an object relation can be established (e.g., car = two hands on the steering wheel); and *non-transparent signs*, where no object relation can be established (Klima & Bellugi, 2010). In addition to individual signs, sign language also has a corresponding finger alphabet. This is equivalent to spoken language and is used in most cases to represent names or unknown words (Padden & Gunsauls, 2003).

2.2 Related Work and Research Objectives

Given the relevance of sign language, there is a corresponding research field that involves various tasks, including detection (to classify whether a person is signing in a video or not), segmentation (to detect the boundaries for signs or phrases in videos), and translation (to output the signs to spoken language).

Sign language recognition has been briefly studied by researchers to determine when in a conversation someone is signing. Borg and Camilleri (2019) used deep learning for training a recurrent neural network for detecting sign languages in videos. The authors demonstrated the feasibility of processing image data and motion data to recognize signers. Moryossef et al. (2020) improved the accuracy of a sign language detector by not only observing hand motion but the human pose. Additionally, they enabled real-time detection as the authors see a sign language detector as particularly useful during communicating, such as during videoconferencing. The authors argue that signers in videoconferences may get ignored or stress other participants that constantly check if someone starts signing.

Continuing the efforts of realizing a sign language interpreter, sign language segmentation is also crucial to detect sign and phrase boundaries for dividing them into meaningful units (Huang et al., 2018; Renz et al., 2020). The speed and simultaneity of sign languages makes it difficult to clearly identify boundaries compared to spoken languages that consist of a linear sequence of words. Current efforts do not leverage prosodic markers in sign languages, i.e., pauses, sign duration, facial expressions, or eye apertures (Bull et al., 2020; Bull et al., 2021; Renz et al., 2020; Renz et al., 2021).

Lastly, sign language translation (SLT) takes up a significant relevance for interpreting signers as its goal is to convert signs to spoken language. Initial research in sign language recognition and translation based on artificial intelligence can be dated back more than 20 years (Parton, 2006). Different works that rely on computer vision, robotics, 3D animation, or neural networks showed that SLT is an active area of research (e.g., Bantupalli & Xie, 2018; Bragg et al., 2019; Rastgoo et al., 2021). Nonetheless, Deep Learning-based approaches seem to achieve the most promising results as they are investigated by multiple researchers, for instance, by using single shot multi-box detection (Abiyev et al., 2020), sequence-to-sequence-models (Ko et al., 2019), or hierarchical long short-term memory models (Guo et al., 2018). Camgoz et al. (2018) were one of the first researchers to see SLT not as a gesture recognition problem but rather as a natural language problem. By applying a gated recurrent unit encoder-decoder architecture with Luong Attention, the authors were able to achieve reasonable results. In subsequent work, Camgoz et al. (2020b) improved their results by implementing a Transformer encoder. With a revised architecture that addresses multiple channels, Camgoz et al. (2020a) crop the signing hand and the face from the video and perform 3D pose estimation to fuse them into a multi-channel Transformer architecture. The resulting architecture overcomes the reliance on gloss annotations during the processing of videos, minimizing the costs for curating datasets to train on. Similarly, Yin and Read (2020) have developed a system based on a Transformer encoder-decoder model with comparable results. One of the most recent approaches is the deep learning-based approach by Adaloglou et al. (2022). The authors focused exclusively on deep neural network-based recognition architectures. However, the results show that further research on sequence learning models and higher semantic representations is needed to ensure reliable results.

Although valuable work has been done (Hammami et al., 2019; Mäkipää et al., 2022), to the best of our knowledge, no work has explicitly targeted an end-to-end sign language interpreter that is capable of detecting, segmenting, and translating sign language to spoken language and vice versa. This assumption is supported by Mäkipää et al.'s (2022) systematic literature review of over 1400 initial meetings, in which the authors identified only one publication that addressed the accessibility issues of people with disabling hearing loss but did not present a holistic solution (Mäkipää et al., 2022). Based on previous findings and technological paradigms, we strive to provide a holistic solution for equal communication of people with hearing handicaps that can be utilized in daily life. We argue that this research is an important piece in leveraging inclusive and sustainable social development.

3 Research Approach: A Design Science Research Study

To achieve our overall goal of including people with hearing inabilities, we are conducting a multi-year DSR project (Hevner et al., 2004), adapting the widely-accepted DSR methodology of Peffers et al. (2007) (see Figure 1). In this project, we aim to build and evaluate an AI-based sign language translator that establishes a pathway for accessible, equal, and inclusive communication for deaf people in their natural languages.

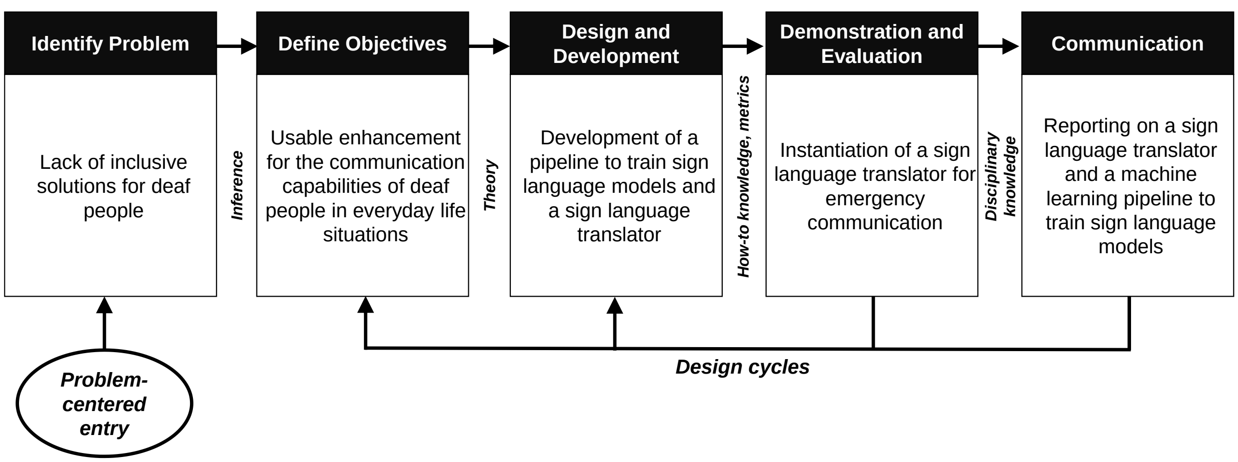

Figure 1. Research Design

3.1 Problem Identification

The starting point of the research process was the consideration of the current situation of the Deaf within society, who still face numerous obstacles concerning equal and inclusive accessibility, participation, and communication with people who are not able to speak sign language (e.g., Brady et al., 2016; Mäkipää et al., 2022; Randolph et al., 2022). The problem was grounded on three main sources: scientific literature, publicly available information from non-governmental organizations (e.g., European Union of Deaf) in the area of the stakeholder group, and empirical insights collected from interviews with professionals (i.e., interpreter/translator). Drawing on these sources, it has become apparent that deaf people have limited opportunities in terms of access to information and the ability to communicate (Mousley & Chaudoir, 2018). As an example (see also the Introduction), in large parts of the European Union, there is still no governmental option for deaf people to place an emergency call in their native language (i.e., sign language), despite the European Union's commitment to providing such an option back in 2018 (European Union of the Deaf, 2021). To concretize this impression and to gain a deeper understanding of the real life of the people affected and the challenges associated with it, we have interviewed professionals in this domain. Based on the discussion with certified sign language translator with more than five years of experience, we were able to mirror our findings and derive concrete requirements to address the problem.

3.2 Objectives of a Solution

To tackle the situation faced, our DSR project sets out to provide a purposeful and useful solution to enhance the communication capabilities of deaf people in everyday life situations (e.g., emergency calls, and governmental services). The ultimate goal is to develop a mobile system to enable daily use and translate sign language into text. Based on insights from the related work (see also Section 2), our own experience within this project (i.e., reflection-on-action), and empirical data collected through an interview with a professional sign language interpreter[2], we were able to sharpen our understanding of the problem and identified three key requirements for the solution (Figure 2).

