CERC2023

Collaborative European Research Conference

Barcelona, Spain
June 9-10, 2023
www.cerc-conf.eu

Proceedings

Editors (in alphabetical order) :
Haithem Afli
Cristina Blasi Casagran
Udo Bleimann
Dirk Burkhardt
Robert Loew
Ingo Stengel
Haiying Wang
Huiru (Jane) Zheng

ISBN 978-3-96187-021-9
https://doi.org/10.48444/h_docs-pub-518

Preface

In an era marked by unprecedented technological advancements, the 2023 Collaborative European Research Conference (CERC) convened in Barcelona, Spain, on June 9-10, 2023, as a hybrid event. This gathering underscored the imperative of interdisciplinary collaboration across Europe, bringing together researchers from diverse fields to address the multifaceted challenges and opportunities presented by rapid innovation.

The conference featured a keynote address that delved into the swift evolution of artificial intelligence (AI) and its profound societal implications. The discourse highlighted the integration of AI across various professions, emphasizing the necessity for human oversight to navigate ethical considerations and mitigate potential risks. The keynote also examined the European Union's proactive stance on AI regulation, particularly through the forthcoming AI Act, which aims to establish a robust framework for the responsible development and deployment of AI technologies.

The proceedings encompass a wide array of research contributions, reflecting the conference's commitment to fostering knowledge transfer and interdisciplinary exchange. Topics span from data processing and machine learning to e-healthcare innovations and the societal impacts of emerging technologies. Notably, discussions on AI's role in healthcare, legal frameworks, and education underscore the critical need for ethical standards and regulatory measures to ensure that technological progress aligns with societal well-being.

As we present the 2023 CERC proceedings, we extend our gratitude to Ana Mar Fernández Pasarín and Santiago Ariel Villar Arias, both from the Universidad Autónoma de Barcelona (Spain) for their exceptional work as Local Chairs. We also deeply appreciate the contributions of our esteemed Co-Chairs, whose dedication and expertise greatly enriched this event: Haithem Afli, Munster Technological University, Ireland; Udo Bleimann, Darmstadt University of Applied Sciences, Germany; Dirk Burkhardt, Software AG, Darmstadt, Germany; Robert Loew, Centre for Applied Informatics (z.a.i.), Darmstadt, Germany; Ingo Stengel, Hochschule Karlsruhe, University of Applied Sciences, Germany; Haiying Wang, Ulster University, United Kingdom; and Huiru (Jane) Zheng, Ulster University, United Kingdom. Their collective efforts were instrumental in making this conference a success.

Dr Cristina Blasi Casagran
Conference and Program Chair, CERC 2023
Barcelona, Spain, August 2023

Conference Chairs

Dr. Cristina Blasi Casagran, Universidad Autónoma de Barcelona, Spain
Ana Mar Fernández Pasarín, Universidad Autónoma de Barcelona, Spain
Santiago Ariel Villar Arias, Universidad Autónoma de Barcelona, Spain
Dr. Haithem Afli, Munster Technological University, Ireland
Prof. Dr. Udo Bleimann, Darmstadt University of Applied Sciences, Germany
Dirk Burkhardt, Software AG, Darmstadt, Germany
Dr. Robert Loew, Centre for Applied Informatics z.a.i. Darmstadt, Germany
Dr. Haiying Wang, Ulster University, UK
Prof. Huiru (Jane) Zheng, Ulster University, UK

International Scientific Committee

Jens-Peter Akelbein, Darmstadt University of Applied Sciences, Germany
Tatjana Archipov, University of Applied Sciences Karlsruhe, Germany
Lu Bai, Ulster University, United Kingdom
Raymond Bond, Ulster University, United Kingdom
Fiona Brown, Datatics Ltd, United Kingdom
Zoraida Callejas Carrión, Granada University, Spain
En-Chi Chang, AMG Digital Marketing & Design Company, Taiwan
Zhibin Chen, Durham University, United Kingdom
Simone Dogu, DB Systel GmbH, Germany
Markus Döhring, Darmstadt University of Applied Sciences, Germany
Lei Duan, Sichuan University, China
Klaus-Peter Fischer-Hellmann, Digamma GmbH, Germany
Marius Faßbender, Daimler AG, Germany
Matthias Feiner, University of Applied Sciences Karlsruhe, Germany
Orla Flynn, Galway-Mayo Institute of Technology, Ireland
Steven Furnell, University of Nottingham, United Kingdom
Jianliang Gao, Imperial College, United Kingdom
Gunter Grieser, Darmstadt University of Applied Sciences, Germany
Vic Grout, Glyndwr University, United Kingdom
Markus Haid, Darmstadt University of Applied Sciences, Germany
Joe Harrington, Munster Technological University, Ireland
Matthias Hemmje, University of Hagen, Germany
Andreas Heinemann, Darmstadt University of Applied Sciences, Germany
Bernhard Humm, Darmstadt University of Applied Sciences, Germany
Martin Knahl, University of Applied Sciences Furtwangen, Germany
Habin Lee, Brunel University, United Kingdom
Olive Lennon, University College Dublin, Ireland
Margot Mieskes, Darmstadt University of Applied Sciences, Germany
Kawa Nazemi, Darmstadt University of Applied Sciences, Germany
Rainer Neumann, University of Applied Sciences Karlsruhe, Germany
Thomas Pleil, Darmstadt University of Applied Sciences, Germany
Nacim Ramdani, Université of Orléans, France
Erwin Rauch, Free University of Bozen-Bolzano, Italy
Stefanie Regier, University of Applied Sciences Karlsruhe, Germany
Sven Rogalski, Darmstadt University of Applied Sciences, Germany
Pablo Mesejo Santiago, Granada University, Spain
Christian Schalles, DHBW Mosbach, Germany

Jan Schikofsky, LaunchMarkets Institute, Germany
Bryan Scotney, Ulster University, United Kingdom
Helen O' Shea, Munster Technological University, Ireland
Melanie Siegel, Darmstadt University of Applied Sciences, Germany
Roy Sleator, Munster Technological University, Ireland
Lina Stankovic, Strathclyde University, United Kingdom
Vladimir Stankovic, Strathclyde University, United Kingdom
Uta Störl, Darmstadt University of Applied Sciences, Germany
Bernhard Thull, Darmstadt University of Applied Sciences, Germany
Ulrich Trick, University of Applied Sciences Frankfurt, Germany
Paul Walsh, Accenture Cork, Ireland
Hui Wang, Ulster University, United Kingdom
Christoph Wentzel, Darmstadt University of Applied Sciences, Germany
Andrea Wirth, University of Applied Sciences Karlsruhe, Germany
Angela Wright, Munster Technological University, Ireland
Tuba Yilmaz Abdolsaheb, Istanbul Technical University, Turky

Table of Content

Keynote

Chapter 1

Big data, computing, modelling and privacy

Chapter 2

New initiatives and laws in the digital market

Chapter 3

Medical, Social and economic sciences : New developments

List of Authors

Keynote

Cristina Blasi Casagran[1]

[1]*Universidad Autónoma de Barcelona*

1. AI developments and emerging tech laws in the current digital era

The rapid evolution of AI is transforming industries, requiring urgent regulatory measures to address its societal impact. Professions are integrating AI tools, necessitating human oversight to mitigate risks such as misinformation, ethical breaches, and copyright violations. In Europe, the EU AI Act stands as a critical regulatory framework, building on existing legislation like the GDPR and the Digital Services Act. These measures aim to ensure transparency, accountability, and the safe use of high-risk AI systems. The emphasis is on developing standards and audits to align AI technologies with democratic values and societal trust while fostering innovation responsibly.

CERC 2023: Collaborative European Research Conference, June 09–10, 2023, Barcelona, Spain
cristina.blasi@uab.cat (C. Blasi Casagran)
0000-0002-4327-2212 (C. Blasi Casagran)

ISBN 978-3-96187-021-9
https://doi.org/10.48444/h_docs-pub-518

Chapter 1

Big data, computing, modelling and privacy

FIT4NER - Towards a Framework-Independent Toolkit for Named Entity Recognition

Florian Freund[a], Philippe Tamla[a], Thoralf Reis[a], Matthias Hemmje[a] and Paul Mc Kevitt[b]

[a]University of Hagen, Faculty of Mathematics and Computer Science, 58097 Hagen, Germany
[b]Ulster University, Faculty of Arts, Humanities and Social Sciences, Londonderry BT48 7JL, North Ireland

Abstract

This paper proposes a Framework-Independent Toolkit for Named Entity Recognition (FIT4NER) to support medical domain experts during extraction of named entities in natural language text. Named Entity Recognition (NER) is an important technique for handling information overload, as it helps to extract relevant information from Electronic Health Records, which often contain free text notes. FIT4NER enables experts to compare multiple NER frameworks, choose the best fit, and overcome computing resource limitations through the use of Cloud-based resources. The paper introduces the context, discusses the current state of the art, and identifies the remaining challenges. A User-Centered System Design approach is used to model the solution, and the models are evaluated through Pattern-Based Architecture Reviews. The evaluation strategies for both the models and the system are discussed, and a conclusion of this work is presented.

Keywords

Natural Language Processing, Named Entity Recognition, Medical Expert Systems, Clinical Decision Support, Cloud-based Resources

1. Introduction and Motivation

This research work aims to use **Named Entity Recognition (NER)** to address the challenge of **Information Overload (IO)** in healthcare. By enabling medical experts to efficiently analyze large volumes of medical texts, **Information Retrieval (IR)** in the medical field will be improved through the use of NER. NER is a **Natural Language Processing (NLP)** technique that aims at extracting **Named Entities (NEs)** from unstructured text documents [1]. In the medical domain, NER can be used *"to analyze and categorize medical conditions, namely symptoms, treatments, diseases, and body conditions"* in **Electronic Health Records (EHRs)** [2, p.151]. Extracting NEs from medical documents is challenging due to the variety and complexity of medical information [3] that is usually stored in natural text documents, such as EHRs [2]. Techniques for NER are generally divided into rule-based, **Machine Learning (ML)**-based, and hybrid approaches involving both [1]. ML-based NER techniques have become very popular because of their ability to efficiently deal with unstructured natural language text, making them one of the preferred methods based on **Artificial Intelligence (AI)** [1]. Most common ML techniques include supervised, unsupervised, and semi-supervised learning. Supervised learning is the most used method, which relies on manually annotated data for

CERC 2023: Collaborative European Research Conference, September 09–10, 2023, Barcelona, Spain
florian.freund@fernuni-hagen.de (F. Freund); philippe.tamla@fernuni-hagen.de (P. Tamla); thoralf.reis@fernuni-hagen.de (T. Reis); matthias.hemmje@fernuni-hagen.de (M. Hemmje); p.mckevitt@ulster.ac.uk (P.M. Kevitt)
https://www.fernuni-hagen.de/multimedia-internetanwendungen/en/ (F. Freund)
0000-0002-7344-6869 (F. Freund); 0000-0002-0786-4253 (P. Tamla); 0000-0003-1100-2645 (T. Reis); 0000-0001-8293-2802 (M. Hemmje); 0000-0001-9715-1590 (P.M. Kevitt)

ISBN 978-3-96187-021-9
https://doi.org/10.48444/h_docs-pub-518

model training [4, 5], unsupervised learning relies only on statistical algorithms to learn patterns from unlabeled data [5]. Semi-supervised learning combines both approaches by learning using a small set of annotated data [5]. Using supervised learning, a custom model needs to be trained to obtain good results in an applied knowledge domain. A typical workflow for creating a ML model includes several experimental steps that are often part of a pipeline, such as data preparation, model training, validation, and evaluation [6]. These steps may require a combination of technology and coding to prepare and generate the training data, train the model, and fine-tune different ML parameters to improve the model performance. This may be challenging for domain experts without software and ML expertise. After outlining the fundamental concepts of NLP, NER, and ML as essential techniques for effective IR in the medical field, this research work is now motivated by the following research projects.

In 2018, **Recommendation Rationalisation (RecomRatio)** was launched with the aim of assisting medical professionals in making informed decisions by leveraging ML-based NER with the assistance of the spaCy framework to extract emerging knowledge from medical literature [7, 8]. **Stanford Named Entity Recognition and Classification (SNERC)**, another project, seeks to aid developers in the applied games field to retrieve information from text documents by utilizing ML-based NER methods from the Stanford CoreNLP framework [9]. Both projects compared NER frameworks and selected a suitable framework to implement their system. However, with the evolution of NER, new and improved techniques, such as Transformer-based NER or NER via Large Language Models [10], have been introduced, leading to several challenges. First, integrating new NER frameworks into existing projects can be challenging due to differences in Application Programming Interfaces or updated NER techniques. Second, it is crucial to compare available NER frameworks for a project and choose the one that best fits the task. Third, previous systems have encountered difficulties when training NER models with large datasets, making a scalable approach essential.

The main objective of this research is to develop a flexible NER system that can efficiently analyze medical texts and provide medical experts with necessary tools to overcome the challenges faced during the process, such as selecting, comparing new NER frameworks and scalability for large datasets. In order to achieve this goal two **Research Questions (RQs)** can be defined. The architecture of an **Information Systems (ISs)** significantly affects the quality attributes of flexibility and scalability as outlined in our main research objective [11]. Therefore, determining the suitable architecture is crucial in addressing these quality attributes in the ISs. As such, the first research question is: *RQ 1: How can the architecture of a scalable and distributed IS be designed to utilize evolving NER frameworks, including the ability to compare and select the appropriate framework?* Once a potential architecture has been formulated, the next step is to verify if the defined quality attributes can be achieved. This leads us to our next research question: *RQ 2: How can such an architecture be evaluated?*

Following Nunamaker's research method [12], this paper organizes the **Research Objectives (ROs)** needed to answer the RQs into several phases, including observation, theory building, system development, and experimentation. RO 1, which involves the review of the state-of-the-art related work, as part of the observation phase, is covered in section 2. The required models, including workflow design and **Use Cases (UCs)**, will be created as part of the theory building in RO 2. The development of a general architecture for an IS, which belongs to the system development phase, is addressed in RO 3 and is covered in section 3 along with RO 2. The final RO 4, which is to experiment with the developed architecture and discuss the evaluation approach for the presented models, belongs to the experimentation phase and is discussed in section 4. Finally, section 5 presents the conclusion.

2. State of the Art in Science and Technology

Artificial Intelligence for Hospitals, Healthcare & Humanity (AI4H3) [13] proposes the use of the KlinGard Smart Hospital EcoSystem Portal, an IS based on the **Content and Knowledge Management Ecosystem Portal (KM-EP)**, a **Knowledge Management System (KMS)**. The KM-EP was developed at the University of Hagen in the Faculty of Mathematics and Computer Science, Chair of Multimedia and Internet Applications [14], and the FTK e.V. Research Institute for Telecommunications and Cooperation [15]. It was successfully used to address IO in multiple research projects, including the **EU Horizon 2020 (EU H2020)** projects **Metaplat** [16], **SenseCare** [17], and RecomRatio [7]. In Metaplat [16], KM-EP was utilized to manage large volumes of data in the genomic research domain and provide the necessary software infrastructure for knowledge management. SenseCare, an EU H2020 project in the medical field, leverages KM-EP *"to capture, analyze, and store information on emotional outputs in the aim of providing effective tools for caregivers and medical professionals to provide more holistic care to people with dementia"* [17, p.2682]. Another related project is RecomRatio [7], which aims to provide medical experts with evidence-based textual arguments in medical literature to support decision-making. Evidence is collected from relevant corpora, to provide arguments for, or against, specific treatments explicit in a knowledge base. Nawroth utilizes NLP, NER, and ML in his work [8] to make emerging knowledge from those corpora available for evidence-based medical argumentation UCs. SNERC [9], which is based on KM-EP, is also closely related to this research. SNERC enables users in the applied gaming domain to train custom NER models using Stanford CoreNLP. Its recent evaluation showed that it's an effective approach for NER support but requires more flexibility, such as supporting different NER frameworks and data formats [18]. To optimize SNERC, support for multiple NER frameworks and data formats is necessary. This work is embedded into the ongoing **Cloud-based Information Extraction (CIE)** project, which deals with the hub architecture of AI4H3 and aims to enhance the scalability of ML-based NER for medical professionals [19]. Specifically, the project leverages cloud services to optimize the performance NER training with improved usability. Many systems have been proposed to help domain experts train and customize their ML-based NER models [9, 20, 21]. However, existing methods are often limited in the usage of a single framework supporting NER tasks. For instance, Magnini et al. [20] and Dernoncourt et al. [21] focus on the automatic generation of training sets. Magnini et al. [20] implemented TextPro [22], a framework for NER based on the Java and C++ programming languages. TextPro was used to generate training data from the domains of news, sports, social media, and educational texts. Dernoncourt et al. [21] combined an Artificial Neural Network system with a graphical user interface based on Brat Rapid Annotation Tool [23] to help domain experts generate and customize their training based on their own defined labels. Their system also relies on TensorFlow and scikit-learn for training custom models using Python. Most of the recent approaches have helped users experiment with a single NER framework which is often limited to a small set of programming languages. Supporting the integration of multiple NER frameworks helps domain experts (such as clinicians) to efficiently extract and operate on their medical data through an easy selection and comparison of various state-of-the-art tools and methods enabling NER. As a result, a **Remaining Challenge (RC) 1** can be stated as follows: *Domain experts need the ability to experiment with various NER frameworks, compare their features and performance, and select the framework that best suits their needs.*

The performance of a ML-based NER model is highly impacted by the volume of the training data. Insufficient training data can lead to poor model performance as it *"will not be able to generalize and the results on unseen data can be catastrophic"* [1, p.22]. Furthermore, training custom NER models on very large datasets can be time-consuming and strain computational resources, potentially compromising the model's quality [4, 24]. Despite the advancements in Transformer-based models, which use at-

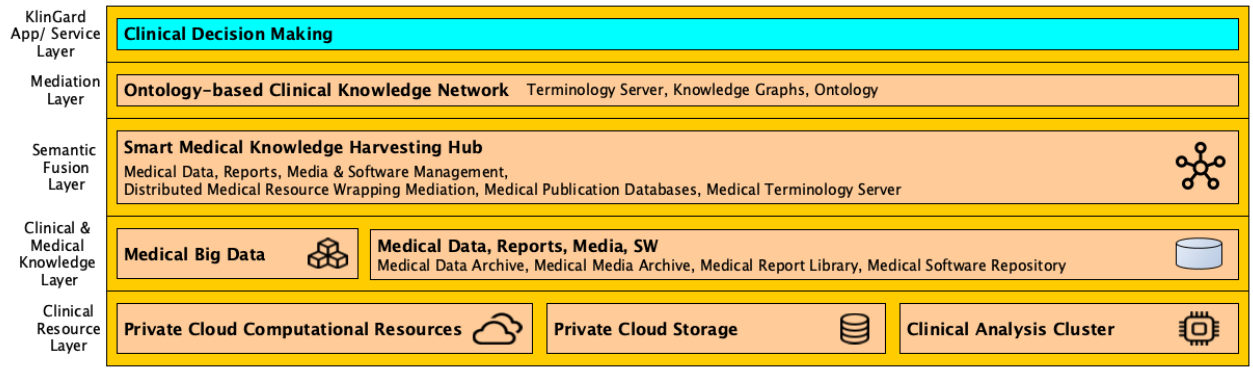


Figure 1: AI4H3 target reference architecture [13]

tention mechanisms to improve quality and reduce training time [25], fine-tuning NER models on big data can still be challenging and may require more computing resources than what a domain expert's personal computer can provide [4]. Utilizing Cloud-based resources offers a solution to the problem of computational limitations, as they provide a vast amount of dynamically available computing power. Managing Cloud resources, such as compute power, storage, and memory, can be challenging for domain experts because it requires expertise in Cloud computing, as well as a deep understanding of the underlying infrastructure and software tools [26]. This includes understanding different Cloud service models (Infrastructure as a Service, Platform as a Service, Software as a Service), setting up platforms (Kubernetes, CloudStack), and being familiar with established vendors Amazon Web Services, Microsoft Azure, Oracle, Google) [27]. The use of commercial Clouds presents a challenge in effectively managing data processing costs. Naive users may struggle to take into account various factors such as instance types, availability zones, and pricing models while provisioning their Cloud resources [28]. Improper allocation of Cloud resources can lead to inefficiencies and increased costs [28]. Hence, RC 2 states: *It is imperative to provide support to make it easier to access and manage Cloud resources and related costs, as different NER tasks may have different requirements.*

It is also anticipated that NER techniques will continue to evolve, and new methods will emerge, given the past and ongoing development in NER. Relying solely on existing NER frameworks is not enough. Instead, it is crucial to allow the use of newly developed tools to improve the performance of the system by taking advantage of new and advanced features in the market. This requires a toolkit that is adaptable to future developments, leading to RC 3: *An integrated and distributed IS is necessary to support the continuous evolution of underlying NER frameworks.* Therefore, the term **Framework-Independent Toolkit for Named Entity Recognition (FIT4NER)** was coined for such an IS.

Developing ISs that support users applying NER requires a flexible architecture and a sound understanding of relevant user stereotypes, their capabilities, and their needs. Reference models standardize users, artifacts, and their relationships and thus provide a sound basis for modeling a new toolkit [29]. AI2VIS4BigData is a reference model for ISs that apply AI for analyzing Big Data. It defines several user stereotypes that cover the whole lifecycle of AI models and provides a standardized reference architecture and implementation [29]. A promising basis for the architecture of FIT4NER was discussed in AI4H3 [13], which suggests utilizing AI techniques for automatically supporting the processing of patients' medical records and medical data, including providing transparent explanations for medical decisions. This should be achieved through a software platform called *KlinGard*, with a multi-layered architecture as shown in Figure 1. AI4H3 proposes a solution to the lack of stan-

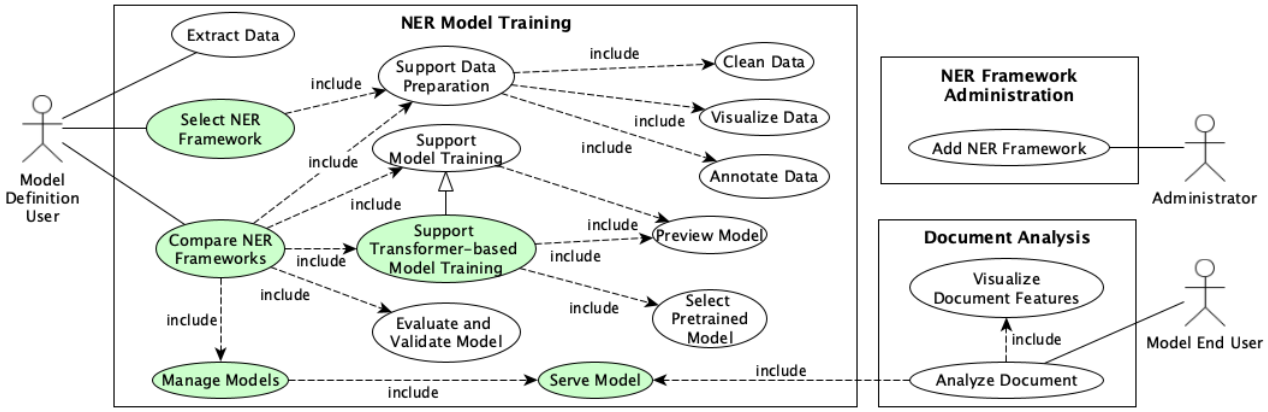


Figure 2: FIT4NER Use Cases

dardization in state-of-the-art AI methods by creating a *"Semantic Fusion Layer"* with a central hub, known as the *"KlinGard Smart Medical Knowledge Harvesting Hub"*. This hub will serve as a point for registering AI modules and facilitating communication among them (Figure 1). The proposed AI4H3 architecture suits FIT4NER, as it allows the integration of different technologies and keeps heterogeneous datasets and AI modules in a distributed design. This facilitates the utilization of multiple NER frameworks, providing the user with the ability to compare and choose the best fit for their needs.

This section discussed the state-of-the-art related work. The next section discusses the design and conceptual modeling of the FIT4NER prototype which addresses all RCs identified in this research.

3. Design and Conceptual Modeling

This aim of this work is to empower medical domain experts to analyze texts through the use of modern ML-based NER frameworks. The proposed solution, FIT4NER, is discussed in the following sections, designed using the **User Centered System Design (UCSD)** approach by Norman and Draper [30]. First, the UCs are defined. Second, the workflow model will be discussed, and, finally, an architecture to support those UCs is provided.

3.1. Use Cases

This section presents the UCs for FIT4NER. The actors are defined using the stereotypes introduced by Reis et al. in the AI2VIS4BigData reference model [29]. The *"Model Definition User"* stereotype in FIT4NER can refer to either a "Model Designer User" or a "Domain Expert User" as defined in AI2VIS4BigData. A Model Designer User is an expert in AI model design, implementation, and training, but lacks domain-specific knowledge. A "Domain Expert User" is an expert in the relevant knowledge domain but lacks technical knowledge in NER. They do not have programming skills or experience in AI but have a deep understanding of their domain (e.g., healthcare) and can identify and label relevant NEs in texts. The *"Model End User"* stereotype interacts with FIT4NER using the trained ML models and may not have deep domain knowledge. This category of users can include medical personnel, patients, and their relatives who use the IS to gain knowledge about treatments and diagnoses. They can examine stored documents to detect NEs, such as diseases or cures. Finally, the *"Administrator"* stereotype specific to FIT4NER is responsible for managing the IS's technical aspects. Figure 2 shows the UC diagram divided into three clusters, with NER-related UCs from previous research [9]

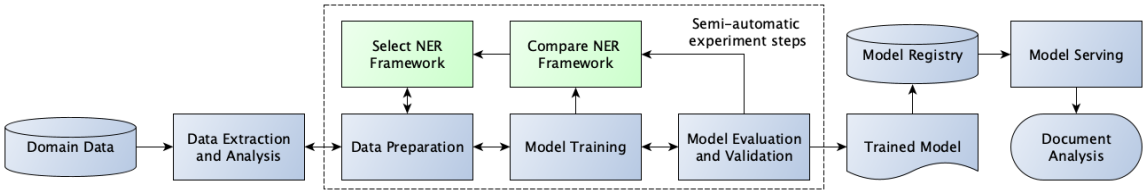


Figure 3: FIT4NER Workflow

extended to support multiple ML-based NER frameworks. The extensions are shown in green. The first cluster **"NER Model Training"** focuses on Model Definition Users and displays FIT4NER's key use cases. Model Definition Users can select a NER framework after extracting the training data. This includes features to help prepare the data for the selected framework, such as visualization, annotation, and cleaning of the data. To evaluate NER frameworks' performance (e.g., accuracy, training time, scalability) in custom knowledge domains, FIT4NER provides support for model training, including access to Cloud-based resources. Model Definition Users should have access to select parameters for NER frameworks, such as compatibility with the system (e.g., programming language, platform integration), support for data type, data protection level, and cost. FIT4NER should also support training with Transformer-based models, including the selection and fine-tuning of pre-trained models. Features for previewing the model with a limited set of training data were added, as seen in previous work [9]. Once the model is trained, Model Definition Users can evaluate and validate it. If the results are unsatisfactory, they can repeat the training process. Models must be managed and served for *"Document Analysis"* UCs. Model Definition Users can refer to the framework-specific features they used during the training process and assess the model's performance using standard metrics like Precision, Recall, and F-Score, or evaluate the models manually using domain-specific documents for comparison of NER frameworks. The "Document Analysis" cluster of UCs applies the NER models to documents, detects named entities, and visualizes document features. The last cluster of UCs, *"NER Framework Administration"*, involves extending FIT4NER by adding a new NER framework. After defining the UCs, the workflow model for FIT4NER is presented in the next section.

3.2. Workflow Model

There are different workflows for ML tasks, which can also be used for ML-based NER [31]. Typically, they follow an iterative process, where data and parameters are prepared, a model is trained and evaluated, and the process repeats if the model's performance does not meet the desired criteria. This research leverages Google's MLOps manual Workflow Model [6] for training custom ML models for NER. To empower domain experts in choosing the optimal NER framework, the model has been extended to include steps to compare and select NER frameworks. The extended workflow model is shown in Figure 3. During data preparation tasks, e.g., data cleanup, data annotation, or defining parameters for model training, domain experts can select a specific NER framework. This can impact the necessary data format and training parameters. After selecting the NER framework, the next step is to train and evaluate the model. Domain experts have the option to repeat the process with a different framework and compare their performance and features such as supported parameters and data formats. This cycle may run multiple iterations until a satisfactory NER framework and resulting model are selected. The architecture supporting the workflow and UCs is discussed in the following.

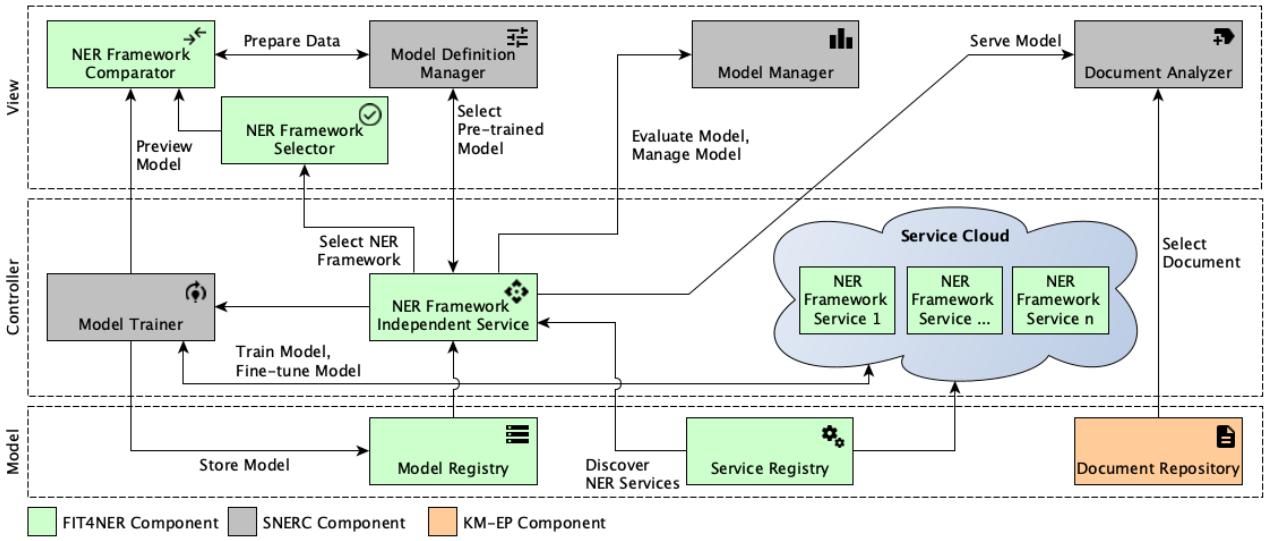


Figure 4: FIT4NER General Architecture

3.3. FIT4NER Architecture

This study is related to the medical field and AI4H3 [13], as introduced in section 2. FIT4NER should support future advancements in ML-based NER, allowing domain experts to experiment with different methods in their specific domain, compare results, and select the best approach for their task. To achieve these goals, a prototype for a Smart Medical Knowledge Harvesting Hub will be designed, as outlined in the Semantic Fusion Layer of the AI4H3 architecture (Figure 1) [13]. KM-EP will simulate the features of other layers, such as the Clinical & Medical Knowledge Layer, Mediation Layer, and KlinGard App/Service Layer. FIT4NER's architecture was developed based on the **Model-View-Controller (MVC)** pattern [32], using SNERC's architecture [9] as a starting point. It was extended to support the proposed UCs and workflow model outlined in sections 3.1 and 3.2. The proposed architecture for FIT4NER is shown in Figure 4, with additional components displayed in green. The orange component, "Document Repository", is provided by KM-EP. The proposed architecture includes the *"NER Framework Selector"* component on the View layer. This component provides a list of all registered NER frameworks and allows the selection of one or multiple frameworks. The selected frameworks are retrieved by the *"NER Framework Independent Service"* on the Controller Layer and compared by the *"NER Framework Comparator"*, which includes features to visualize differences in NER framework performance and features. The "Model Definition Manager" is responsible for setting the parameters for model training using the chosen NER framework. In case a Transformer-based model needs fine-tuning, the Model Definition Manager is provided a list of available pre-trained models by the NER Framework Independent Service. The NER Framework Independent Service acts as a central interface and layer of abstraction for all connected NER services, serving as a prototype for the Smart Medical Knowledge Harvesting Hub as outlined by AI4H3 [13]. It accesses various NER framework services from a service Cloud through a "Service Registry" component on the Model layer. The "Service Cloud" is a collection of multiple and diverse NER services provided by various NER frameworks or Cloud-based services. These services are used for training and executing NER models with a specific NER framework. To ensure scalability and dynamic resource allocation, it is recommended to implement the Service Cloud using a Cloud-based approach. The NER Framework Independent Service obtains parameters for model training from the Model Definition Manager and passes them

to the "Model Trainer" component. This component selects the necessary NER services through the NER Framework Independent Service and utilizes them directly. The NER Framework Comparator component receives a preview of the trained model using a small dataset. This allows domain experts to evaluate the expected outcome and make adjustments with a quick turnaround time. If satisfied, they initiate model training with the full dataset. The result is then stored in the "Model Registry" for future management and evaluation through the "Model Manager" component on the View layer The trained model is used by the "Document Analyzer" component to analyze documents from the "Document Repository" component provided by KM-EP. The next section discusses evaluation of the UCs, workflow model, and architecture.