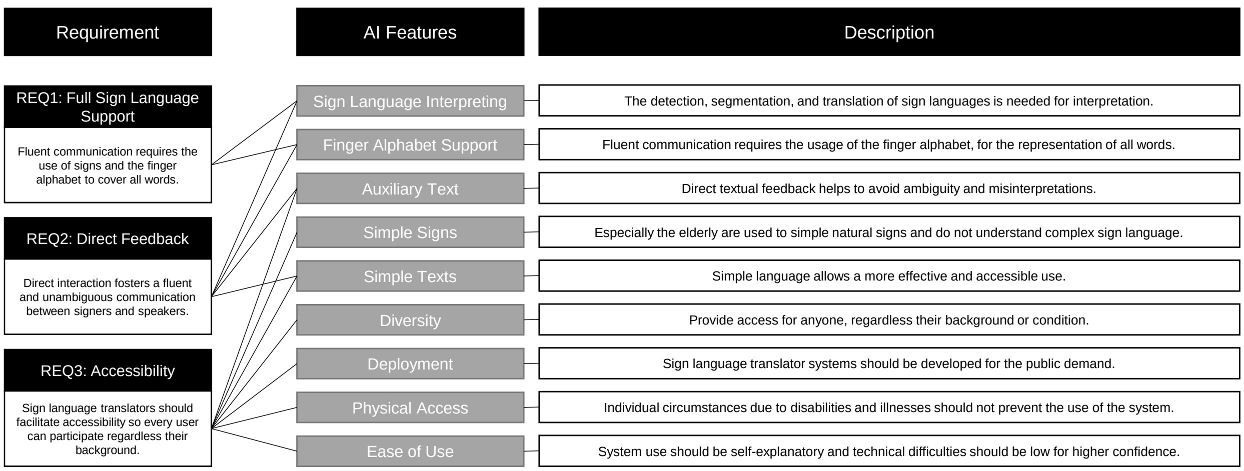

Figure 2. Objectives of a Solution

(REQ1) Full sign language support – As illustrated in the research background, fluent communication requires not only the use of signs but also the use of the finger alphabet, since corresponding signs do not exist for all words (e.g., proper names). **(REQ2) Direct feedback** – Similar to spoken language, sign language also has the possibility of ambiguity, so in order to provide the user with more security and transparency, "*an additional text to the signs would help to avoid misinterpretation in case of ambiguities.*" **(REQ3) Accessibility** – To enable every user in participating with sign language translators, accessibility features should be integrated. Hence, access is provided for diverse users, regardless their background or condition. The use of simple language allows a more effective and accessible system usage. This applies to both language types, the textual representation as "*learning written language is difficult for many deaf people due to different grammar and translations compared to spoken language*" and the selection of signs used because "*especially older people do not understand complex sign language and primarily use natural signs.*" Ultimately, sign language translators should be developed for the public demand and provide a self-explanatory user interface that everyone can operate without instruction, for instance by an "*[automatic] recognition when signing is in progress.*"

3.3 Design and Development

We translated the experience and requirements gathered from literature and practice into two artifacts: **(A)** a machine learning pipeline for training AI models to recognize sign language, and **(B)** a sign language translator. To do so, we drew on approaches from AI and particularly made use of Transformer neural

[2] We conducted a semi-structured face-to-face interview with a professional sign language teacher and interpreter. The expert has more than five years of professional experience and the language level C2 for German Sign Language. The interview lasted 54 minutes and was structured along areas for daily challenges of deaf people, equal communication between deaf and non-deaf people, communication hurdles, situations that are especially hard to handle, sign language vs. text recognition of deaf people, and requirements for inclusive translations.

networks (TNN). The structure and configuration of the TNN Transformers were adapted to meet the technical requirements for sign language translation. Thus, the number of encoders and decoders was reduced from six in the standard configuration to three. In addition, the normalization was pulled in front of the coders, which reduced the training time and optimized the overall system pipeline (Camgoz et al., 2020b; Xiong et al., 2020). To provide a tangible, applicable, and holistic approach, the prototyped lightweight sign language interpreter application **(B)** uses the models trained by the pipeline **(A)** as a basis for sign language recognition.

3.4 Demonstration and Evaluation

For evaluating both artifacts, we created our dataset, transformed it into a sign language recognition model with the help of the pipeline **(A)**, integrated it into the sign language prototype **(B)**, and evaluated it against traditional machine learning metrics as well as a quality metrics specialized for machine translation.

4 Artifact Description: Sign Language Translation

In the following, we present our main results. We first outline our underpinning AI-based approach of TNNs. We then describe **(A)** our machine learning pipeline (see Figure 4) that enables users to create datasets, preprocess them, and train TNN-based sign language recognition models, as well as **(B)** a sign language translator that translates the finger alphabet and sign language into text (see Figure 5).

4.1 Underpinning Technological Approach: Transformer Neural Networks

Following the insights from prior research (see Section 2), we decided to build upon Transformer neural networks for our solution. In general, artificial neural networks describe a class of machine learning algorithms based on the information processing of nervous systems in biological systems (McCulloch & Pitts, 1943; Rosenblatt, 1958). Due to their basic modular structure and the associated flexibility in the combination of processing units, artificial neural networks can be used in a variety of tasks in different domains and forms of expression (Janiesch et al., 2021). TNNs are one type of neural network, and they are based on attention mechanisms and facilitate the recognition of dependencies as well as the processing of sequential, position-dependent data (see Figure 3).

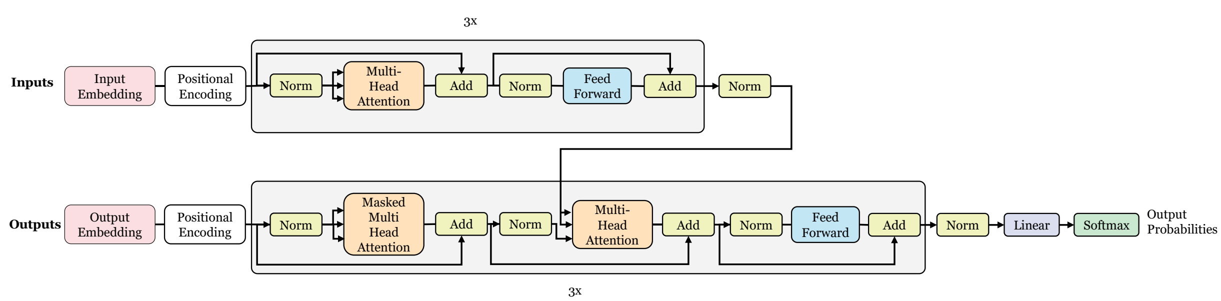

Figure 3. Transformer Neuronal Network (Xiong et al., 2020)

These capabilities are essential for the processing of visually represented sentences of a language (Camgoz et al., 2020b; Vaswani et al., 2017). To ensure that the order of the positional data within the sequence is not lost during processing, it is supplemented by positional encoding as an order plotted on a sine or cosine curve before it is passed to the Transformer (Vaswani et al., 2017). The Transformers are composed of an encoder and a decoder, which possess corresponding multi-head attention (MHA). The idea of attention is to recognize relationships between components of one or more data inputs and to create a valid output from them. Therefore, several heads are used per MHA, each receiving different parts of the input (or output). Each head has its own linear layer for converting the output to the size of the vocabulary, allowing variations in the inputs to be weighted and, thus, different aspects of them to be considered. Furthermore, this compensates for the increase in computational effort as the number of heads increases (Vaswani et al., 2017). The encoder itself is responsible for the pure processing of the input and comprises two components. Within the encoder, the inputs are weighted by the MHA and then processed further by a feedforward neural network. Multiple encoders can be combined with the output of one encoder, mapping the input of the following one (Vaswani et al., 2017). The decoder, however, has an additional masked MHA component. This component masks and preprocesses the previous "shifted" output to ensure that only known words are processed, and the next word can be determined to match the previous output. The further design is similar to the encoder with a feedforward neural network that receives the output of the previous MHA component as input and processes it further (Vaswani et al., 2017).

4.2 Machine Learning Pipeline

The basis in all machine learning-based recognition approaches is the data with which the model is trained and from which it learns (Padmanabhan et al., 2022). Source material for training models based on the pipeline forms video footage of sign language and corresponding transcripts of the signed content. For our training, we use individual video sequences that contain one sentence per recording in order to train as effectively and modularly as possible and to avoid the overhead for sentence separation. The decisive quality feature is not the video quality in the form of the resolution of the footage but rather the speed at which signing takes place. For this reason, videos produced for younger target groups or sign language beginners are well suited for training.