4. FIT4NER Evaluation

The evaluation of FIT4NER's workflow, UCs, and architecture was performed using the **Pattern-Based Architecture Reviews (PBARs)** guidelines [11], a lightweight evaluation methodology for software architecture assessment. It was set up by planning the necessary resources. The PBAR recommends that reviewers should *"have expertise in architecture, architecture patterns, quality attributes, and a general knowledge of the domain"* [11, p.67]. The evaluation was conducted by two post-docs with expertise in NER and ML. Reviewer R1 has experience in medical domain research, and reviewer R2 is specialized in document classification and Cloud-architectures. Both reviewers were provided access to the FIT4NER models in advance for preparation, and review sessions were scheduled.

The PBAR sessions were held with each post-doc separately. Both sessions identified important quality attributes of the system, focusing on its extensibility and flexibility to accommodate different knowledge domains and future NER advancements. The system's scalability to handle large data for state-of-the-art NER was also considered. The detailed discussion included FIT4NER workflow model, UCs, and MVC-based architecture. During the meeting with Reviewer R1, the proposed models were evaluated to assess their ability to detect **emerging Named Entities (NEs)**, aligned with the expert's prior work [8]. The lack of support for Transformer-based NER frameworks was noted, leading to improvements in the UCs and architecture to address this. Thus, the UC "Support Transformer-based Model Training" was added to Figure 2 and a link between the components "Model Definition Manager" and "NER Framework Independent Service" was included in Figure 4 to allow the selection of pre-trained models. A second PBAR session was held with Reviewer R2, with a focus on verifying the proposed architecture's ability to support **Information Extraction (IE)** using Cloud computing and storage resources. The second session highlighted the need to have a better understanding of the new components that offer innovative capabilities for effective IE in KM-EP. This includes features for selecting and comparing NER frameworks in KM-EP and Cloud services to integrate new NER frameworks. The architecture was improved by using different colors in Figure 4 to distinguish between existing KM-EP components (orange), SNERC components (dark gray), and the newly added functions (green and light green). A comprehensive evaluation strategy will be outlined in future work, but the main ideas for the next evaluation steps are now shared. As soon as a working prototype is created, a functional evaluation of the integrated NER frameworks can be performed. Thus, a widely recognized gold-standard corpus, such as Colorado Richly Annotated Full Text [33], can be utilized to demonstrate the successful integration using standard evaluation metrics (Precision, Recall, F-Score). The effectiveness of the IS should be determined through a user study with the target user group. The study should assess if FIT4NER satisfies the research objective and aids healthcare professionals in effectively using ML-based NER to extract knowledge from free-text medical documents. After sharing the evaluation strategy, the following section concludes and covers future work.

5. Conclusion and future work

Clinicians in the medical field struggle with IO as crucial information is frequently concealed in vast amounts of free text notes in EHRs. It is a significant challenge to equip them with intelligent ISs to aid them in finding the information necessary to treat patients promptly. Techniques such as NER can be employed to extract knowledge from medical data, but they are limited to experts in the field The proposed IS, FIT4NER, empowers medical domain experts to extract knowledge from medical texts using NER. By integrating into a KMS and adopting the architecture of AI4H3, FIT4NER allows for the use of multiple, state-of-the-art ML-based NER frameworks for analysis and comparison. The NER Framework Independent Service component acts as a prototype for the KlinGard Smart Medical Knowledge Harvesting Hub in FIT4NER. Additionally, FIT4NER leverages Cloud resources to overcome computing limitations and optimize analysis. This paper presented the approach's context, discussed the state-of-the-art and related work, identified the RCs, and presented developed models including the workflow, UCs, and architecture. The completed evaluation steps and outlook for future tasks were also discussed. In future work, an expert survey will be conducted to identify the crucial selection parameters a system should support for comparing and selecting NER frameworks. After that, a working FIT4NER prototype can be developed as an experimental environment to test if the approach addresses the identified challenges.

References

[1] I. M. Konkol, Named Entity Recognition, Ph.D. thesis, University of West Bohemia, 2015.
[2] N. S. Pagad, N. Pradeep, Clinical named entity recognition methods: An overview, in: International Conference on Innovative Computing and Communications, Springer, 2022, pp. 151–165.
[3] Z. Liang, J. Chen, Z. Xu, Y. Chen, T. Hao, A Pattern-Based Method for Medical Entity Recognition From Chinese Diagnostic Imaging Text, Frontiers in Artificial Intelligence 2 (2019).
[4] J. Li, A. Sun, J. Han, C. Li, A Survey on Deep Learning for Named Entity Recognition, IEEE Transactions on Knowledge and Data Engineering 34 (2022) 50–70.
[5] Ariruna Dasgupta, Asoke Nath, Classification of Machine Learning Algorithms, Figshare (2016).
[6] Google, MLOps: Continuous delivery and automation pipelines in machine learning, https://Cloud.google.com/architecture/mlops-continuous-delivery-and-automation-pipelines-in-machine-learning, 2020.
[7] B. University, RecomRatio, http://ratio.sc.cit-ec.uni-bielefeld.de/projects/recomratio/, 2017.
[8] C. Nawroth, Supporting Information Retrieval of Emerging Knowledge and Argumentation, Ph.D. thesis, FernUniversität in Hagen, Hagen, 2020.
[9] P. Tamla, F. Freund, M. Hemmje, Supporting Named Entity Recognition and Document Classification for Effective Text Retrieval, in: The Role of Gamification in Software Development Lifecycle, IntechOpen, 2021, p. 24. doi:10.5772/intechopen.95076.
[10] S. Wang, X. Sun, X. Li, R. Ouyang, F. Wu, T. Zhang, J. Li, G. Wang, Gpt-ner: Named entity recognition via large language models, arXiv preprint arXiv:2304.10428 (2023).
[11] N. Harrison, P. Avgeriou, Pattern-Based Architecture Reviews, IEEE Software 28 (2011) 66–71.
[12] J. F. Nunamaker Jr, M. Chen, T. D. M. Purdin, Systems Development in Information Systems Research, Journal of Management Information Systems 7 (1990) 89–106.
[13] Artificial Intelligence for Hospitals, Healthcare & Humanity (AI4H3), R&D White Paper, FTK, Dortmund, Germany, 2020.
[14] Chair of Multimedia and Internet Applications, http://www.lgmmia.fernuni-hagen.de, 2023.

[15] FTK e.V. Research Institute for Telecommunications and Cooperation, http://www.ftk.de, 2023.
[16] B. Vu, Y. Wu, H. Afli, P. Mc Kevitt, P. Walsh, F. Engel, M. Fuchs, M. Hemmje, A metagenomic content and knowledge management ecosystem platform, in: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2019, pp. 1–8.
[17] R. Donovan, M. Healy, H. Zheng, F. Engel, B. Vu, M. Fuchs, P. Walsh, M. Hemmje, P. Mc Kevitt, Sensecare: Using automatic emotional analysis to provide effective tools for supporting, in: 2018 IEEE International Conference on Bioinformatics and Biomedicine, IEEE, 2018, pp. 2682–2687.
[18] P. Tamla., F. Freund., M. Hemmje., P. M. Mc Kevitt., Evaluation of a system for named entity recognition in a knowledge management ecosystem, in: Proceedings of the 14th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management - KEOD„ INSTICC, SciTePress, 2022, pp. 19–31. doi:10.5220/0011374000003335.
[19] P. Tamla, B. Hartmann, N. Nguyen, C. Kramer, F. Freund, CIE: A Cloud-Based Information Extraction System for Named Entity Recognition in AWS, Azure, and Medical Domain (2023).
[20] B. Magnini, A.-L. Minard, M. R. H. Qwaider, M. Speranza, TextPro-AL: An Active Learning Platform for Flexible and Efficient Production of Training Data for NLP Tasks, COLING (2016).
[21] F. Dernoncourt, J. Y. Lee, P. Szolovits, NeuroNER: An easy-to-use program for named-entity recognition based on neural networks, 2017.
[22] E. Pianta, C. Girardi, R. Zanoli, The TextPro tool suite, in: Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08), European Language Resources Association (ELRA), Marrakech, Morocco, 2008, p. 5.
[23] P. Stenetorp, S. Pyysalo, G. Topić, T. Ohta, S. Ananiadou, J. Tsujii, BRAT: A web-based tool for NLP-assisted text annotation, in: Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics, 2012, pp. 102–107.
[24] S. Doyen, N. B. Dadario, 12 Plagues of AI in Healthcare: A Practical Guide to Current Issues With Using Machine Learning in a Medical Context, Frontiers in Digital Health 4 (2022) 765406.
[25] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017).
[26] M. Vouk, Cloud Computing – Issues, Research and Implementations, Cloud Computing (2008).
[27] W. Kim, Cloud Computing: Today and Tomorrow, CLOUD COMPUTING 8 (2009) 8.
[28] R. Chard, K. Chard, K. Bubendorfer, L. Lacinski, R. Madduri, I. Foster, Cost-Aware Elastic Cloud Provisioning for Scientific Workloads, in: 2015 IEEE 8th International Conference on Cloud Computing, IEEE, New York City, NY, USA, 2015, pp. 971–974. doi:10.1109/CLOUD.2015.130.
[29] T. Reis, M. X. Bornschlegl, M. L. Hemmje, AI2vis4bigdata: A reference model for AI-based big data analysis and visualization, in: T. Reis, M. X. Bornschlegl, M. Angelini, M. L. Hemmje (Eds.), Advanced Visual Interfaces. Supporting Artificial Intelligence and Big Data Applications, volume 12585, Springer International Publishing, 2021, pp. 1–18.
[30] D. A. Norman, S. W. Draper, User Centered System Design; New Perspectives on Human-Computer Interaction, L. Erlbaum Associates Inc., USA, 1986.
[31] J. Françoise, B. Caramiaux, T. Sanchez, Marcelle: Composing Interactive Machine Learning Workflows and Interfaces, in: The 34th Annual ACM Symposium on User Interface Software and Technology, ACM, Virtual Event USA, 2021, pp. 39–53. doi:10.1145/3472749.3474734.
[32] A. Kılıçdağı, H. İ. Yilmaz, Laravel Design Patterns and Best Practices: Enhance the Quality of Your Web Applications by Efficiently Implementing Design Patterns in Laravel, 2014.
[33] K. B. Cohen, K. Verspoor, K. Fort, C. Funk, M. Bada, M. Palmer, L. E. Hunter, The Colorado Richly Annotated Full Text (CRAFT) Corpus: Multi-Model Annotation in the Biomedical Domain, in: N. Ide, J. Pustejovsky (Eds.), Handbook of Linguistic Annotation, Springer Netherlands, Dordrecht, 2017, pp. 1379–1394. doi:10.1007/978-94-024-0881-2_53.

CRISP4BigData: Qualitative evaluation of a cross-industry process reference model supporting Big Data analysis in virtual research environments

Martin Bley[1], Kevin Berwind[1] and Matthias Hemmje[1]

[1]*Chair of Multimedia and Internet Applications / University of Hagen*

Abstract

This paper aims to present, firstly, the multi-published Cross Industry Standard Process for Big Data (CRISP4BigData) based on the Big Data Management Reference Model (BDMcube), the IVIS4BigData Reference Model, and the Cross Industry Standard Process for Data Mining (CRISP-DM), and secondly, the Qualitative Evaluation of the abovementioned process model and the associated software solution. CRISP4BigData could be used as a project reference model as well as a process reference model for Big Data project processing and analysis. The corresponding graphical user interface (CRISP4BigData UI) supports and guides the end user through the project and analysis process and allows to manage and archive the corresponding scientific resources.

Keywords

CRISP4BigData, Big Data, Big Data Analytics, Big Data Process, Big Data Analytics Process, Cross Industry Standard Process for Big Data, IVIS4BigData, Reference Model for Big Data Management, BDMcube, Virtual Research Environments, CRISP- DM, Cross Industry Standard Process for Data Mining, Business Intelligence, Project Management, Process Management, Software Evaluation, Process Evaluation

1. Introduction

Every year, the volume of data in corporate data centers increases by 35 to 50 percent [2]. In addition to the expansion of computing capacities, especially in computation, new, far-reaching breakthroughs in information and sensor technology will promote intensive and analytical methods. The mass data collected comes from internal and external businesses and consists of unstructured data such as text data, processing information, (database) tables, images, videos [2], emails, feeds, and sensor data [11].

Big Data defines the collection of decision-relevant big data. Originally, the term was defined by companies that had to deal with a large amount of rapidly growing data. Gartner Inc, an information technology research and consulting firm, defines the term as follows: "*Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making*" [12].

In contrast the market research company IDC defines Big Data as follows: "*Big Data technologies as a new generation of technologies and architectures designed to economically extract value*

CERC 2023: Collaborative European Research Conference, September 09–10, 2023, Barcelona, Spain
martin.bley@studium.fernuni-hagen.de (M. Bley); kevin.berwind@fernuni-hagen.de (K. Berwind); matthias.hemmje@fernuni-hagen.de (M. Hemmje)

ISBN 978-3-96187-021-9
https://doi.org/10.48444/h_docs-pub-518

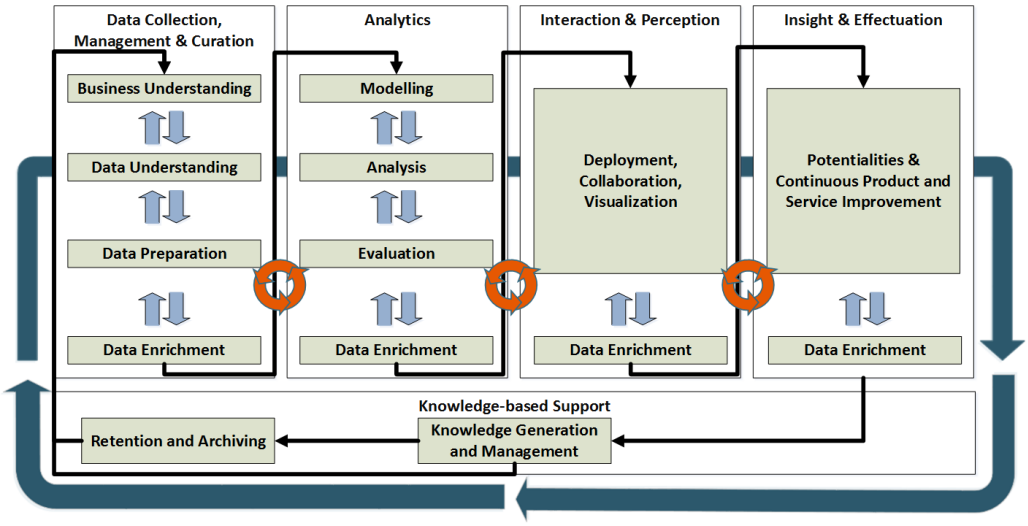


Figure 1: CRISP4BigData Reference Model Version 1.0 [3]

from very large volumes of a wide variety of data by enabling high-velocity capture, discovery, and/or analysis" [8]. Some companies had the pressure to deal with such Big Data because their business is fundamentally built on collecting and analyzing this large amount of data.

This means that, in addition to their core business, companies must also take care of technological and functional use cases to expand their competitive advantage [1]. To do this, they need the appropriate technical personnel, in this case data scientists, among others, to be able to solve these tasks [10].

The Cross Industry Standard Process for Big Data (abbreviated CRISP4BigData) reference model presented below is intended to provide companies with a framework for formulating the requirements for an analysis process and for carrying it out to be able to highlight new insights for companies and thus gain a competitive advantage.