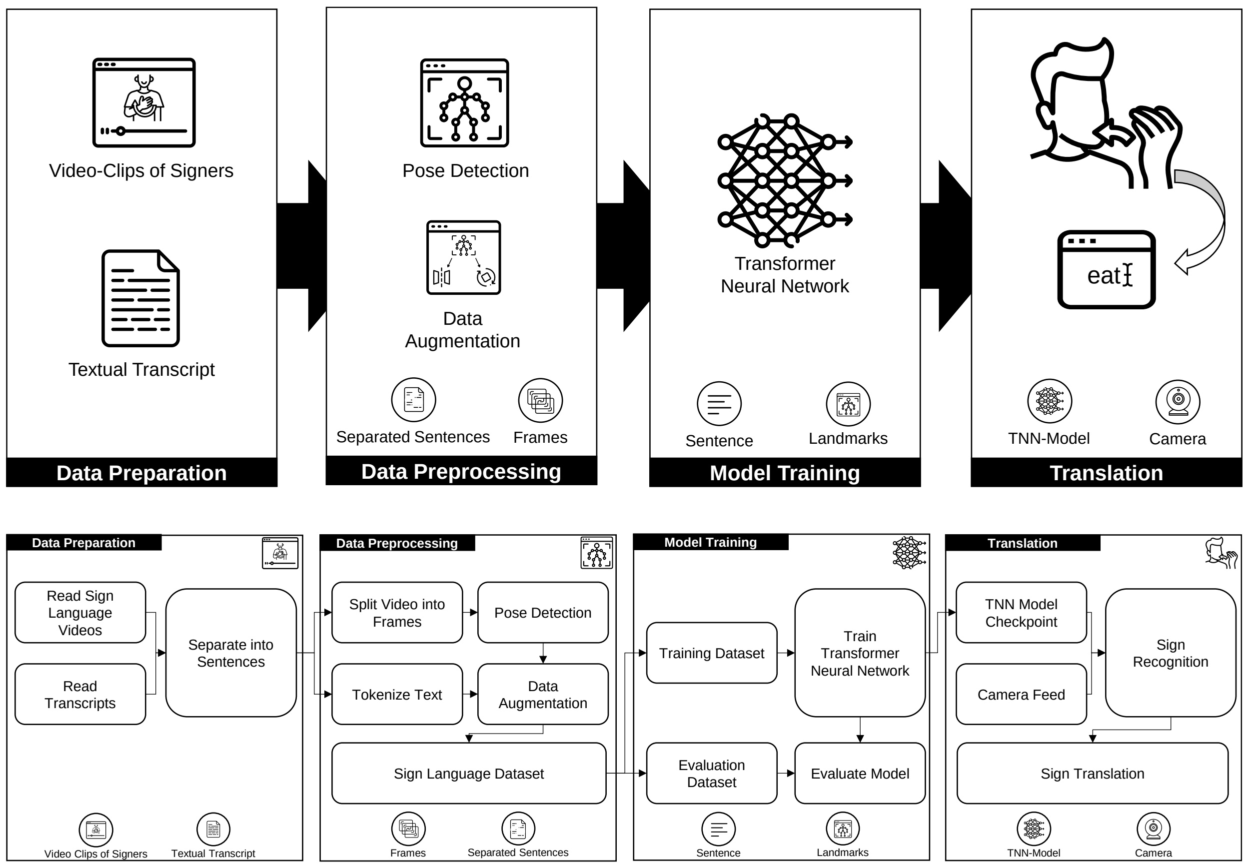

Figure 4. Sign Language Translator Pipeline

The first step of model training in the pipeline is **data preparation**, in which the transcript was split into individual sentences through natural language processing, and the video sequences were automatically separated into the according sentences. The processed data was organized into an appropriate folder structure and passed to the second level of the pipeline: data preprocessing. During the **preprocessing**, the individual images of a video sequence were analyzed for 543 position points (i.e., landmarks) via the pose detection framework Mediapipe. These landmarks act as features for training the TNN. The majority of the points (468) were located on the face for analyzing the mimic, and 21 points were used for hand position. The remaining 33 points described the pose of the signing person (Figure 5-right). For each landmark, the respective x, y, z, and visibility values are captured. To augment our dataset, we apply data augmentation techniques (i.e., random horizontal flips and shift-scale-rotate) on each sequence n=5 times. In the first stage of the prototype, tracking was limited to the hands. This was sufficient for recognizing single letters during fingerspelling; however, it was insufficient for recognizing gestures based on their multiple basic components; see also Section 2 Research Background for the importance of considering several components of sign language. In addition to the image data, the annotating text information was also pre-processed. The goal was to extract duplicates and train only unique tokens. Within the **model training** process, the acquired and pre-processed sign language data (i.e., landmarks and annotations) is fed into the TNN. The goal of model training is to generalize the dataset by learning from the samples, that is to heuristically adjust the TNN's weights. We split our dataset into a training and an evaluation dataset to reduce overfitting by exposing the model to new, unseen data (Sanjay & Sanyam, 2016). Training is performed with a batch size of 32 and the usage of an Adam optimizer (default parameters) for an average of 2000 epochs with an early stop mechanism based on three encoders as well as decoder layers and eight heads within the MHA components. Exposing one iteration of the full dataset to the neural network is called an epoch (Goodfellow et al., 2016). Also, the model is evaluated after every five epochs by metrics such as BLEU, translation error rate, or ChrF score (see Evaluation in Section 5). After training, the metrics are compared and the best model checkpoint is selected, that is the model state with the best metrics, minimized underfitting or overfitting. This model file (with the respective weights at the time of control) is integrated into the sign language interpreter and serves as the basis for the recognition and **translation** of the signs.

4.3 Sign-Language Translator

The sign language translator is a standalone tool that incorporates the model trained with the pipeline and based on it, interprets the signs performed by the user. The translator interprets fingerspelling and signs sentence by sentence (REQ1).

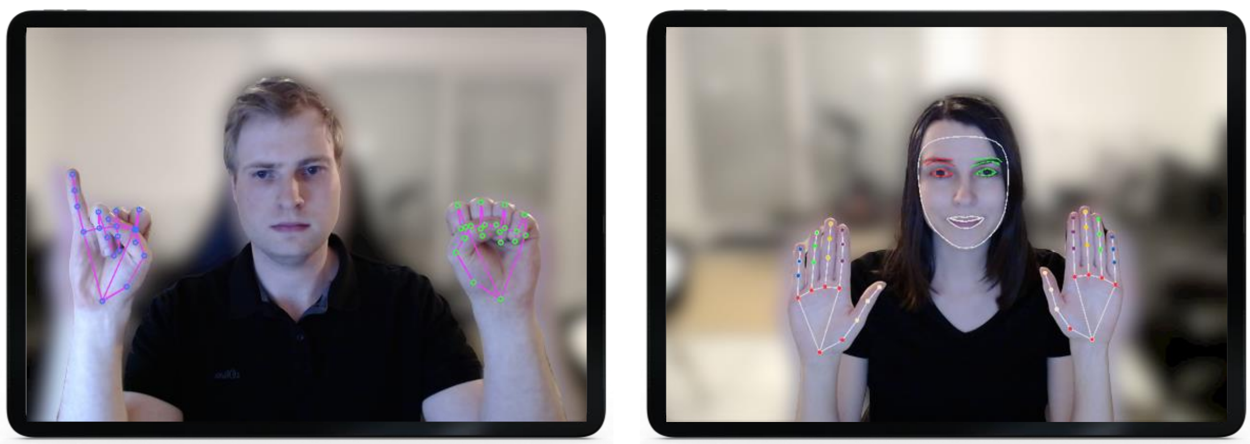

Figure 5. Finger Alphabet (left) and Sign Language (right)

The current prototype is designed as a web application, which allows device or operating system independent access. The only device requirement is a camera which serves as an input of the video stream. The application automatically detects the beginning of a gesture and starts the translation process. During the recording, the individual frames of the video stream are analyzed by the pose detection framework (i.e., Mediapipe) to generate the respective landmarks. Due to the direct pose detection on the device during the recording process, it is possible to detect whether all parts of the body relevant for the detection are captured on the camera image and, if necessary, prompt the user to adjust their position. These recordings and analyses are also done sentence by sentence to allow a consistent comparison with the trained models (REQ2). After the comparison, the corresponding translation will be returned to the UI and presented to the user in text. The disadvantage of the approach is that, depending on the model used, the prototype generates a relatively large footprint and is, therefore, correspondingly computationally intensive. Regardless, the sign language translator provides a usable approach to support and improve the communication process between signers and speakers.

5 Demonstration and Evaluation

To evaluate the usability of the research results, we trained a sign language recognition model using the pipeline based on our dataset and integrated the model into the sign language translator. The objective was to instantiate a sign language interpreter for communication in emergency situations. A starting point for the creation of the dataset was the official emergency fax, which deaf people can use for emergency communication until now. This fax is based on the classic W-questions (*What happened?*, *Where is the emergency?*, *What injuries?,* etc.), which are also asked in an emergency call in spoken language. Based on this framework, we identified over 200 phrases (sentences and keywords) necessary for communication and trained the model on 1638 publicly available augmented sign language videos (e.g., YouTube). In addition, the model was trained with basic communication phrases (*Hello, My name is, Goodbye*) to enable basic communication (REQ3). As the goal of training a TNN is to achieve high generalizability by learning and abstracting from the dataset, we procured a diverse dataset of signers from publicly available sign language videos. The phrases were signed by multiple individuals and care was taken to ensure that the signers were as diverse as possible. This approach enabled an inclusive model and ensured optimal recognition of every individual for translation. A complete list of communication phrases can be found in Appendix A.

To evaluate the model, we use four metrics. First, the loss or cost functions quantify the deviation between the output of the translator and the expected result value. For classification problems, *cross entropy loss* is considered the de-facto standard, and it calculates the difference between individual probability distributions for a set of events based on the information-theoretic notion of entropy (Goodfellow et al., 2016; Shannon, 1948). As illustrated in Figure 6 (upper-left), during training, the model learns to assign the appropriate output words for the corresponding gestures with approximately the correct probabilities (decreasing loss). After about 200 epochs, the loss is already below 10 and drops below the threshold of one in epoch 700. Although a loss of zero has not yet been reached in epoch 2000, a clear trend toward the zero value is recognizable and achievable with longer training.

Second, the *BiLingual Evaluation Understudy (BLEU)* score is a language-independent evaluation metric for machine translation. The metric is based on the N-gram comparison of a machine output with a human reference. N-grams represent fragments of a text and range from single letters to whole sentences. Following Papineni et al. (2002), the reference is divided into N-fragments and then checked the extent to which these fragments were present in the machine output. Typically, this is accomplished by averaging one to four grams and plotting a score between zero and one. The higher the value, the more similar the machine output is to the human reference, with a value of one symbolizing the exact overlap with the reference. BLEU score illustrates that the model started to formulate the first simple sentences that corresponded to the referenced solution after about 200 iterations through the training dataset (see Figure 6 upper-right). From this point on, the score and the associated translation ability of the translator continued to increase until it reached a value of 0.23 at approximately 1250 epochs, and it alternated around this value until the end of training.

The third metric to evaluate the quality of the translation outputs is the *translation edit rate* (TER). It is also a machine translation evaluation metric and measures the amount of editing that a human would have to perform so that the machine output matches the reference translation (Snover et al., 2006). Compared to the BLEU score, TER provides an easy-to-explain metric for non-experts in the field of machine translation. The lower the value, the better the results, as less editing is needed to achieve the ground truth. Our model yielded the first outputs after 200 epochs but showcased a high TER of around 30, as illustrated in Figure 6 (lower-left). After approximately 400 epochs, TER improved to less than three and continued with a slow decline until the end of the training, reaching TER of around 0.6.

Lastly, the *character n-gram F-score* (chrF) provides an automatic evaluation of machine translation output. Popović (2015) argues that n-gram-based F-scores represent good correlations with human judgements and outperform metrics such as BLEU and TER. As chrF is language-independent and tokenization-independent, it also provides a good evaluation metric for comparing machine translations to human judgments. As the first translations are produced after 200 epochs, initial chrF scores can be reported in the lower 0.1 area. As training continues and results improve, the chrF also increases up to 0.65 until the end of training (see Figure 6 lower-right).

Based on the qualitative evaluation metrics, it is noticeable that the TNN is learning the dataset signs and can formulate sentences, with the quality of the sentences improving over time. Moreover, a manual analysis revealed that in some cases, the translated sentences featured mostly correct grammar and contained words with similar meanings as in the reference sentence. Take, for example, the sentence "i need help. it is an emergency.," which the translator interpreted from the signs as "help me. emergency!" This comparison demonstrates that the basic intention of the sentence was recognized and correctly translated.

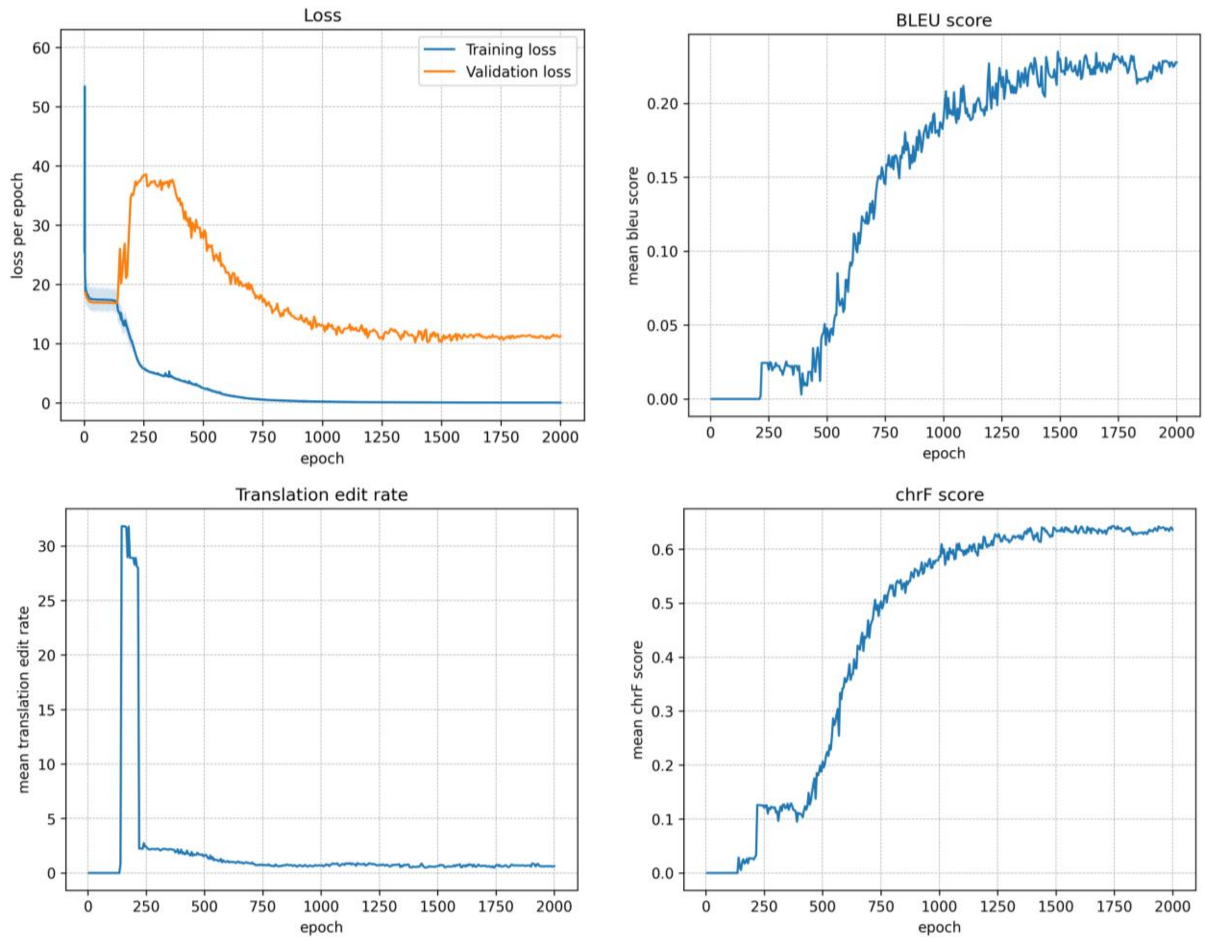

Figure 6. Loss (upper-left), BLEU score (upper-right), translation edit rate (lower-left), and chrF score (lower-right)

6 Discussion

6.1 Summary and Lessons Learned

Sign language is a fully recognized, complex, and natural language used by more than 80,000 people in Germany and about 750,000 people in the European Union (Council of Europe, n.d.). However, even though we are in the age of digitalization, people with hearing and communication handicaps still face major challenges, such as sending a fax[3] for emergencies or relying on external private services (e.g., TESS). In line with prior research that stressed the IS's power to "*foster the social inclusion of people regardless of factors such as handicaps, gender, culture, and incomes*" (Schoormann & Kutzner, 2020, p. 9), we report on a DSR project aiming to support people dealing with sign language. Therefore, this paper presents a machine learning-based pipeline and a sign language translator that is capable of translating sign language based on 543 analyzed position data points into non-sign language sentences.