2. CRISP4BigData Reference Model

The CRISP4BigData reference model [4] [3] [5] is based on Kaufmann's Big Data Management Cube (BDM^{cube}) [14] and Bornschlegl's IVIS4BigData reference model [7]. CRISP4BigData is an extension of the classic Cross Industry Standard Process for Data Mining (CRISP-DM), which was developed by the CRISP-DM consortium with the objective of managing the complexity of Data Mining projects [9]. The CRISP4BigData Reference Model (see fig. 2) is based on Kaufmann's six phases of Big Data Management Cube (BDMcube), which include Datafication, Integration, Analytics, Interaction, Effectuation, and Intelligence, as well as the CRISP-DM model's standard four-layer methodology.

The CRISP4BigData methodology represents a hierarchical process model with four distinct layers (based on the CRISP-DM methodology): Phase, Generic Task, Specialized Task, and Process Instance. [3] [5]. Within each of these levels, there is a set of phases (e.g., Business Understanding, Data Understanding, Data Preparation) that are analogous to the standard description of the original CRISP-DM model, but with some additional phases (e.g., Data

Enrichment, Retention, and Archiving) [3] [5].

Each of these phases consists of typical second-level activities. This layer is called generic because it is broad enough to cover all potential use cases. The third level, Specialized Task, explains how the activities of the generic tasks should be handled and how they should vary in different contexts.

For example, the Specialized Task layer manages how data is handled, such as "cleaning numeric values versus cleaning categorical values, or whether the problem type is clustering or predictive modeling." [9] The fourth layer, the process instance, is a record collection of ongoing activities, decisions, and Big Data analysis results. According to the layer, "a process instance is organized according to tasks defined at higher layers but represents what actually happened in a particular engagement rather than what happens in general." [9]

In the following sections, the various steps of the CRISP4BigData reference model are explained in detail: Business Understanding refers to the goals and challenges that arise from a project plan for the targeted use of Big Data analytics technologies.

The Data Understanding phase addresses the collection of internal and external data and the definition of data types and source systems.

Data Preparation manages the transformation or cleansing of data (or updating, enhancing, and reducing) to improve data quality. All process elements of CRISP4BigData Data Enrichment are responsible for enriching the database with valuable metadata based on the entire process, gaining insight into process information to improve process quality, or updating the knowledge-based support phase with information. [3] [5].

The Modeling aspect discusses the creation and use of a statistical, mathematical, or process model based on a suitable data model. The Analysis element explains the use of a Big Data analysis method or algorithms to examine the available data (model) and is based on the statistical, mathematical, or process model created in the Modeling element. Evolution assesses the outcome of the analysis and the process. This phase also assesses the result's correctness, usefulness, uniqueness, and relevance.

The Deployment, Collaboration, and display phase deals with the deployment and display of the analytic results, as well as managing the distribution of the correct results to the right people (collaboration) [3] [5].

The Potentialities & Continuous Product and Service Improvement component manages potentials and oversees a continuous improvement approach to identify new prospects for the firm.

The information Generation and Management process guarantees that the created insights and information are not lost. The Retention and Archiving stage ensure that data is preserved and archived indefinitely; it also classifies data into hot, cold, and warm tiers. Finally, runs through each phase demonstrate that the CRISP4BigData reference model is not a timely process. In certain circumstances, repeating the entire process is valuable or essential to process new information received during the project or to enhance the overall process [3] [5].

3. User Study

In order to prove the usefulness of CRISP4BigData, a user study was conducted [6]. The usefulness of a system must first be distinguished from the higher-level issue of system acceptance, which stands for all user needs and requirements. The category of usefulness describes whether a system can be used to achieve a desired goal and is divided into the to subcategories of *utility* and *usability* "where *utility* is the question whether the functionality of the system in principle can do what is needed, and *usability* is the question of how well users can use this functionality." [16]

3.1. Method

The utility was analysed using an expert survey. This type of evaluation is an established and recognized evaluation technique in research [20].

The survey was conducted in the form of an anonymous online survey. The questionnaire addressed to the participants was divided into 6 groups and included a total amount of 28 questions that could be answered via multiple-choice and free-text fields [6].

- The first group of questions (*Professional Background* with 3 questions) serves to classify the expertise and scientific background of the participants. The tasks in the individual phases of the process model make demands on the knowledge and skills of the users.
- The questions in the group *Big Data Application Area* (4 questions), aim to get a context to the participants' environment.
- The 5 questions in *Properties of Data* are intended to provide a classification of the data base according to the definition of Laney [15].
- In the question group *Phases and Tasks* with 7 questions it is to be determined, how the Big Data application of the participants could be integrated into the CRISP4BigData model.
- Which technologies and methods are used here is addressed in a question- group with 6 questions. This is necessary to evaluate the interoperability with the prototype.
- The skills required by the end user for the application of the process model are determined in the question group *Skills of the end user* with 3 questions.

Usability is not a single, one-dimensional property, but has several components, which are associated with the five properties of learnability, efficiency, memorability, fault tolerance and subjective satisfaction [16]. The selection of the evaluation method was made on the basis of a comparative study of evaluation methods conducted by Jeffries et al [13]: An exclusion criterion for a heuristic evaluation, which provided the best results in this study, is the dependency on several UI specialists, who were not available for this work. General usability tests, on the other hand, are ruled out primarily because of their high cost. Furthermore, in the study many serious problems could not even be found using this method. The same applies to the method of revising guidelines, which is a sensible alternative for software developers, but in the end is more likely to be considered as a substitute for a heuristic evaluation. Weighing up the goals, the knowledge sought and the the available resources therefore a Cognitive Walkthrough (CW)

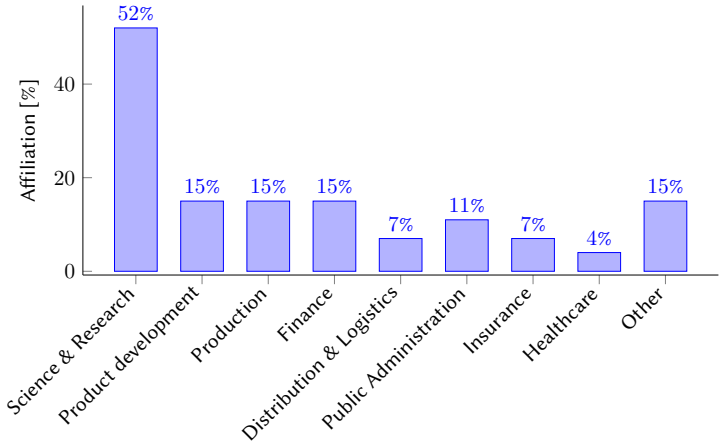


Figure 2: Industry classification [6]

has been applied. Selecting this method also avoids the risk of finding too many problems that have less relevance to the CRISP4BigData process model [13].

The procedure corresponds to a proposal by Wharton et al [19], which describes the practical implementation of the method and is based on the theoretical foundations of the work by Polson et al [17]. The CW focuses on an evaluation regarding the learnability of a user interface. This is done by exploring the user interface, simply following the usual work tasks [19].

The usability was first evaluated in a preliminary study to identify potential improvements. In a further step, these improvements were modeled and implemented in the system. Through this reimplementation of individual components, a further CW has been made afterwards.

3.2. Results of utility evaluation

3.2.1. Demographics

The expert survey had a total of 62 participants. 35 of them terminated the survey prematurely, 27 completed the questionnaire in full. In the evaluation only the 27 participants who completed the questionnaire in full are included.

Participants who did not complete the survey to the end did so after nearly every completion stage (7 without answering a question, 12 in the first, 5 in the second, 6 in the third, two in the fourth, and three in the fifth group of questions). This shows, that there was not a general comprehension problem in a particular question or a group of questions.

The majority, namely 14 participants, work in science and research. They are followed by 4 each in product development, production and finance. 3 participants work in the public administration sector and 2 each in the distribution & logistics and insurance sectors. One participant felt that he belongs to the healthcare sector. The following information was entered in the free text field: Transportation, digital services, aerospace, and energy. Since this question was offered as a multiple choice, the sum of the responses is 38, which means that some of the participants are active in several sectors (cf. fig. 2). This gives a good range of the Big Data applications to which the following answers refer to [6].

89 % of the participants have an academic degree (Bachelor 15 %, Master 26 %, Diplom 19 %

Magister 4 %, Ph.D. 26 %). Only 11 % have not completed an academic degree. The composition thus reflects the distribution in the industry affiliation.

In terms of team size, 44 % work in a small team of 1-5 people, 22 % in a medium-sized team of 5-10 people, and 33 % in a large team of more than 5 people. persons, 22 % in a medium-sized team with 5-10 persons and 33 % in large teams with more than than 10 people. This means that the different team sizes are are sufficiently represented. This is important because the way of working and the composition of the tasks differs with an increasing team size [6].

3.2.2. Scope of application

The categories of analysis methods are distributed over 178 % due to the possibility of multiple selection. At 70 %, data mining is the most common method. This is followed by business intelligence and knowledge management, each with 52 %. The Six Sigma management method was specified once in a free text field, which corresponds to 4 % of the responses.

Regarding the maturity level, it was stated that 37 % of the applications have already been launched. 22 % are in the implementation phase, 19 % are in the validation phase, 4 % are still in the design phase, and 11 % are in the concept phase. Only 7 % have only an idea for the application so far.

The user stereotypes (multiple selection possible) are distributed over a total of 133 % with 33 % academic, 48 % technical, and 52 % management. The model can be used – as intended – by all selected stereotypes.

Responses to the question about what technical skills users need to have in participants' Big Data applications were also offered via multiple choice, which is why the total is 256 %. For 78 % business and technical skills are required. Analytical skills are required by 13 %. 9 % each require knowledge of programming, databases and visualization. 8 % must be familiar with statistics [6].

3.2.3. Data characteristics

The amount of data used in the applications of the participants is 44 % in the range of mebibytes, 33 % Gebibytes, 15 % Tebibytes and 7 % Pebibytes. 22 % of the data is generated in real time, 44 % continuously, and 33 % on demand. The data is then processed and published in real time for 15 %, continuously for 41 %, and on demand for 44 % of the applications.

The number of different data sources used is between 2 and 5 for 48 % of applications, 15 % use 6 to 10, and 37 % use more than 10 sources (cf. Fig. 3). Databases are the most common (89 %) followed by spreadsheets and structured text (67 % each), and unstructured text (33 %). Image files have a share of 52 %, video files 11 %, and audio data 7 % (cf. Fig. 4) [6].

3.2.4. Phases and tasks

How the survey participants rated the necessity of the individual phases and tasks of the CRISP4BigData model for their own applications is shown in Table 1. It can be seen that the first two phases are rated as very necessary by 96 % and 85 %. Even the third phase is still considered necessary by almost three quarters of the participants. Only the last two phases receive only 44 % approval each. These are precisely the phases that actually provide real added

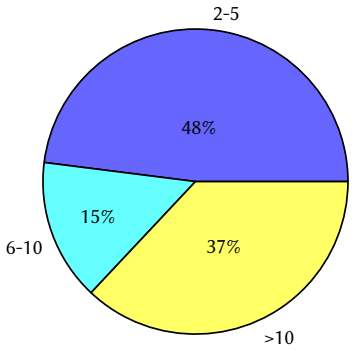


Figure 3: Variation of Datasources [6]

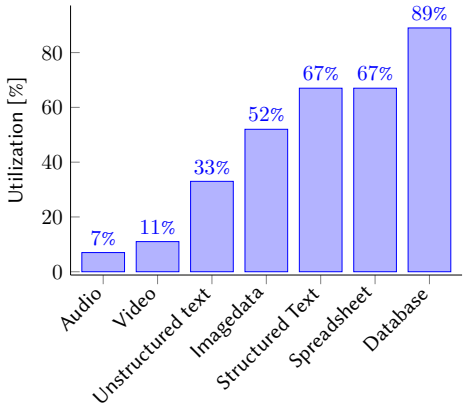


Figure 4: Utilized Datasources [6]

Phase	Task	Usage
Data Collection, Management & Curation (96 %)	Business Understanding	63 %
	Data Understanding	89 %
	Data Preparation	81 %
	Data Enrichment	67 %
Analytics (85 %)	Modelling	74 %
	Analysis	78 %
	Evaluation	63 %
	Data Enrichment	44 %
Interaction & Perception (74 %)	Deployment, Collaboration, Visualization	63 %
	Data Enrichment	33 %
Insight & Effectuation (44 %)	Potentialities & Continuous Product and Service Improvement	33 %
	Data Enrichment	19 %
Knowledge-based Support (44 %)	Knowledge Generation and Management	37 %
	Retention and Archiving	37 %

Table 1
Applied phases and tasks of the CRISP4BigData model [6]

value compared to other models. The question arises here as to whether the participants have actually dealt sufficiently with the requirements such as data enrichment, archiving and finding experts for an analysis [6]. These aspects were ultimately decisive for the initiation of the new process model [3] [5].

3.2.5. Technologies and methods

When querying the frameworks used, it became apparent that the given answer options do not reflect the actual diversity; visible by the use of the free text field by 41 % of the participants. However, one of these could be clearly assigned to Relational Database Management Systems

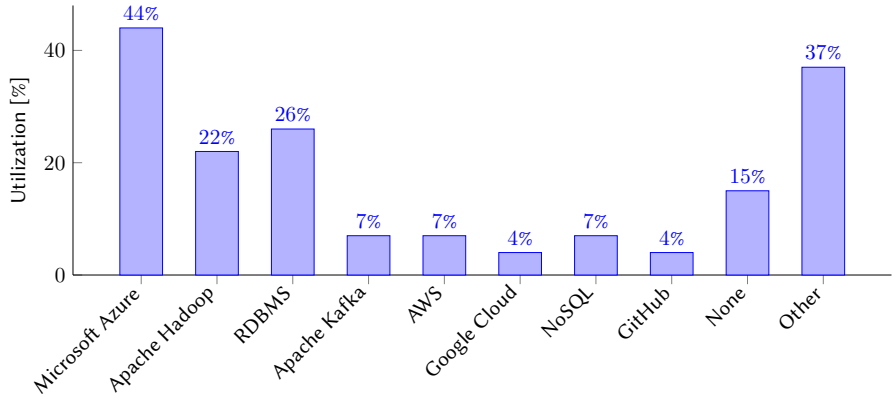


Figure 5: Utilized Frameworks [6]

(RDBMS). For others, however, a summary was possible, or only software tools were mentioned that were not software tools that are not Big Data frameworks. The raw data was therefore cleaned accordingly for the evaluation. The result is shown in Figure 5.

Microsoft Azure is most frequently cited with 44 %, followed by RDBMS (26 %), Apache Hadoop (22 %), and Kafka (7 %). Cloud services such as Amazon Web Services (AWS) (7 %) and Google Cloud (4 %) are also used. NoSQL (7 %), GitHub (4 &) and the framework TileDB (4 %) were also mentioned under "Other". 15 % of the entries are not frameworks.

37 % of the participants use a process model for their Big Data analysis. CRISP-DM and KDD Process were named most frequently, each with 15 %. CRISP4BigData or their own process model are used by 11 %.

When asked if the CRISP4BigData process model is sufficiently complete, to represent all the necessary steps within your research or business activities. 78 % answered yes and 19 % answered no. The remaining 3 % did not respond to the question [6].

3.2.6. End user capabilities

Regarding the degree of adaptability by the users, it was indicated that 22 % of applications have a high degree of adaptability, 26 % a medium degree of adaptability, and 48 % a high degree of adaptability. Only 4 % of the applications did not allow for indicated. 78 % allow customizations in the integration of data sources, 70 % for visualizations, 56 % for views. Modifications to workflows are possible for 52 % and the types of archiving can be varied for 33 %. The integration of different software frameworks is possible for 22 %.

In relation to the CRISP4BigData process model, a high need for adaptability is expressed. The relatively low demand for the adaptability of the Knowledge-based Support phase arises almost automatically from its function as a documentation and archiving phase [6].