Based on the design and evaluation of artifacts as well as the interaction with interpreter experts and the affected group, we are able to derive a set of content-related and technical lessons learned from our project. In Table 1, we summarize and elaborate on the main aspects to provide additional guidance for the field as well as point to aspects that should be reflected on.

[3] Emergency Fax - https://t1p.de/ssa9c

Table 1. Lessons Learned

Observation		Selected approaches for handling
Mode of communication (content-related)	During the project, it becomes apparent that relying on one primary mode (e.g., exclusively sign language, finger alphabet, or speech-to-text) is in many cases not sufficient.	Communication tools should provide modalities that are adaptable and accessible to diverse groups and individual circumstances. • Use auxiliary text • Use multi-modalities
Evolution of language (content-related)	Language is changing over time, and so does sign language. While spoken languages tend toward simplicity (e.g., due to smartphones), we experienced a contrary development in sign languages where especially older generations tend to use natural signs.	Communication tools should take linguistic changes into account and should allow for extendibility and adaptability. • (Re-)Update repository of languages • Consider generational differences
Appearance of users (content-related)	Our approach depends on the detection of pose landmarks, specifically the face and upper body, which might exclude people who do not fit these specific requirements.	Communication tools should take (partly) disguised people (e.g., face masks) into account. • Use different recognition points • Consider cultural differences
Style of communication (technical)	Depending on the actual domain and use cases (e.g., newspapers, scientific papers), different styles of language are used, which ultimately affects the translation quality.	Consider different styles of languages and their possible ambiguity. • Train data accordingly • Annotate data accordingly
Data input (technical)	AI-based translators are dependent on the data and training generalist models already requires a huge amount of data and computing power; Sign languages are multi-dimensional.	Consider the complexity and large amount of data to be processed. • Focus on a certain use case • Transfer/integrate results

6.2 Contributions and Implications

Our work has important theoretical and practical implications. **First** and foremost, from a practical view, our prototype can have a distinct impact on society, as it helps to support deaf people, and people having hearing issues in general, in the digital era. In doing so, we advance the possibilities for them to engage and participate in social (e.g., communication with friends and family) but also governmental (e.g., citizen services, political participation, education) and economic (e.g., making use of offerings) activities. Creating a more inclusive world responds to this special issue's call to "move from discussion to action" (Corbett et al., 2022) and general calls for promoting social development (e.g., improve well-being and remove barriers) (United Nations, n.d.), as well as has effects on several SDGs (United Nations, 2015), including quality education (SDG 4), well-being (SDG 3), and reduction of inequalities (SDG 10). Generally, following the results from Asadikia et al. (2021) who examined the power of single SDGs, social aspects are among the most impactful ones because addressing them can ultimately affect the entire SDG score.

Second, from an academic viewpoint, we contribute to the emerging stream of research using AI and digital technologies to promote equality and inclusion (e.g., Andrade & Doolin, 2016; Schoormann et al., 2021; Schoormann et al., 2023; Vinuesa et al., 2020). By focusing on the actual communication between people (i.e., ear-handicapped and non-ear-handicapped), we complement the valuable body of IS research that helps other parts of society, such as refugees (e.g., AbuJarour et al., 2019). Also, our results can be applied in both digital and physical (e.g., via smartphones) situations and thus helps to reduce accessibility barriers when using information technology artifacts. This is important because according to Mäkipää et al.'s (2022) review of the accessibility of technological artifacts, "only one study that focused fully on the accessibility issues of people with disabling hearing loss" (p. 681).

Third, our machine learning pipeline and prototype can be adapted to and contextualized in various fields of application, including education (e.g., Hammami et al., 2019; Lang et al., 2022), communication-intensive work tasks (e.g., Dennis et al., 2008), and equal access to citizen and administration services (e.g., Sidani et al., 2022). In line with accumulating knowledge and artifacts (vom Brocke et al., 2020), the pipeline is intended to be used, extended, and refined by other researchers to address specific demands.

Fourth, our prototype serves as a situational instantiation that already provides specific knowledge and impulses on how to inclusively design a specific class of augmentative and alternative communication systems (Randolph et al., 2022) as well as socio-technical systems in general (Olbrich et al., 2015). In the next steps, after conducting additional evaluations in more naturalistic settings, we plan to generalize the insights gathered during the project to formulate more abstract design knowledge (Gregor & Hevner, 2013).

6.3 Limitations and Future Directions

Although the AI-based prototype shows promising results for learning and translating signs, the study is not free of limitations, offering several opportunities for future research within the research project. We see three major areas for improvement in future research: **First,** from both a technical and application point of view, we are restricted to the specialization of the training and evaluation dataset used. The model learned thematically focused gestures and is limited in its type of language. Currently, it is proficient in American Sign Language. A multi-language approach and the integration of the international variant of sign language are conceivable. The primary problem would be the syntax differentiation of the corresponding spoken languages depending on the language and the associated variance in the translation. **Second**, user-centered artifact development is a key quality criterion in DSR. Hence, we have already integrated domain experts in the form of interpreters into the research process during requirements elicitation. Moreover, the objective is to integrate deaf people through cooperation with boarding schools and associations for the Deaf, evolve the prototype, and cooperatively develop further scenarios besides the emergency model for everyday scenarios. **Third**, the interpreter currently translates from signs to text in one direction only so that the Deaf, as the primary target group, will be able to communicate in their native languages. Whether bidirectional communication is desired or useful at this point will become apparent in the initial testing with affected individuals.

7 Conclusion

Following the paths towards more inclusive and sustainable social environments, this paper provides a digital solution to support deaf people in barrier-free and equal communication in daily situations and even in extreme situations, such as in case of emergencies. The solution covers two main parts, namely a machine learning-based pipeline to describe a possible procedure from processing data to recognizing sign language and a software prototype that serves as a situational instantiation and illustrates how the pipeline can be implemented. Initial discussions with translation experts/interpreters of sign language as well as insights from the more technical evaluation indicate the promising potential of supporting the group of ear-handicapped people. Ultimately, we hope to boost the emerging stream that is concerned with making use of digital innovations for inclusive, equal, and fair environments in both settings, digital and analog, by providing our piece of that large puzzle.

References

Aarts, E., Fleuren, H., Sitskoorn, M., & Wilthagen, T. (2021). *The New Common*. Springer International Publishing.

Abiyev, R. H., Arslan, M., & Idoko, J. B. (2020). Sign Language Translation Using Deep Convolutional Neural Networks. *KSII Transactions on Internet and Information Systems*, *14*(2).

AbuJarour, S., Wiesche, M., Díaz-Andrade, A., Fedorowicz, J., Krasnova, H., Olbrich, S., Tan, C.-W., Urquhart, C., & Venkatesh, V. (2019). ICT-enabled Refugee Integration: A Research Agenda. *Communications of the Association for Information Systems*, *44*, 40.

Acosta-Vargas, P., Guaña-Moya, J., Acosta-Vargas, G., Villegas-Ch, W., & Salvador-Ullauri, L. (2021). Method for Assessing Accessibility in Videoconference Systems. In D. Russo, T. Ahram, W. Karwowski, G. Di Bucchianico, & R. Taiar (Eds.), *Advances in Intelligent Systems and Computing. Intelligent Human Systems Integration 2021* (Vol. 1322, pp. 669–675). Springer International Publishing.

Adaloglou, N., Chatzis, T., Papastratis, I., Stergioulas, A., Papadopoulos, G. T., Zacharopoulou, V., Xydopoulos, G. J., Atzakas, K., Papazachariou, D., & Daras, P. (2022). A Comprehensive Study on Deep Learning-based Methods for Sign Language Recognition. *IEEE Transactions on Multimedia*, *24*, 1750–1762.