3.3. Results of usability evaluation

In the first iteration of the CW, some problems were found that could be solved by reimplementing the user interface. As a result, some changes were made on elements in the UI belonging to all phases of the CRISP4BigData model. Also a new function has been implemented, which

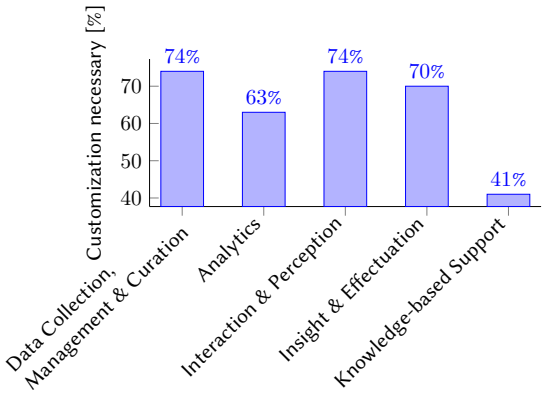


Figure 6: Necessary adaptability of the individual CRISP4BigData phases [6]

enables verification of the integrity and authenticity of data. This enforces a better security. In addition, the basic structure of the navigation as well as individual input elements were restructured thematically to ensure a better mapping to the process model. In terms of the criteria of usability named in chapter 3.1, the improvements concerning the criteria of efficiency, memorability, learnability, and error tolerance. Thus, the reimplementation based on the CW has been approved the usability significantly.

With the successful completion of all tasks in the post implementation CW, the overall usability of the prototype can be confirmed [6].

4. Conclusion

Overall, the results of the expert survey can confirm the usefulness of the CRISP4BigData process model. The profile of the participants shows a good distribution of industries, educational qualifications and team composition. The same is true for the methods used, the maturity level, and the user stereotypes of the applications. The characteristics and diversity of the data justify the architecture and, in particular, the mediator/wrapper, that was implemented by Schmatz in a master thesis [18]. It is part of the first phase Data Collection, Management & Curation and provides a consistent and flexible user interface. This ensures a high level of interoperability when importing and exporting data. The necessity of the individual phases and their tasks is also confirmed, even if the assessment of the need for the Insight & Effectuation and Knowledge-based Support phases is rather low. This is attributed to the fact that the participants have not yet sufficiently addressed the added value of the process model. This assumption is supported by the fact that most of the participating experts have not yet used a process model for their analyses.

The completeness of the model is confirmed by almost four-fifths of the experts. The remaining participants partially express concerns about the complexity of the model, which in fact has no influence to the completeness. On the other hand, one objection is justified with regard to the data integrity: An element that enabling an authenticity check of data is not yet available in the model. In addition to this fact, the high level of adaptability of almost all phases and the large number of frameworks used.

The findings on the utility of the model obtained in the evaluation show that the requirements placed on CRISP4BigData are indeed met. The implementations carried out also transfer the theoretical model into a practical application and thus demonstrate its feasibility. The evaluation in the overall context can also verify this [6].

5. Discussion and Outlook

The results of this evaluation show that CRISP4BigData supports the analysis of Big Data in a project- or process-oriented approach to increase manageability and transparency for the end user. Due to the openness and flexibility of the presented process model, not only different cross-industry projects but also different technological frameworks can be supported. The CRISP4BigData UI provides additional support for the implementation of these projects.

The evaluations have also shown that the automation solution for controlling the modeled processes is a special feature that offers a promising perspective for future applications. These aspects open options for future research that should focus on a consistent approach to implementation and integration.

Even if the number of participants in the survey (27 participants) and the participants in the cognitive walkthrough (3 participants) is limited to a total of 30 people, the expert status, the industry classification and the distribution among various industry sectors as well as the experts surveyed within the cognitive walkthrough provide information that CRISP4BigData can support experts in the context of Big Data in designing and implementing Big Data analysis processes better, faster, more transparently and more sustainably for improving their own competitive advantage.

Both the presented CRISP4BigData UI and the process offer research approaches for the complete automation of integration processes across different technologies.

6. Acknowledgments and Disclaimer

This publication has been produced in the context of the MetaPlat project. The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 690998. However, this paper reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains.

References

[1] Anuradha, J., et al.: A brief introduction on big data 5vs characteristics and hadoop technology. Procedia computer science **48**, 319–324 (2015)

[2] Beath, C., Becerra-Fernandez, I., Ross, J., Short, J.: Finding value in the information explosion., http://sloanreview.mit.edu/article/finding-value-in-the-information-explosion/, accessed on 5/22/2023

[3] Berwind, K., Bornschlegl, M., Hemmje, M., Kaufmann, M.: Towards a cross industry standard process to support big data applications in virtual research environments. In:

Collaborative European Research Conference. pp. 82–91. CERC2016, Cork Institute of Technology, Cork, Ireland (2017)
[4] Berwind, K.: Design of use case diagrams and personas based on the crisp4bigdata process and conceptual design and implementation of a graphical user interface for crisp4bigdata. BIRDS - Bridging the Gap between Information Science, Information Retrieval and Data Science (2021)
[5] Berwind, K., Voronov, A., Schneider, M., Bornschlegl, M., Engel, F., Kaufmann, M., Hemmje, M.: Big data reference model
[6] Bley, M.: Evaluation und Reimplementierung der web-basierten Benutzungsschnittstelle zum CRISP4BigData-System. Masterthesis, Fernuniversität Hagen, Fakultät für Mathematik und Informatik (2023)
[7] Bornschlegl, M.X.: Advanced Visual Interfaces Supporting Distributed Cloud-Based Big Data Analysis. Ph.D. thesis, University of Hagen, Hagen (2019)
[8] Carter, P.: Future architectures, skills and roadmaps for the cio, http://www.sas.com/resources/asset/BigDataAnalytics-FutureArchitectures-Skills-RoadmapsfortheCIO.pdf, accessed on 7/1/2016
[9] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R.: CRISP-DM 1.0: Step-by-step data mining guide, ftp://ftp.software.ibm.com/software/analytics/spss/support/Modeler/Documentation/14/UserManual/CRISP-DM.pdf, accessed on 6/9/2016
[10] Davenport, T.H., Patil, D., Scientist, D.: The sexiest job of the 21st century. Harvard Business Review **8** (2012)
[11] Freiknecht, J.: Big Data in der Praxis: Beispiellösungen mit Hadoop und NoSQL. Daten speichern, aufbereiten, visualisieren. Carl Hanser Verlag GmbH Co KG (2014)
[12] Gartner Inc.: It-glossary big data, http://www.gartner.com/it-glossary/big-data/, accessed on 29/5/2023
[13] Jeffries, R., Miller, J.R., Wharton, C., Uyeda, K.: User interface evaluation in the real world: A comparison of four techniques. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. p. 119–124. CHI '91, Association for Computing Machinery, New York, NY, USA (1991)
[14] Kaufmann, M., Eljasik-Swoboda, T., Nawroth, C., Berwind, K., Bornschlegl, M.X., Hemmje, M.L.: Modeling and qualitative evaluation of a management canvas for big data applications. In: DATA. pp. 149–156 (2017)
[15] Laney, D.: 3d data management: Controlling data volume, variety and velocity. https://idoc.pub/documents/3d-data-management-controlling-data-volume-velocity-and-variety-546g5mg3ywn8, accessed on 10/23/2021
[16] Nielsen, J.: Usability engineering. Academic Press (1993)
[17] Polson, P.G., Lewis, C., Rieman, J., Wharton, C.: Cognitive walkthroughs: a method for theory-based evaluation of user interfaces. International Journal of man-machine studies **36**(5), 741–773 (1992)
[18] Schmatz, K.D.: Konzeption, Implementierung und Evaluierung einer datenbasierten Schnittstelle für heterogene Quellsysteme basierend auf der Mediator-Wrapper-Architektur innerhalb eines Hadoop-Ökosystems. Masterthesis, Fernuniversität Hagen, Fakultät für

Mathematik und Informatik (2018)
[19] Wharton, C., Rieman, J., Lewis, C., Polson, P.: The cognitive walkthrough method: A practitioner's guide. In: Usability inspection methods, pp. 105–140. John Wiley & Sons (1994)
[20] Zeepedia: Evaluation paradigms and techniques. https://zeepedia.com/read.php?evaluation_paradigms_and_techniques_human_computer_interaction&b=11&c=29, accessed on 03/08/2023

Improving Multiclass Classification of Cardiac Arrhythmias with Photoplethysmography using an Ensemble Approach of Binary Classifiers

Katharina Post[a,b], Marisa Mohr[a], Max Koeppel[a] and Astrid Laubenheimer[b]

[a] *inovex GmbH*
[b] *Karlsruhe University of Applied Sciences*

Abstract
Cardiac arrhythmias are a leading cause of death worldwide, emphasizing the need for early detection. Electrocardiograms (ECGs) are a major tool to detect arrhythmias, but they require specialized equipment and trained personnel. To support doctors in diagnosing ECG data, machine and deep learning techniques have been developed. Recently, smart devices with photoplethysmography (PPG) are increasingly used as an alternative for monitoring cardiovascular health outside of clinical settings. While some smartwatches can detect individual arrhythmia classes, approaches to classify other abnormal heartbeats and rhythms using PPG are limited. To address this issue, this study aims to improve the identification of various heartbeat and rhythm abnormalities in PPG signals using an ensemble of multiple binary classifiers. The study achieved an F1-score of 89% in the classification of five classes, outperforming other multiclass classification methods in the domain of PPGs. Transfer learning was also assessed in the study, showing that pre-training convolutional neural networks on ECG data can further improve classifier performance on PPG data.

Keywords
multiclass arrhythmia classification, PPG signal, transfer learning, ensemble of binary classifiers

1. Introduction

Cardiac arrhythmias are one of the leading causes of death worldwide, and early detection and classification are crucial for timely intervention [16]. To detect cardiac arrhythmias in patients, primarily recordings of electrocardiograms (ECGs) are used, which are measured via electrodes connected to the patient's body allowing to record signaling-channels, so-called leads. ECG require special medical equipment and trained personal to interpret the results. To support doctors, promising approaches have been undertaken to detect and classify cardiac arrhythmias using machine learning and deep learning techniques [5, 9, 17], providing doctors with more information for better decision-making when diagnosing ECG data.

To extend the options to monitor cardiac health away from medical facilities, an increasing number of smart devices, such as smartwatches, are equipped with photoplethysmography (PPG) as a potential alternative for monitoring cardiovascular health [3]. While some smartwatches can already detect individual arrhythmia classes, approaches to classify other abnormal heartbeats and rhythms using PPG are limited. However, as certain kinds of arrhythmia posses a higher risk for severe complications than others, their timely detection and correct classification remains a key challenge.

CERC 2023: Collaborative European Research Conference, June 09–10, 2023, Barcelona, Spain
katharina.post@inovex.de (K. Post); marisa.mohr@inovex.de (M. Mohr); max.koeppel@inovex.de (M. Koeppel); astrid.laubenheimer@h-ka.de (A. Laubenheimer)

ISBN 978-3-96187-021-9
https://doi.org/10.48444/h_docs-pub-518

This work aims to improve the identification of various heart rhythm abnormalities by a combination of multiple binary classifiers, also known as ensemble classifiers. A model structure is first constructed and trained on ECG data to achieve this goal. Later, this model structure is adapted to PPG data. The research primarily focuses on evaluating the effectiveness of the ensemble approach compared to a multiclass classifier. Furthermore, the usage of transfer learning is assessed.

2. Related Work

Considerable research has been conducted to explore the effectiveness of statistics, machine learning, and deep learning techniques in automatic detection of pathologic changes to the heartcycle on ECGs. In 2021, a comprehensive review of the existing literature on arrhythmia classification was conducted by Neha et al. [13]. The researchers highlight the number of articles that used ECG and PPG to classify arrhythmias. The results of the review show that there were significantly more studies using ECGs than PPGs, with a difference of over eleven times.

Previously described models can be categorized according to the techniques used, the number of labels, and the choice of classes and samples. In the past, statistical methods and machine learning have been emphasized, but recently the emergence of deep learning has introduced new possibilities. Signals of arrhythmia have different morphologies, even within a single patient, which are most likely easier captured by deep learning methods, compared to more traditional machine learning models based on hand-crafted features [8].

The classification of individual types of arrhythmia has shown promising results. For example, Zhang et al. [18] proposed a method to detect atrial fibrillation in ECGs, and Bashar et al. [2] developed a model capable of identifying this class in PPGs. Because certain arrhythmias also pose a high risk to patients, a multiclass classification model capable of identifying a range of arrhythmias is preferable to a model that detects a single abnormal heartbeat or -rhythm.

Focusing on multiclass classification, Dindin et al. [5] proposed a modular multichannel neural network to identify different arrhythmias based on ECG data. The model combines topological data analysis (TDA), handcrafted features, Fast Fourier Transformation, and deep learning. The first component of the model is an autoencoder that uses unsupervised learning to identify sinus rhythm. This autoencoder performs binary classification, so categorizing data as either normal or abnormal. Another component of the model is the generation of Betti curves, which are constructed using TDA, specifically persistent homology [4]. Betti curves are highly robust to variations in ECG signal patterns and are not affected by expansion or contraction along the time axis [5]. These curves represent sub-level and upper-level set filtrations and help to account for inter-patient differences in the data, thereby improving the generalisability of the model to new patients. In addition, the ECG signal is processed using a CNN.

Initial approaches using PPG data focused on the classification of a single arrhythmia class [14, 1]. Neha et al. [13] argued for further development of methods capable of multiclass categorization. Subsequently, in 2022, Liu et al. [10] presented a deep convolutional neural network (DCNN) based on the VGGNet-16 architecture to classify six rhythm types including five different arrhythmias. The authors achieved good results in classifying sinus rhythm and atrial fibrillation. However, they acknowledged the need to improve the classification of other heart rhythm abnormalities. While they were able to obtain an overall accuracy of 85%, this might have partially been caused by a relative overrepresentation of samples with sinus rhythm or atrial fibrillation. By using such an imbalanced data set the resulting high accuracy score may not reflect the true performance of a classifier due to its tendency to predict the majority class.

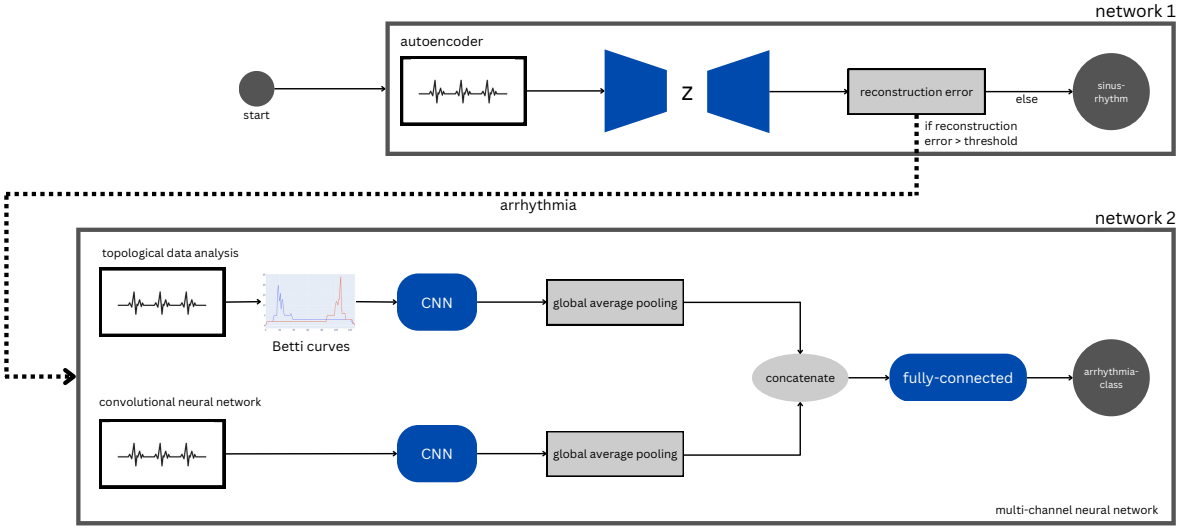


Figure 1: Overview of the proposed model architecture. Signals are processed in network 1, which consists of an autoencoder. If the reconstruction error is less than a dynamically calculated threshold, the signal is classified as normal. Otherwise, the signal is further handled in network 2, composed of two channels. The first channel uses topological data analysis to generate Betti curves, prior to passing them through a CNN. The other channel takes the raw signal as input to another CNN. Due to the multichannel approach, both channels are used to determine an arrhythmia class.

3. System Design

This paper extends previous approaches to detect arrhythmias from ECG and PPG signals using a two-stage deep learning model. Our model is designed to identify normal and abnormal rhythms and subsequently classify the abnormal signals into various arrhythmia classes. For differentiation between various types of arrhythmia, this work compares two classification methods: an ensemble approach, which combines binary classifiers, one for each class, and a multiclass model. Our model components are chosen based on the publication of Dindin et al. [5]. However, the division into two separate networks and the architecture of the models differ from their approach, allowing us to classify normal and abnormal signals before further processing the abnormal signals.

3.1. Model Architecture

To classify arrhythmias, we chose a two-step model architecture starting with an autoencoder network that distinguishes between normal and abnormal heartbeats and rhythms, shown in Figure 1. We train the model on normal sinus rhythm to learn its structure. After training, the weights of the model are frozen. If an abnormal rhythm is introduced into the network during inference, the model will not be able to reconstruct it well and the reconstruction error will be larger than for normal sequences. The reconstruction error is determined by the mean squared error. To distinguish between normal and abnormal sequences, we use a dynamic threshold based on the loss distributions of normal and abnormal samples. This threshold is determined by identifying the intersection of the loss distribution of normal and abnormal sequences. An exemplary illustration is shown in Figure 2. Any sample with a reconstruction error below this threshold is classified as normal, while others above are considered abnormal. Separating the autoencoder from the other model components has the advantage that normal heartbeats do not need to be processed further, thus saving computing power.

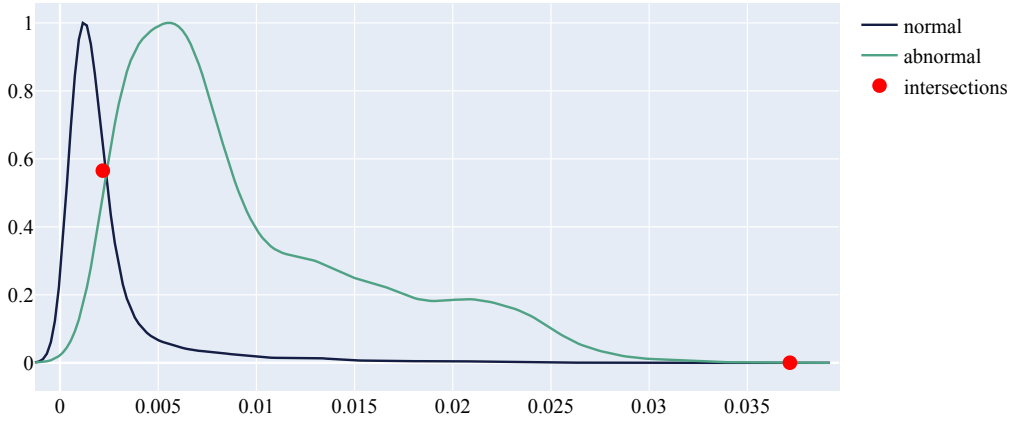


Figure 2: Distribution of loss curves for normal (blue) and abnormal (green) samples. The intersection (red) furthest to the left of the two distributions is used as the threshold. Samples to the left of the threshold are classified as normal and samples to the right are classified as abnormal.

The autoencoder consists of five fully connected layers in both the encoder and decoder, which reduce the input signal from 720 to 16 and reconstruct it back to its original size. The Rectified Linear Unit (ReLU) is used as the activation function, except for the last layer where Sigmoid is applied. The model is trained for 1500 epochs with a batch size of 128 and a learning rate of 0.001.

The second step of our approach involves a two-channel network. The first channel, hereafter referenced as Betti-CNN, generates Betti curves. The two generated Betti curves each of size 128, representing the upper/lower set filtration, are further processed in parallel in a one-dimensional CNN. The second channel digests the signal directly in a one-dimensional CNN. After applying global average pooling to both channels respectively, they are concatenated and further processed in fully connected layers to determine the arrhythmia class.

The model architectures are comparatively smaller than those found in other publications. The Betti-CNN and CNN models consist of only three convolutional layers followed by the global average pooling operation and three fully connected layers. The number of filters increases from 16 to 32 and 64, respectively. While the filter size remains constant at three for the Betti-CNN, it changes from three to five and up to seven for the CNN model. In addition, max-pooling is applied after each convolutional operation, with a filter size of three and a stride of two. ReLU is used as the activation function, and the Softmax activation function is applied in the last layer. The models are trained with the Cross-Entropy loss.

3.2. Transfer Learning

The proposed model structure is designed to accept both ECG and PPG sequences as input data. In addition to developing separate models for each data type, transfer learning techniques are also explored. In particular, the model is first trained on ECG data and then further tuned on PPG data. Because no weights are frozen, the entire model is fine-tuned with PPG data. Transfer Learning aims to enable the network to learn basic structures from the ECG data during training, which can then be applied to the PPG data. An advantage is also the increase in the number of training data, as the model is trained on more samples.

Table 1
Available number of samples per class for ECG and PPG.

Class	ECG	PPG
sinus rhythm (N)	672,141	77,755
premature ventricular contraction (PVC)	2,316	22,605
ventricular tachycardia (VT)	2,722	6,135
premature atrial contraction (PAC)	612	13,010
supraventricular tachycardia (SVT)	876	21,420

3.3. Ensemble of Multiple Binary Classifiers

In addition to creating a multiclass classifier, we evaluate the use of an ensemble of multiple binary classifiers on the channels of the second network. In the ensemble of multiple binary classifiers, the original multiclass problem is divided into subtasks whose combined output performs the classification [11]. The underlying concept is that solving binary classification tasks is comparatively less challenging than addressing a multiclass problem. However, the challenge is to combine the binary classifiers to ensure correct classification afterwards [6]. To binarize the multiclass problem the one-vs-all approach is chosen. Therefore, a binary model is created for each arrhythmia, distinguishing it from all others. The individual binary models are combined by using binary relevance to form an ensemble. The result of the prediction is the union of all classifiers. In our case, the return value was always one class at a time.

4. Experimental Evaluation

In this section, we present the data used for the experimental evaluation and its preprocessing, before discussing the results.

4.1. Data Selection and Preprocessing

To approve the ability of our system, we use one-channel ECGs of 86 patients from the MIT-BIH database and take samples from the MIT-BIH Normal Sinus Rhythm Database [15], the MIT-BIH Arrhythmia Database [12], and the MIT-BIH Malignant Ventricular Ectopy Database [7]. In order to be able to apply transfer learning, ECG data is selected that contains the same classes that are present in the PPG data, namely N, PAC, PVC, VT and SVT. ECG data processing steps include the removal of paced rhythms and signals recorded other than from lead II, resampling to 360 Hz, detrending, and scaling between zero and one. The 30-minute-long ECG signals are divided into two-second sections without overlap and each section is assigned the heartbeat and rhythm label that occurs in that period. In some cases, more than one label is assigned to each sequence. If this is a sinus rhythm label together with an arrhythmia class, the arrhythmia class is selected as the label. In addition, the rhythm label is preferred to the beat label because, for example, a VT consists of at least three PVCs. Sequences still containing multiple arrhythmia labels are removed. The number of seconds is determined by evaluating the occurrence of peaks and the resulting number of samples per class. To define a rhythm, at least three consecutive beats of a class should occur. This means that there should be three peaks per sequence. Using longer sequences would result in fewer samples per class, which would be disadvantageous given the paucity of data for classes A and SVT.

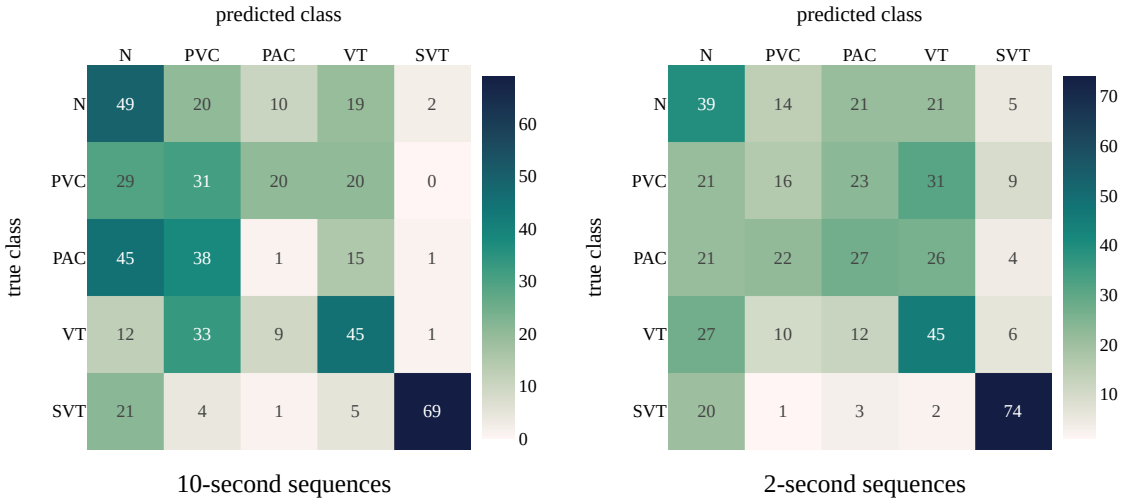


Figure 3: Confusion matrices of the classification of five heartbeat and rhythm classes with the model architecture proposed in Liu et al. [10] for ten-second sequences (left) and two-second sequences (right). The comparison shows an improvement in the classification of PAC and SVT and misclassifications to sinus rhythm are less frequent.

Furthermore, for the model to be used in practice, its high predictive accuracy on previously unseen patient data is essential. However, the morphology of an ECG varies from patient to patient [8]. To test the transferability to new patients, an inter-patient approach is used in which we strictly separated patient sequences for training and testing. Thus, there are data from 69 patients (ECG) and 73 patients (PPG) for training, and 17/18 patients for testing, resembling a training-testing split of 80% to 20%. The number of samples per class is shown in Table 1.

As PPG samples, we use the validation and test data published by Liu et al. [10]. This data is recorded from 91 patients. The signals are resampled to 360 Hz and split from ten-second to two-second sequences to have the same time window as the ECG data. Splitting the sequences into smaller windows might have the potential for incorrect label assignment due to a split in which an arrhythmia-label is assigned to the now independent sinus-rhythm-signal surrounding the pathological heartbeat in an initial ten-second window. Also, cases of PVC or PAC, in which sinus rhythms have to be included in between pathological beats for correct classification might be incorrectly labeled. However, comparing the performance of the ten-second and two-second sequences showed that the potential for mislabeling is negligible, as the F1-scores of both models (tested with the model structure from the publication by Liu et al. [10]) are almost identical (37.32% and 37.92%). The comparison of the confusion matrices (see Figure 3) shows that the division into two-second sequences is advantageous, as the classes PAC and SVT can be recognized better than with ten-second sequences. In addition, misclassifications to sinus rhythm are less frequent, which is important in the medical context.

4.2. Results

The individual models, namely the autoencoder, Betti-CNN, and CNN, are implemented and tested independently before being integrated into the overarching architecture. Training and testing the models on ECG data serve to optimize the models and prepare them for use with PPG data. An objective of the project is to evaluate the performance of transfer learning, where the model is trained with ECG data and fine-tuned with PPG data. However, the challenge in building the models with ECG data is the underrepresentation of certain classes, which prevents the models from achieving the

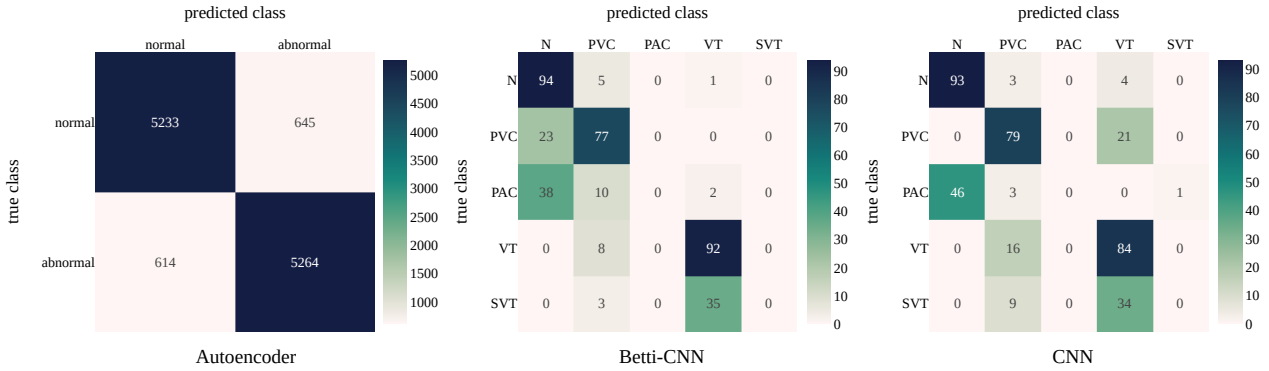


Figure 4: Confusion matrices of the on ECG trained autoencoder (left), the Betti-CNN (middle) and the CNN (right) in the multiclass approach.

benchmark performance levels. To overcome this obstacle, we explore the potential of using multiple one-vs-all binary classifiers rather than a single multiclass classification model. We find that this form of classification outperforms the original approach, particularly in situations involving limited data sets and consequently reduced model structures. Therefore, this aspect is discussed in more detail below.

4.2.1. Multiclass Model

The initial classification into normal and abnormal rhythm is done by the autoencoder, trained on 20,000 normal samples each. This results in an F1-score of 89% for ECG and 65% for PPG data, the corresponding confusion matrix showing the performance on ECG data is presented in Figure 4. When evaluating which classes are misclassified, the PAC class stands out with 11.44% incorrect classifications, while misclassification of other classes are negligible with less than 0.01%.

The Betti-CNN achieve a classification score of 62% on ECG data when considering five classes (N, PVC, PAC, VT and SVT). However, analysis of the confusion matrix (Figure 4) shows that no samples are assigned during inference to the two underrepresented classes (PAC and SVT). Similar performance is observed with the CNN, which obtains a score of 56%. In contrast, when classifying three classes (N, PVC, VT), the Betti-CNN and CNN achieve scores of 90% and 92%, respectively.

When trained on PPG data the Betti-CNN achieves an F1-score of 35%. The confusion matrix shows that a great number of arrhythmias is classified as normal. This behavior is also observed with the autoencoder. In contrast, the CNN is unable to learn any useful patterns and the resulting F1-score is 7%. This is demonstrated by the random assignment of samples to one class during inference.

Combining the two models in a multichannel approach results in an F1-score of 61% on five classes for ECG data and a score of 18.33% for PPG data.

4.2.2. Ensemble of Multiple Binary Classifiers

To deal with the uneven distribution of the classes in ECG data, the combination of several binary classifiers is investigated. Specifically, the one-vs-all approach is used along with binary relevance. A Betti-CNN and a CNN are trained separately for each class, and during inference, test samples are sequentially passed through each binary model. Each model then predicts whether the sample belongs to its associated class. The classification F1-score of each model is shown in Table 2. Notably, the performance of the classifier with the fewest samples (PAC) is the weakest. The overall F1-score

Table 2
F1-scores of each model in the binary ensemble trained on ECG data for each class and the number of samples.

Class	Betti-CNN	CNN	Number of Samples
sinus rhythm	91%	96%	> 5,000,000
premature ventricular contraction	93%	92%	2,316
ventricular tachycardia	92%	89%	2,731
premature atrial contraction	66%	50%	612
supraventricular tachycardia	91%	88%	867

Table 3
Comparison of F1-score performance …

Model	Multiclass Approach	Binary Ensemble
Betti-CNN	35%	59%
CNN	7%	71%
Multichannel Model	18%	89%

(a) … between a multiclass approach and an ensemble of binary classifiers for the classification of PPG signals.

Model	PPG-only	Transfer Learning
Autoencoder	64.88%	66.6%
Betti-CNN	34.94%	35.4%
CNN	6.67%	38.82%
Multichannel Model	18.33%	29.98%

(b) … between a multiclass model trained solely on PPG signals and the use of transfer learning, where the model is first pre-trained on ECG data and then fine-tuned with PPG signals.

for the binary multichannel model is 91%, showing a 30% improvement over the multiclass approach on ECG data.

Furthermore, the performance of binary classifiers and a multiclass model is also compared for PPG data. Similar to the findings for ECGs, the binary approach shows a significant improvement in results. The results of the approaches are shown in Table 3 (a). The findings show a remarkable 71% improvement in the performance of the multichannel model, indicating that it outperforms the use of a single model. This improvement is remarkably strong, highlighting the potential benefits of incorporating multiple channels into the model architecture. It is worth noting that utilizing an ensemble of several binary classifiers can lead to a significant improvement in classification accuracy.

4.2.3. Transfer Learning

Finally, through transfer learning, we refine a model originally trained on ECG data with PPG data. By doing so, we increase the number of training samples. By using the basic structures and features learned from the ECG data, the model is able to train effectively on PPG data. While we observe small improvements in the autoencoder and Betti-CNN, the most significant improvement is in the CNN, as shown in Table 3 (b). As a result, the non-learning model evolves into a classifier that achieves a 38.82% probability of correctly predicting the class. The performance of the multichannel network also improves. Although some of the improvements achieved by transfer learning may be minor, there is always a noticeable performance increase within the limits of imprecision.