American Speech-Language-Hearing Association. (n.d.). *Augmentative and Alternative Communication*. Retrieved from https://www.asha.org/Practice-Portal/Professional-Issues/Augmentative-and-Alternative-Communication/

Andrade, A. D., & Doolin, B. (2016). Information and Communication Technology and the Social Inclusion of Refugees. *MIS Quarterly*, *40*(2), 405–416.

Asadikia, A., Rajabifard, A., & Kalantari, M. (2021). Systematic prioritisation of SDGs: Machine learning approach. *World Development*, *140*.

Bantupalli, K., & Xie, Y. (2018). American Sign Language Recognition using Deep Learning and Computer Vision. In *2018 IEEE International Conference on Big Data (Big Data)* (pp. 4896–4899). IEEE.

Barbosa, A. F., Cappi, J., Oyadomari, W., & Winkler, I. (2001). Electronic Government in Brazil - Measuring E-Gov Appropriation by Citizens and Enterprises. In *CONF-IRM 2011 Proceedings*.

Borg, M., & Camilleri, K. P. (2019). Sign Language Detection "in the Wild" with Recurrent Neural Networks. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1637–1641). IEEE.

Boyes-Braem, P., & Sutton-Spence, R. (2001). *The hands are the head of the mouth: The mouth as articulator in sign languages*. *International studies on sign language and communication of the deaf: v. 39*. Signum.

Brady, N. C., Bruce, S., Goldman, A., Erickson, K., Mineo, B., Ogletree, B. T., Paul, D., Romski, M. A., Sevcik, R., Siegel, E., Schoonover, J., Snell, M., Sylvester, L., & Wilkinson, K. (2016). Communication Services and Supports for Individuals With Severe Disabilities: Guidance for Assessment and Intervention. *American Journal on Intellectual and Developmental Disabilities*, *121*(2), 121–138.

Bragg, D., Koller, O., Bellard, M., Berke, L., Boudreault, P., Braffort, A., Caselli, N., Huenerfauth, M., Kacorri, H., Verhoef, T., Vogler, C., & Ringel Morris, M. (2019). Sign Language Recognition, Generation, and Translation. In J. P. Bigham, S. Azenkot, & S. K. Kane (Eds.), *The 21st International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 16–31). ACM.

Brennan, M. (2003). Deafness, Disability and Inclusion: The Gap between Rhetoric and Practice. *Policy Futures in Education*, *1*(4), 668–685.

Buck, C., Kreuzer, T., Oberländer, A., Röglinger, M., & Rosemann, M. (2022). Four Patterns of Digital Innovation in Times of Crisis. *Communications of the Association for Information Systems*, *50*(1).

Bull, H., Afouras, T., Varol, G., Albanie, S., Momeni, L., & Zisserman, A. (2021). *Aligning Subtitles in Sign Language Videos.*

Bull, H., Gouiffès, M., & Braffort, A. (2020). Automatic Segmentation of Sign Language into Subtitle-Units. In A. Bartoli & A. Fusiello (Eds.), *Lecture Notes in Computer Science. Computer Vision – ECCV 2020 Workshops* (Vol. 12536, pp. 186–198). Springer International Publishing.

Camgoz, N. C., Hadfield, S., Koller, O., Ney, H., & Bowden, R. (2018). Neural Sign Language Translation. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Camgoz, N. C., Koller, O., Hadfield, S., & Bowden, R. (2020a). *Multi-channel Transformers for Multi-articulatory Sign Language Translation.*

Camgoz, N. C., Koller, O., Hadfield, S., & Bowden, R. (2020b). Sign Language Transformers: Joint End-to-End Sign Language Recognition and Translation. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 10020–10030). IEEE.

Corbett, J., Dennehy, D., & Carter, L. (2022). Special Issue — Call for Papers: Digital Innovation for Social Development and Environmental Action. *Communications of the Association for Information Systems*.

Corbett, J., & Mellouli, S. (2017). Winning the SDG battle in cities: how an integrated information ecosystem can contribute to the achievement of the 2030 sustainable development goals. *Information Systems Journal*, *27*(4), 427–461.

Council of Europe. (n.d.). *European Day of Languages: FAQs on sign language*. Retrieved from https://edl.ecml.at/Facts/FAQsonsignlanguage/tabid/2741/language/en-GB/Default.aspx

Davydov, M., & Lozynska, O. (2017). Information system for translation into ukrainian sign language on mobile devices. In *2017 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)* (pp. 48–51). IEEE.

Dennis, A. R., Fuller, R. M., & Valacich, J. S. (2008). Media, Tasks, and Communication Processes: A Theory of Media Synchronicity. *MIS Quarterly*, *32*(3), 575–600.

Dostal, H. M., & Wolbers, K. A. (2014). Developing language and writing skills of deaf and hard of hearing students: A simultaneous approach. *Literacy Research and Instruction*, *53*(3), 245–268.

Drechsler, K., Gregory, R., Wagner, H.-T., & Tumbas, S. (2020). At the Crossroads between Digital Innovation and Digital Transformation. *Communications of the Association for Information Systems*, *47*(1).

European Union of the Deaf (Ed.). (2021). *Accessibility of Information and Communication*. Retrieved from https://www.eud.eu/eud/position-papers/accessibility-of-information-and-communication/

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. *Adaptive computation and machine learning series*. MIT Press.

Gregor, S., & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly*, *37*(2), 337–355.

Guo, D., Zhou, W., Li, H., & Wang, M. (2018). Hierarchical LSTM for Sign Language Translation. *Proceedings of the AAAI Conference on Artificial Intelligence*, *32*(1).

Hammami, S., Saeed, F., Mathkour, H., & Arafah, M. A. (2019). Continuous improvement of deaf student learning outcomes based on an adaptive learning system and an Academic Advisor Agent. *Computers in Human Behavior*, *92*, 536–546.

Harini, R., Janani, R., Keerthana, S., Madhubala, S., & Venkatasubramanian, S. (2020). Sign Language Translation. In *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)* (pp. 883–886). IEEE.

Hersh, M., Leporini, B., & Buzzi, M. (2020). Accessibility evaluation of video conferencing tools to support disabled people in distance teaching, meetings and other activities. In *ICCHP open access compendium "Future Perspectives of AT, eAccessibility and eInclusion"*.

Hevner, A. R., March, S. T., Park, J [Jinsoo], & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75.

Huang, J., Zhou, W., Zhang, Q., Li, H., & Li, W. (2018). Video-Based Sign Language Recognition Without Temporal Segmentation. *Proceedings of the AAAI Conference on Artificial Intelligence*, *32*(1).

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685–695.

Jones, M. (2002). Deafness as Culture: A Psychosocial Perspective. *Disability Studies Quarterly*, *22*(2).

Kahlon, N. K., & Singh, W. (2021). Machine translation from text to sign language: A systematic review. *Universal Access in the Information Society*(22), 1–35.

Karpov, A., Kipyatkova, I., & Zelezny, M. (2016). Automatic Technologies for Processing Spoken Sign Languages. *Procedia Computer Science*, *81*, 201–207.

Klima, E. S., & Bellugi, U. (2010). *The signs of language*. Harvard Univ. Press.

Ko, S.-K., Kim, C. J., Jung, H., & Cho, C. (2019). Neural Sign Language Translation Based on Human Keypoint Estimation. *Applied Sciences*, *9*(13), 2683.

Kranz, J., Zeiss, R., Beck, R., Gholami, R., Sarker, S., Watson, R., & Whitley, E. (2022). Practicing What We Preach? Reflections on More Sustainable and Responsible IS Research and Teaching Practices. *Communications of the Association for Information Systems*, *51*(1).

Kritzinger, J., Schneider, M., Swartz, L., & Braathen, S. H. (2014). "I just answer 'yes' to everything they say": Access to health care for deaf people in Worcester, South Africa and the politics of exclusion. *Patient Education and Counseling*, *94*(3), 379–383.

Lang, M., Freeman, M., Kiely, G., & Woszczynski, A. B. (2022). Special Issue Editorial: Equality, Diversity, and Inclusion in IS Education. *Journal of Information Systems Education*, *33*(1), 1–6.