5. Conclusion

Our study focused on improving the multiclass classification of cardiac arrhythmias in PPG signals. Using an ensemble of multiple one-vs-all binary classifiers, we achieve an F1-score of 89% in the classification of five classes, outperforming other multiclass classification methods in the domain of PPGs. Utilizing an ensemble of binary classifiers offers several advantages, such as acquiring benchmark results with the use of small model structures and less training data. Furthermore, this approach permits the use of multiple labels per sequence, thus second-long sequences without complex preprocessing can be used as model input. This, together with the smaller model size, makes the approach practical and allows it to be further developed for use in smart device applications.

In addition, we investigated the effectiveness of transfer learning by fine-tuning ECG-pre-trained models on PPG data, which yield a slight enhancement in the performance of some models. Notably, the CNN demonstrated a significant improvement when pre-trained on ECG data, resulting in an F1-score increase of 30%.

The scope of our study is restricted by the limited PPG data available. We used the data published by Liu et al. [10], as there are only a few data sets with labelled PPG data. The classes available in this PPG database must be used for the ECG approach in order to perform transfer learning. Some of these classes are under-represented in the otherwise large database of ECGs. This lack of data prevents us from utilizing very large network architectures to avoid overfitting during the training. Consequently, the CNN model fails to learn when applied to PPGs, and transfer learning results in a significant gain in performance. To determine how a CNN improves with a larger training data set and a more complex structure, and whether transfer learning continues to provide an advantage, further verification is required.

Overall, our results demonstrate that the use of an ensemble of binary classifiers is a practical and effective method for improving the classification of multiple classes of arrhythmias in PPG signals. Additionally, our investigations into the application of transfer learning have shown that pre-training CNNs on ECG data is a promising way to improve classifier performance on PPG data. However, further research is needed to validate our findings on larger data sets and model architectures. Thereby patient metadata such as gender and BMI should be included to ensure the applicability of our findings to a broader patient population.

References

[1] Aschbacher, K., Yilmaz, D., Kerem, Y., Crawford, S., Benaron, D., Liu, J., Eaton, M., Tison, G.H., Olgin, J.E., Li, Y., Marcus, G.M.: Atrial fibrillation detection from raw photoplethysmography waveforms: A deep learning application. Heart Rhythm O2 **1**(1), 3–9 (2020). https://doi.org/10.1016/j.hroo.2020.02.002, https://www.sciencedirect.com/science/article/pii/S2666501820300040

[2] Bashar, S.K., Han, D., Ding, E., Whitcomb, C., McManus, D.D., Chon, K.H.: Smartwatch based atrial fibrillation detection from photoplethysmography signals. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) pp. 4306–4309 (2019). https://doi.org/10.1109/EMBC.2019.8856928

[3] Biswas, D., Everson, L., Liu, M., Panwar, M., Verhoef, B.E., Patki, S., Kim, C.H., Acharyya, A., Van Hoof, C., Konijnenburg, M., Van Helleputte, N.: Cornet: Deep learning framework for ppg-based heart rate estimation and biometric identification in ambulant envi-

ronment. IEEE Transactions on Biomedical Circuits and Systems **13**(2), 282–291 (2019). https://doi.org/10.1109/TBCAS.2019.2892297

[4] Carlsson, G.: Topology and data. Bulletin of The American Mathematical Society - BULL AMER MATH SOC **46**, 255–308 (04 2009). https://doi.org/10.1090/S0273-0979-09-01249-X

[5] Dindin, M., Umeda, Y., Chazal, F.: Topological data analysis for arrhythmia detection through modular neural networks. In: Goutte, C., Zhu, X. (eds.) Advances in Artificial Intelligence. pp. 177–188. Springer International Publishing, Cham (2020)

[6] Galar, M., Fernández, A., Barrenechea, E., Bustince, H., Herrera, F.: An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. Pattern Recognition **44**(8), 1761–1776 (2011). https://doi.org/10.1016/j.patcog.2011.01.017, https://www.sciencedirect.com/science/article/pii/S0031320311000458

[7] Greenwald, S.D.: The mit-bih malignant ventricular arrhythmia database (1992). https://doi.org/10.13026/C22P44, https://physionet.org/content/vfdb/

[8] Hoekema, R., Uijen, G., van Oosterom, A.: Geometrical aspects of the interindividual variability of multilead ecg recordings. IEEE Transactions on Biomedical Engineering **48**(5), 551–559 (2001). https://doi.org/10.1109/10.918594

[9] Irfan, S., Anjum, N., Althobaiti, T., Alotaibi, A.A., Siddiqui, A.B., Ramzan, N.: Heartbeat classification and arrhythmia detection using a multi-model deep-learning technique. Sensors **22**(15) (2022). https://doi.org/10.3390/s22155606, https://www.mdpi.com/1424-8220/22/15/5606

[10] Liu, Z., Zhou, B., Chen, X., Li, Y., Tang, M., Miao, F.: Multiclass arrhythmia detection and classification from photoplethysmography signals using a deep convolutional neural network. Journal of the American Heart Association **11** (2022). https://doi.org/10.1161/JAHA.121.023555

[11] Lorena, A., Carvalho, A., Gama, J.: A review on the combination of binary classifiers in multiclass problems. Artificial Intelligence Review **30**, 19–37 (12 2008). https://doi.org/10.1007/s10462-009-9114-9

[12] Moody, G.B., Mark, R.G.: Mit-bih arrhythmia database (1992). https://doi.org/10.13026/C2F305, https://physionet.org/content/mitdb/

[13] Neha, Sardana, H.K., Kanwade, R., Tewary, S.: Arrhythmia detection and classification using ecg and ppg techniques: a review. Physical and Engineering Sciences in Medicine **44**(4), 1027–1048 (Dec 2021). https://doi.org/10.1007/s13246-021-01072-5

[14] Polania, L., Mestha, L., Huang, D., Couderc, J.P.: Method for classifying cardiac arrhythmias using photoplethysmography. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. vol. 2015, pp. 6574–6577 (08 2015). https://doi.org/10.1109/EMBC.2015.7319899

[15] The Beth Israel Deaconess Medical Center, T.A.L.: The mit-bih normal sinus rhythm database (1990). https://doi.org/10.13026/C2NK5R, https://physionet.org/content/nsrdb/

[16] WHO: World health organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. Lancet Global Health **7** (2019). https://doi.org/10.1016/S2214-109X(19)30318-3

[17] Wu, M., Yang, W., Wong, S.: A study on arrhythmia via ecg signal classification using the convolutional neural network. Frontiers in computational neuroscience **14** (2021). https://doi.org/10.3389/fncom.2020.564015

[18] Zhang, X., Li, J., Cai, Z., Zhang, L., Chen, Z., Liu, C.: Over-fitting suppression training strategies for deep learning-based atrial fibrillation detection. Medical & Biological Engineering & Computing **59** (01 2021). https://doi.org/10.1007/s11517-020-02292-9

Analyzing CNN Architectures for Secure and Private Image Classification with Homomorphic Encryption and Differential Privacy

Robin Baumann[a], David Böhm[b], Raoul Saipt[b] and Astrid Laubenheimer[a]

[a]*Intelligent Systems Research Group, Karlsruhe University of Applied Sciences*
[b]*Karlsruhe University of Applied Sciences*

Abstract

Homomorphic encryption (HE) enables privacy-preserving deep learning, but it comes with significant performance overheads. In this study, we evaluate the impact of model architectures on the utility and efficiency of deep learning models under differential privacy (DP) and HE settings. Our experiments reveal that dedicated model architectures are crucial for maintaining model utility when using DP. Moreover, we observe that aligning complex model architectures like ResNets for HE by replacing ReLU with square activation, max pooling with average pooling, and group norm with batch norm strongly deteriorates model utility and results in architectures with sharp minima that fail to generalize. Training such models with DP, however, yields a regularizing effect that improves model utility. Our study contributes to the understanding of the role of model architecture on the applicability of DP and HE.

Keywords

Homomorphic Encryption, Differential Privacy, Deep Learning

1. Introduction

With the increasing importance of data privacy, there is a growing need for secure and private machine learning techniques that can protect sensitive data while still allowing for useful insights to be generated. Homomorphic encryption and differential privacy are two such techniques that have emerged as powerful tools for enabling secure and private machine learning [4, 6].

Homomorphic encryption (HE), first introduced in [30], allows for data to be encrypted in a way that preserves its mathematical structure, enabling computations to be performed on the encrypted data without first decrypting it. Differential privacy (DP) [10], on the other hand, provides a rigorous framework for protecting the privacy of individual data points, by adding random noise to the output of an algorithm in a way that preserves overall statistical properties of the data. Both techniques fulfill different aspects in privacy-preserving deep learning. HE works towards protecting the sensitive data, as well as the model weights from being exposed to adversaries during inference while DP obfuscates the training data in order to protect individual data owners from being exposed through the means of specifically designed attacks, like model inversion or membership inference. Following [4], we denote the former as input and model secrecy, respectively, and the latter as data privacy.

We consider a system framework in which both aspects are explicit requirements. More specifically, our framework comprises a trusted environment where we can train our AI models on sensitive data in plaintext with the consent of data owners. However, since we deal with sensitive personal data, we

CERC 2023: Collaborative European Research Conference, June 09–10, 2023, Barcelona, Spain
robin.baumann@h-ka.de (R. Baumann); astrid.laubenheimer@h-ka.de (A. Laubenheimer)
0009-0004-0961-4655 (R. Baumann); 0000-0001-7955-4521 (A. Laubenheimer)

ISBN 978-3-96187-021-9
https://doi.org/10.48444/h_docs-pub-518

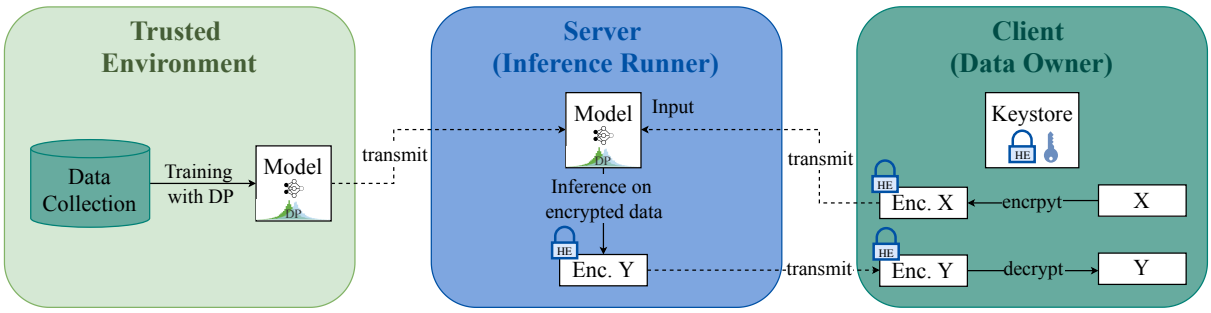


Figure 1: Overview of the investigated application scenario. We assume that all models are trained in a trusted environment with differential privacy (DP) in order to protect data owners. At inference time, we encrypt the data and evaluate the model on the ciphertext rather than on plaintext. The model can be encrypted additionally to prevent the weights from being exposed to adversaries.

still want to protect individuals and thus apply differential privacy during model training to ensure plausible deniability. The trained models can then be used for inference on a centralized server. In order to preserve the privacy of users during model inference, we apply homomorphic encryption. An overview of this setting is depicted in Figure 1. Using this design, the model can be deployed on a server that can be considered as an honest-but-curious adversary but still provide input and output privacy, as well as model secrecy.

Designing privacy-preserving AI solutions comprises several trade-offs between utility (i.e. prediction quality) and privacy, privacy and efficiency, and utility and efficiency, where efficiency includes both, speed and hardware utilisation. These trade-offs are already well-studied [4, 32] and several optimizations have been proposed in order to tighten the gap between plaintext and homomorphically encrypted image classification models [3, 5, 12, 33, 36] or designing more accurate model architectures for DP [8, 16, 20, 29, 31]. Little work has been published on the combined setting, i.e. optimizing deep learning architectures for accurate predictions under DP and fast inference using HE.

In this work, we compare several convolutional neural network (CNN) architectures regarding their utility on two image classification datasets with different data complexity, both with and without DP applied. We then exchange several building blocks of those networks in order to make them compatible with HE and repeat the experiments. We observe significant deterioration in model utility when aligning model architectures for HE that were originally designed for the application of DP. However, the application of DP seems to have regularizing effects on the model training, thus improving the results on the HE-aligned model architectures. Overall, this paper contributes to the growing body of work on secure and private machine learning, and provides insights into the practical considerations involved in designing and implementing these systems.

The remainder of this paper is structured as follows. Section 2 reviews related work in the areas of homomorphic encryption, differential privacy and analysis of deep learning model architectures. Section 3 describes the investigated model architectures and training procedure. In Section 4, we report our results and interpret and discuss them in Section 5 while also pointing out the limitations of this work. Finally, Section 6 concludes this work.

2. Related Work

In the following, we review literature published on homomorphic encryption and differential privacy with a specific focus on image classification applications.

2.1. Homomorphic Encryption

First introduced in [30], HE enables evaluating functions over encrypted data. Dowlin et al. [12] introduced CryptoNets, the first deep learning architecture specifically designed for HE. The architecture uses squared activation functions and scaled max pooling, but only consists of two convolutional layers. Chabanne et al. [5] extended CryptoNets to six layers by using a polynomial approximation of the ReLU activation function preceded by a batch normalization layer [15]. Although normalization in theory incorporates the computation of a square root, the approximation of this can be mitigated by reparameterizing the layer weights and biases prior to the normalization layer [14]. Badawi et al. [3] introduced the first homomorphically encrypted CNN that can run on a GPU. Nandakumar et al. [26] proposed a fully homomorphically encrypted training algorithm for deep neural networks. While all of the previously mentioned works evaluated their approaches only on MNIST or CIFAR-10, Wingarz et al. [33] investigated the scalability of HE for datasets with higher input and output dimensionality by leveraging parameter quantization and pre-processing for faster encryption. In general, HE induces an enormous computational overhead, especially on high resolution image data [33]. Therefore, all of the network architectures introduced in the publications mentioned above are significantly smaller than state of the art architectures in visual computing.

2.2. Differential Privacy

Differential privacy [10] is arguably the most popular data privacy mechanism, providing a mathematically rigorous privacy guarantee in the form of a privacy budget (ϵ, δ). This budget depends on the dataset size and number of training epochs. In deep learning, DP is usually applied to the gradients during optimization with the DP-SGD optimization algorithm [1], which obfuscates the exact gradients with noise sampled from a normal distribution. Due to this obfuscation, the application of differential privacy results in neural networks with a lower prediction accuracy opposed to non-private counterparts. Several model architectures have been proposed to reduce the loss in accuracy under DP. Proposed optimizations to the training process of established network architectures in order to enable DP on large-scale vision datasets include the application of weight standardization [28], replacing batch normalization with group normalization [34], increasing the batch size [24], applying parameter averaging [27] and pre-training on a non-private dataset [8, 20]. Remerscheid et al. proposed SmoothNets [29], a model family found by a neural architecture search subjected to various observations about the implications of specific hyperparameter choices on the model utility under DP. Especially the width-depth ratio when scaling neural networks seems to be of importance when using DP [8, 29].

2.3. Analyzing Deep Learning Architectures

Since deep learning optimization is highly non-convex, many local optima, as well as saddle points and plateaus exist in the loss landscape [7]. Li et al. [22] proposed a method to visualize the loss landscape of neural networks and investigate the implications of several architectural decisions, like the inclusion of skip connections, on the structure of the loss landscape. Dinh et al. [9] showed that flatness of minimizers is not necessarily correlated to the generalization ability of the network. Keskar et al. [17] reported a tendency of large batch optimization to converge to sharp minimizers, whereas small batch sizes seem to have a regularizing effect and converge to flat minimizers. Besides the topological structure of the loss landscape, other theories of generalization focus on norms expressed over the weight space [21], model compression under the PAC-Bayes Framework [23] and formulation of data-dependend error bounds [11].