Leal Filho, W., Tripathi, S. K., Andrade Guerra, J. B. S. O. D., Giné-Garriga, R., Orlovic Lovren, V., & Willats, J. (2019). Using the sustainable development goals towards a better understanding of sustainability challenges. *International Journal of Sustainable Development & World Ecology*, *26*(2), 179–190.

López-Ludeña, V., San-Segundo, R., Ferreiros, J., Pardo, J. M., & Ferreiro, E. (2013). Developing an information system for deaf. In *Interspeech 2013* (pp. 3617–3621). ISCA.

Mäkipää, J.-P., Norrgård, J., & Vartiainen, T. (2022). Factors Affecting the Accessibility of IT Artifacts: A Systematic Review. *Communications of the Association for Information Systems*, *51*, 666–702.

Marshall, C. R., & Morgan, G. (2015). From gesture to sign language: Conventionalization of classifier constructions by adult hearing learners of British Sign Language. *Topics in Cognitive Science*, *7*(1), 61–80.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, *5*(4), 115–133.

Ministry of the Interior of the State of NRW. (2022). *nora - Official Emergency Call App of the German Federal States*. Retrieved from https://www.nora-notruf.de/en-en/startpage

Moryossef, A., Tsochantaridis, I., Aharoni, R., Ebling, S., & Narayanan, S. (2020). *Real-Time Sign Language Detection using Human Pose Estimation.*

Mousley, V. L., & Chaudoir, S. R. (2018). Deaf Stigma: Links Between Stigma and Well-Being Among Deaf Emerging Adults. *Journal of Deaf Studies and Deaf Education*, *23*(4), 341–350.

NDIA. (2019). *National Digital Inclusion Alliance - You searched for digital inclusion*. Retrieved from https://www.digitalinclusion.org/?s=digital+inclusion

Núñez-Marcos, A., Perez-de-Viñaspre, O., & Labaka, G. (2023). A survey on Sign Language machine translation. *Expert Systems with Applications*, *213*, 118993.

Olbrich, S., Trauth, E. M., Niedermann, F., & Gregor, S. (2015). Inclusive Design in IS: Why Diversity Matters. *Communications of the Association for Information Systems*, *37*.

Ortega, G. (2017). Iconicity and Sign Lexical Acquisition: A Review. *Frontiers in Psychology*, *8*, 1280.

Padden, C., & Gunsauls, D. (2003). How the Alphabet Came to Be Used in a Sign Language. *Sign Language Studies*, *4*(1), 10–33.

Padmanabhan, B., Xiao, X., Sahoo, N., & Burton-Jones, A. (2022). Machine Learning in Information Systems Research. *MIS Quarterly*, *46*(1), iii–xix.

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). BLEU: A Method for Automatic Evaluation of Machine Translation. In *ACL '02, Proceedings of the 40th Annual Meeting on Association for Computational Linguistics* (pp. 311–318). Association for Computational Linguistics.

Park, J [Junghyun] (2020). Unraveling the Invisible but Harmful Impact of COVID-19 on Deaf Older Adults and Older Adults with Hearing Loss. *Journal of Gerontological Social Work*, *63*(6-7), 598–601.

Parton, B. S. (2006). Sign language recognition and translation: A multidisciplined approach from the field of artificial intelligence. *Journal of Deaf Studies and Deaf Education*, *11*(1), 94–101.

Patel, B. D., Patel, H. B., Khanvilkar, M. A., Patel, N. R., & Akilan, T. (2020). ES2ISL: An Advancement in Speech to Sign Language Translation using 3D Avatar Animator. In *2020 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)* (pp. 1–5). IEEE.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, *24*(3), 45–77.

Pietrabissa, G., & Simpson, S. G. (2020). Psychological Consequences of Social Isolation During COVID-19 Outbreak. *Frontiers in Psychology*, *11*, 2201.

Popović, M. (2015). chrF: character n-gram F-score for automatic MT evaluation. In *Proceedings of the Tenth Workshop on Statistical Machine Translation* (pp. 392–395). Association for Computational Linguistics.

Randolph, A. B., Petter, S. C., Storey, V. C., & Jackson, M. M. (2022). Context-aware user profiles to improve media synchronicity for individuals with severe motor disabilities. *Information Systems Journal*, *32*(1), 130–163.

Rastgoo, R., Kiani, K., Escalera, S., & Sabokrou, M. (2021). Sign Language Production: A Review. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

Recio-Barbero, M., Sáenz-Herrero, M., & Segarra, R. (2020). Deafness and mental health: Clinical challenges during the COVID-19 pandemic. *Psychological Trauma: Theory, Research, Practice, and Policy*, *12*(1), 212-213.

Renz, K., Stache, N. C., Albanie, S., & Varol, G. (2020). *Sign language segmentation with temporal convolutional networks.*

Renz, K., Stache, N. C., Fox, N., Varol, G., & Albanie, S. (2021). *Sign Segmentation with Changepoint-Modulated Pseudo-Labelling.*

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, *65 6*, 386–408.

Sanjay, Y., & Sanyam, S. (2016). Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification. In *2016 IEEE 6th International Conference on Advanced Computing (IACC)*.

San-Segundo, R., Barra, R., Córdoba, R., D'Haro, L. F., Fernández, F., Ferreiros, J., Lucas, J. M., Macías-Guarasa, J., Montero, J. M., & Pardo, J. M. (2008). Speech to sign language translation system for Spanish. *Speech Communication*, *50*(11), 1009–1020.

Schoormann, T., & Kutzner, K. (2020). Towards Understanding Social Sustainability: An Information Systems Research-Perspective. In *Proceedings of the 41st International Conference on Information Systems*.

Schoormann, T., Strobel, G., Möller, F., & Petrik, D. (2021). Achieving Sustainability with Artificial Intelligence—A Survey of Information Systems Research. In *Proceedings of the 42nd International Conference on Information Systems*.

Schoormann, T., Strobel, G., Möller, F., Petrik, D., & Zschech, P. (2023). Artificial Intelligence for Sustainability—A Systematic Review of Information Systems Literature. *Communications of the Association for Information Systems*, *52*(1).

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(3), 379–423.

Sidani, D., Veglianti, E., & Maroufkhani, P. (2022). Smart Cities for a Sustainable Social Inclusion Strategy - A Comparative Study between Italy and Malaysia. *Pacific Asia Journal of the Association for Information Systems.*, *14*(2), 3.

Skelton, T., & Valentine, G. (2003). Political participation, political action and political identities: Young d/deaf people's perspectives. *Space and Polity*, *7*(2), 117–134.

Snover, M., Dorr, B., Schwartz, R., Micciulla, L., & Makhoul, J. (2006). A Study of Translation Edit Rate with Targeted Human Annotation. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers* (pp. 223–231). Association for Machine Translation in the Americas.

Steinberg, A. G., Sullivan, V. J., & Montoya, L. A. (1999). Loneliness and Social Isolation in the Work Place for Deaf Individuals During the Transition Years: A Preliminary Investigation. *Journal of Applied Rehabilitation Counseling*, *30*(1), 22–30.

Stokoe, W. C. (2005). Sign language structure: An outline of the visual communication systems of the American deaf. 1960. *Journal of Deaf Studies and Deaf Education*, *10*(1), 3–37.

Sutton-Spence, R., & Woll, B. (1998). *The Linguistics of British Sign Language*. Cambridge University Press.

Taylor, G. (1999). Empowerment, Identity and Participatory Research: Using social action research to challenge isolation for deaf and hard of hearing people from minority ethnic communities. *Disability & Society*, *14*(3), 369–384.

United Nations. (n.d.). *United Nations Social Development Network (UNSDN)*. Retrieved from https://www.un.org/development/desa/dspd/about-us/united-nations-social-development-network-unsdn.html

United Nations. (1995). *WSSD 1995 Agreements*. Retrieved from https://www.un.org/development/desa/dspd/world-summit-for-social-development-1995/wssd-1995-agreements.html

United Nations. (2015). *A/RES/70/1—Transforming our world: The 2030 Agenda for Sustainable Development*. Retrieved from https://sustainabledevelopment.un.org

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, u., & Polosukhin, I. (2017). Attention is All You Need. In *NIPS'17, Proceedings of the 31st International Conference on Neural Information Processing Systems.* Curran Associates Inc.

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications*, *11*(1), 233.

vom Brocke, J., Winter, R., Hevner, A., & Maedche, A. (2020). Special Issue Editorial –Accumulation and Evolution of Design Knowledge in Design Science Research: A Journey Through Time and Space. *Journal of the Association for Information Systems*, *21*(3), 520–544.

Vygotskij, L. S. (Ed.). (1981). *Mind in society: The development of higher psychological processes*. Harvard Univ. Press.

Wang, X., Brooks, S., & Sarker, S. (2015). Understanding Green IS Initiatives: A Multi-theoretical Framework. *Communications of the Association for Information Systems*, *37*(1).

Watson, R. T., Elliot, S., Corbett, J., Farkas, D., Feizabadi, A., Gupta, A., Iyer, L., Sen, S., Sharda, R., Shin, N., Thapa, D., & Webster, J. (2021). How the AIS can Improve its Contributions to the UN's Sustainability Development Goals: Towards A Framework for Scaling Collaborations and Evaluating Impact. *Communications of the Association for Information Systems*, *48*(1), 476–502.

Weerakkody, V., Dwivedi, Y. K., El-Haddadeh, R., Almuwil, A., & Ghoneim, A. (2012). Conceptualizing E-Inclusion in Europe: An Explanatory Study. *Information Systems Management*, *29*(4), 305–320.

Wilson, C., & Secker, J. (2015). Validation of the Social Inclusion Scale with Students. *Social Inclusion*, *3*(4), 52–62.

Xiong, R., Yang, Y., Di He, Zheng, K., Zheng, S., Xing, C., Zhang, H., Lan, Y., Wang, L., & Liu, T.-Y. (2020). On Layer Normalization in the Transformer Architecture. In *Proceedings of the 37th International Conference on Machine Learning*.

Yin, K., & Read, J. (2020). Better Sign Language Translation with STMC-Transformer. In *Proceedings of the 28th International Conference on Computational Linguistics.*

Zeshan, U. (2006). Sign Languages of the World. In *Encyclopedia of Language & Linguistics* (pp. 358–365). Elsevier.

Zoom. (n.d.). *Accessibility*. Retrieved from https://explore.zoom.us/en/accessibility/

Appendix A: Emergency Dataset

In the following, we provide an overview of the individual phrases that our emergency dataset contains.

#ID	Phrases	#ID	Phrases
01	How old are you?	28	heart attack
02	I don't feel good. I feel lousy.	29	I appreciate your help.
03	I don't understand.	30	It is an emergency.
04	I feel OK.	31	Where is the toilet?
05	I feel tired.	32	Can I use the toilet?
06	I feel wonderful.	33	No drinking allowed.
07	I feel sick with a sore throat and an annoying headache.	34	I need a cold drink.
08	I have a question.	35	He has an internal injury.
09	Sorry.	36	He has a first degree burn
10	Thank you.	37	It is a constant pain.
11	There is no sign for that, you need to fingerspell it.	38	Can I help you?
12	What are you doing?	39	May I help you?
13	What did you/they say?	40	How can I help you?
14	What do you think?	41	I am not hearing.
15	What does that mean?	42	I am hard of hearing.
16	What is your name?	43	Are you hard of hearing?
17	What is your phone number?	44	Are you deaf or hearing?
18	What time?	45	Who?
19	Where do you live?	46	What is your address?
20	Where do you work?	47	number
21	Who do you live with?	48	floor
22	Yes, please.	49	city
23	Yes, thank you.	50	fire
24	You are welcome.	51	fire alarm
25	Nice to meet you	52	fire detector
26	robber	53	fire extinguisher
27	pharmacist	54	rescue
55	We have had a car accident.	86	He hears. He does not speak yet.
56	Did anybody see the car accident?	87	He can hear. He does not talk.
57	accident	88	hearing
58	Call the ambulance.	89	hard of hearing
59	ambulance	90	disabled
60	paramedic	91	Hey, what is your name?
61	injury	92	What is the name of your doctor?
62	finger injury	93	What is the name of your hospital?
63	ill	94	In which street are you living?
64	illness	95	What is your last name?
65	I will call the police.	96	How do you spell your last name?
66	police	97	How do you spell your name?

67	assault	98	What is the name of your partner?
68	gun	99	What is the name of your wife?
69	visiting hours	100	What is your exact address?
70	I need a doctor.	101	How many floors does this building have?
71	Where can I find a doctor?	102	In what city do you live in?
72	You should go to the doctor.	103	Who is that person?
73	doctor	104	Do you understand him/her?
74	I need a dentist.	105	Are you deaf?
75	dentist	106	Are you married?
76	ear	107	Do you have children?
77	nose	108	Do you need to go to the bathroom?
78	throat	109	Do you need to go to the doctor?
79	I have a sore throat.	110	Can you drive?
80	optometrist	111	Are you driving from here to home?
81	emergency	112	Where is your home?
82	This is an emergency!	113	How do you feel?
83	Do you want to go to the emergency room?	114	Do you want to go home now?
84	Hey! An emergency is happening right now!	115	Can you use your phone?
85	Losing my hearing	116	Do you feel sick?
117	There is a fire! Call 9-1-1	147	Are you allergic to any medicine?
118	A fire happened, call 9-1-1	148	medicine
119	Fire-alarm	149	Do you take any medicine pills?
120	Did you have an accident?	150	I suppose my blood-pressure is high, what should I do?
121	Why did the car accident happen?	151	Can you breathe?
122	Do you need an ambulance?	152	short of breath
123	sick	153	You are breathing hard.
124	The police will arrive soon.	154	calm down
125	cop	155	chest
126	Did someone break into your apartment before?	156	tight chest
127	robbery	157	cold temperature
128	criminal	158	cough
129	violence	159	dizzy
130	violent	160	Do you feel dizzy?
131	address	161	I am thirsty, where is the water?
132	You need to contact your doctor.	162	Are you thirsty?
133	See the doctor	163	Are you tired?
134	ear hurts	164	exhausted
135	bloody nose	165	tired
136	sore throat	166	vomit
137	eyes	167	eat
138	pharmacy	168	food
139	help	169	hospital
140	Help me!	170	Do you want to go to the hospital?
141	I am here to help you	171	How are you?

142	Do you need help?	172	hot
143	Show me where your pain is.	173	Are you hungry?
144	pain	174	Do you want to eat?
145	allergic	175	hungry
146	Are you allergic to anything?	176	Are you hurt?
177	Did he hurt you?	189	allergy
178	hurt	190	bandage
179	Where are you hurt?	191	bones
180	heartache	192	breathe
181	injection	193	burn
182	Do you want to lie down?	194	call 9-1-1
183	lie down	195	car accident
184	nurse	196	disease
185	take pill	197	emergency room
186	sleepy	198	nauseous
187	ache	199	seizure
188	headache	200	shot
		201	swallow

.

About the Authors

Gero Strobel is an assistant professor at the Chair of Information Systems and Software Engineering at the University of Duisburg-Essen, Germany. His research interests include the design of information systems, smart services, and ecosystems as well as the design of digital assistants in the field of healthcare. Gero's work has been published in leading IS conferences such as *ICIS* and *ECIS* as well as in corresponding journals such as *Communications of the Association for Information Systems* and *Electronic Markets*.

Thorsten Schoormann is a postdoctoral researcher at the University of Hildesheim, Germany. Thorsten's research focuses on business model innovation, design science research, and supporting (digital) tools that foster economic, ecological, and social sustainability. His work has been published in academic journals including *Journal of Management Information Systems*, *European Journal of Information Systems*, *Business & Information Systems Engineering*, *Electronic Markets*, and *Communication of the Association for Information Systems*, as well as has been presented at leading conferences such as *ICIS*, *ECIS*, and *DESRIST*.

Leonardo Banh is a research assistant and PhD student at the University of Duisburg-Essen, Germany. His research interests at the Chair of Information Systems and Software Engineering revolve around the design of information systems, artificial intelligence, and particularly generative AI.

Frederik Möller is a junior professor for information systems and data-driven enterprise at TU Braunschweig, Germany and a researcher at Fraunhofer ISST. Frederik's research explores the nature of how companies use data internally as well as in interorganizational data-sharing settings. His work has been published in academic journals, including *Business & Information Systems Engineering*, *Electronic Markets*, *Communications of the Association for Information Systems*, *IEEE Transactions on Engineering Management*, *Journal of Enterprise Information Management*, and leading IS conferences, such as *ICIS*, *ECIS*, *DESRIST*, and *HICSS*.