3. Methodology

From the literature review we identified several orthogonal neural network architecture design decisions when optimizing either for DP or HE. While wide architectures, like wide ResNets [37], seem to outperform other architectures for DP [8], HE architectures are usually restricted in their width and depth in order to limit the number of multiplications in a forward pass. In addition, the application of the ReLU non-linearity requires a polynomial approximation in HE architectures, with square activation being the approximation with the lowest degree. Furthermore, max pooling is not applicable in HE and thus is usually replaced with (scaled) average pooling. Batch normalization layers can be reparameterized and are therefore applicable in both scenarios. However, several DP-optimized architectures include group normalization [34] instead of batch norm. Unfortunately, the reparameterization trick [14] does not generalize from batch normalization to group normalization, since the latter depends on on-the-fly computed statistics over channel groups, involving the computation of a square root. Using group normalization in HE therefore requires the approximation of a square root. In our experiments, we replaced group norm with batch norm when training models for HE.

3.1. Baseline Architectures

We implemented two baseline architectures, one optimized for DP and the other optimized for HE. In order to compare the effects of orthogonal model architecture designs on the opposing privacy regime, we adjusted the respective baseline architectures to meet the best practices from the other regime. Specifically, we removed restrictions on activation functions and pooling layers from the HE baseline and added those to the DP baseline. Furthermore, we implemented two variations of these baseline architectures as described below.

Our DP baseline model is inspired by [8] and is a Wide Residual Network (WRN) [37] with width scaling $k = 2$. Like [8], we replaced batch normalization with group normalization. The number of groups of the group normalization layers in the WRN Blocks is $\frac{C_{out}}{4}$, where C_{out} denotes the number of output channels of the preceding convolutional layer. See Table 1 for the concrete architecture in the plaintext and HE settings. Our variation of this model includes a thinner model with a width multiplicator of $k = \frac{1}{2}$ while also halving the number of filters in the first convolution layer (which in original Wide ResNet does not depend on k) in order to reduce the number of multiplications and make the application of HE feasible. We refer to these models as Wide ResNet and Tiny ResNet, respectively.

For our HE baseline, we adopted two distinct versions of CryptoNet [12]: the original version as described in [12], and an adapted variation that incorporates findings from the DP deep learning literature suggesting that increased width is a beneficial factor for DP training. Specifically, our adapted version features an increase in width by a factor of 3 for all convolutional layers of the baseline CryptoNet. We refer to the former model as CryptoNet and the latter as CryptoNet-L for the remainder of this paper.

3.2. Model training

We built and trained all models with TensorFlow and used the HeLayers Library for homomorphic encryption [2]. We performed our experiments on the Fashion-MNIST [35] and CIFAR-10 [19] datasets. Table 2 summizes the characteristics of both datasets. We trained all networks with batch sizes of 128 and learning rates of 10^{-3}. For non-DP experiments, we used the Adam Optimizer with default parameters [18], and for DP experiments, we used the differentially private counterpart of Adam im-

Table 1
DP-optimized Baseline architecture (left) and our alignment in order to support HE. Numbers in brackets denote number of filters in convolutional layers. Like [37], we use filter sizes of 3 × 3 in all Conv layers. We trained models for $k = 2$ and $k = \frac{1}{2}$. Our fully connected network (FCN) classifier consists of three layers with ReLU/Square activations and 64, 16, and 10 units respectively.

Layer	DP-optimized Architecure	HE-aligned Architecture	(HE-)WRN Block
1	Conv (16)	Conv (8)	
2	GroupNorm(4)	BatchNorm()	
3	ReLU()	Square()	
4	WRN Block 1 (16 × k)	HE-WRN Block 1 (16 × k)	
5	WRN Block 2 (32 × k)	HE-WRN Block 2 (32 × k)	
6	WRN Block 3 (64 × k)	HE-WRN Block 3 (64 × k)	
7	Average Pooling	Average Pooling	
8	Classifier FCN	Classifier FCN	

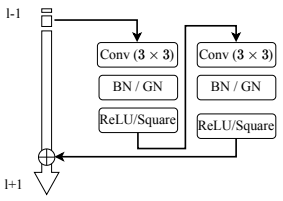


Table 2
Characteristics of the dataset under study.

Dataset	Contents	Image Dimensions	# Images (Train/Test)	# Classes
Fashion-MNIST	Clothing	28 × 28 × 1	60.000/10.000	10
CIFAR-10	Animals and Vehicles	32 × 32 × 3	50.000/10.000	10

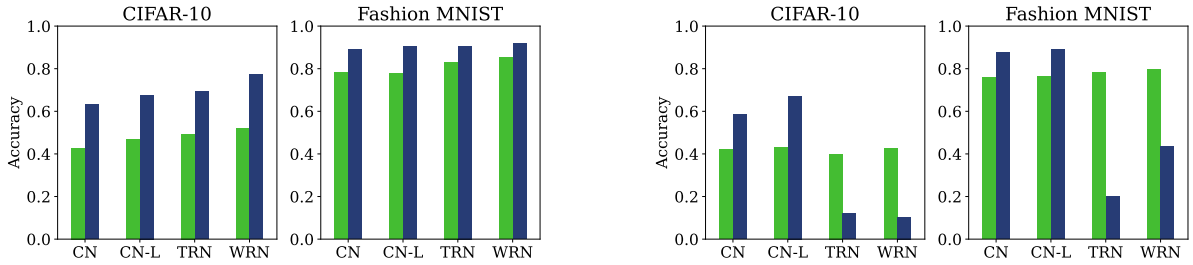
plemented in TensorFlow Privacy [13]. We trained all DP Models to suffice $(8, 10^{-5})$-DP on all datasets in the Rényi-DP setting [25].

4. Results

We conducted three experiments for which we report results in the following subsection. First, we trained all DP-optimized models, i.e. the baseline Wide ResNet, Tiny ResNet and both DP-aligned CryptoNets, with and without DP. After that, we repeated the process for our HE-optimized CryptoNets and the HE-aligned Wide ResNets. We also encrypted those models and evaluated their performance and relative inference speed drop compared to the plaintext models. Finally, we analyzed the loss landscape of the local minima identified for all models.

4.1. DP-optimized Architectures

Figure 2a summarizes the results on the DP-optimized Wide ResNets and the DP-aligned CryptoNets evaluated over the test sets of Fashion MNIST and CIFAR-10. We observe no significant differences in the performance of these models on the Fashion MNIST dataset when trained without DP. As expected, the performance gap is smaller for Wide ResNet and Tiny ResNet since those models are optimized for DP. Wide ResNet is performing best. On CIFAR-10, this finding replicates, but the gap between non-DP and DP models is larger across all models, indicating a correlation between dataset complexity and model performance under the application of DP.



(a) DP-optimized

(b) HE-aligned

Figure 2: Accuracies of all trained Models on the Fashion MNIST and CIFAR-10 datasets. ■ w/ DP, ■ w/o DP. CN(-L): CryptoNet(-L), WRN: Wide ResNet, TRN: Tiny ResNet

Table 3
RAM usage and ratio of inference time for inference on encrypted data. Both, model and data are encrypted.

Model	Fashion MNIST		CIFAR-10	
	RAM used	Inf. time scaling factor	RAM used	Inf. time scaling factor
CryptoNet	14.7 GB	158.0	12.7 GB	41.5
CryptoNet-L	31.3 GB	141.0	31.5 GB	68.7

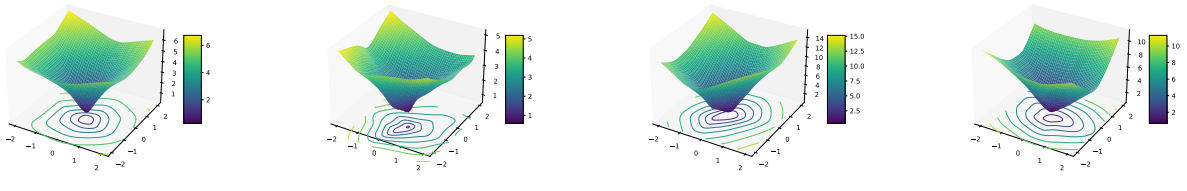
4.2. Homomorphic CNNs

Figure 2b shows the results for the HE-optimized CryptoNets and the HE-aligned Wide ResNet and Tiny ResNet. Notably, the performance of both CryptoNets does not change drastically when using HE building blocks instead of DP counterparts. In contrast, both ResNet architectures perform terrible in the HE-aligned setting. For CIFAR-10, the non-DP models completely fail to make accurate predictions on the test set. The predictions for Fashion MNIST are also poor, as can be seen by the large drop in prediction accuracy. Interestingly, the DP models perform better on both datasets, indicating a regularizing effect of DP on these model architectures. In Table 3 we report the relative difference in inference speed for all homomorphically encrypted networks compared to their plaintext counterpart, as well as used RAM for the prediction. We were unable to compute results for both ResNet architectures, since our workstation[1] ran out of memory for those models. We only report the average relative increase in inference time for a single data sample, since wall clock time depends strongly on the hardware used for evaluation.

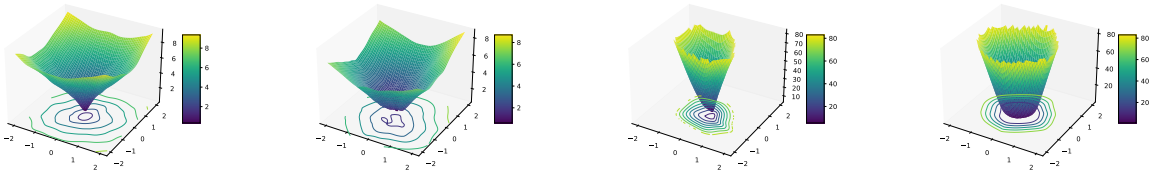
4.3. Visualizations of local minima

We applied the visualization method proposed by Li et al. [22] to all models and evaluated the implications of DP/HE-alignment on the structure of the loss landscapes. The results are depicted in Figure 3. We observe different effects for HE-aligned models, as well as for models trained with DP and without DP. For CryptoNet and Tiny ResNet, the application of DP introduces visible distortions to the loss-landscape in the model architectures without HE-alignment (subfigures 3b and 3f) compared to the non-DP models (subfigures 3a and 3e). This can be particularly observed in the contour lines that are projected onto the xy-planes. When applying both, HE and DP, the loss surface is less distorted (subfigures 3d and 3h). Across all models, the application of HE yields steeper valleys and thus, sharper minimizers (subfigures 3c, 3d, 3g and 3h), while the application of DP in this setting

[1]Intel® Xeon® Silver 4114 CPU 2.20 GHz and 64 GB of DDR4 RAM



(a) CN,−HE,−DP (b) CN,−HE,DP (c) CN,HE,−DP (d) CN,HE,DP



(e) TRN,−HE,−DP (f) TRN,−HE,DP (g) TRN,HE,−DP (h) TRN,HE,DP

Figure 3: Loss-Landscapes for CryptoNets (a-d) and Tiny ResNet (e-h). Best viewed digitally and zoomed in. All z-values are on log-scale. Abbreviations in subfigure captions are: CN: CryptoNet, TRN: Tiny ResNet, HE: Architecture for HE, DP: model trained with DP. −HE and −DP means without HE or DP, respectively. Models are trained and evaluated on CIFAR-10. Loss values are computed using categorical cross-entropy.

seems to attenuate this effect and thus regularize the optimization in the HE setting (subfigures 3d and 3h), especially for Tiny ResNet.

5. Discussion

Our results confirm the importance of dedicated model architectures for the application of DP in order to close the performance deterioration induced by DP. Wide ResNet and Tiny ResNet did perform better in the non-HE setting than both CryptoNets in our experiments. We note at this point that our Tiny ResNet comprises ~$29k$ parameters, whereas CryptoNet comprises ~$28k$ parameters. Both models have significantly less parameters than CryptoNet-L (~$119k$) and Wide ResNet (~$336k$). Since Tiny ResNet outperforms CryptoNet-L with and without DP, the parameter count of a model does not seem to be the most important architectural property for model utility. Our results indicate that the network topology has a direct impact on utility when applying DP. Especially on Fashion MNIST, the utility gap between the DP and non-DP models is smaller compared to the CryptoNets. However, on CIFAR-10 this result is not as straightforward. While both ResNets outperform CryptoNets, the gap between DP and non-DP models is approximately similar for all models, indicating that regarding more complex datasets other tools are required to close this gap. Towards this, [8, 20] have leveraged data augmentation, learning rate schedules, and other tweaks to the training procedure in order to obtain tighter utility gaps. Hence, architecture alone can help to close this gap, but does not suffice on its own.

Aligning more complex model architectures like ResNets for homomorphic encryption strongly deteriorates model utility. We have observed that replacing ReLU with square activation, max pooling with average pooling, and group norm with batch norm results in model architectures with sharp minima which in our experiments completely failed to generalize to the test set. Training these models with DP, however, seems to yield a regularizing effect that increases model utility to a level that is competitive with the results of DP-trained CryptoNets. However, encrypted evaluation of both ResNet

architectures failed using the HELayers library [2] because of memory restrictions. Presumably, this is due to the presence of skip connections. Without these, each layer including its corresponding inputs can be loaded separately into the memory. Concerning skip connections, the intermediary results of multiple layers need to be kept in memory, thus resulting in higher RAM allocations that exceeded our 64GB main memory even for the Tiny ResNet. We note that one could evade this problem through engineering effort, i.e., providing more efficient implementations for the inference of skip connections. However, we have not conducted any further investigation and leave it to future work.

Overall, we note that the increase in inference time and RAM usage is significant for the application of HE on deep learning, even for small models as CryptoNets on small toy datasets such as Fashion-MNIST and CIFAR-10. While related work such as [33] have investigated the scalability of HE-enabled deep learning on datasets with higher complexity and dimensionality, they do not consider complex network architectures as our Tiny ResNet, which yield better results in the plaintext setting while having an approximately similar amount of parameters. In the computer vision literature, residual networks and other complex architectures with millions of parameters yield state of the art results. Given the computational complexity and resource requirements for our small networks, applying HE to state of the art models is intractable in real world use cases.

5.1. Limitations

We have limited our study to minimal requirements for homomorphic encryption. Existing work on HE-enabled deep learning comprises model architectures with polynomial approximations of activation functions of higher degree. Using these approximations could close the observed gap between the DP-optimized architectures and their HE-aligned counterpart. On the other hand, it may lead to further increase in computational complexity of the inference with homomorphically encrypted data. Future work should investigate upon that. In addition to that, the loss landscape visualization that we have used only provides a high level idea about the generalizability of the model architectures. Other methods have been proposed in the literature that may give more insight. However, this is still a very active area of research.

6. Conclusion

In conclusion, our research findings underline the importance of dedicated model architectures for the application of differential privacy (DP) in order to close the performance deterioration induced by DP. Specifically, we have observed that Wide ResNet and Tiny ResNet perform better in the non-HE setting than CryptoNets. Moreover, the network topology seems to have a great impact on utility when applying DP. While Tiny ResNet outperforms CryptoNet-L with and without DP, our results indicate that the number of model parameters is not the most significant architectural property regarding model utility. Furthermore, we have observed that aligning more complex model architectures like ResNets for homomorphic encryption strongly deteriorates model utility. While related work has investigated the scalability of HE-enabled deep learning on datasets with higher complexity and dimensionality, they do not consider complex network architectures like our Tiny ResNet, which yield better results in the plaintext setting while having an approximately similar amount of parameters. However, evaluating ResNets in an encrypted setting drastically increases the resource requirements compared to CryptoNets, thus rendering homomorphic encryption impractical for these model architectures.

References

[1] Abadi, M., Chu, A., Goodfellow, I., McMahan, H.B., Mironov, I., Talwar, K., Zhang, L.: Deep learning with differential privacy. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. p. 308–318. CCS '16, Association for Computing Machinery, New York, NY, USA (2016). https://doi.org/10.1145/2976749.2978318

[2] Aharoni, E., Adir, A., Baruch, M., Drucker, N., Ezov, G., Farkash, A., Greenberg, L., Masalha, R., Moshkowich, G., Murik, D., Shaul, H., Soceanu, O.: HeLayers: A Tile Tensors Framework for Large Neural Networks on Encrypted Data. Privacy Enhancing Technology Symposium (PETs) 2023 (2023), https://petsymposium.org/2023/paperlist.php

[3] Al Badawi, A., Jin, C., Lin, J., Mun, C.F., Jie, S.J., Tan, B.H.M., Nan, X., Aung, K.M.M., Chandrasekhar, V.R.: Towards the alexnet moment for homomorphic encryption: Hcnn, the first homomorphic cnn on encrypted data with gpus. IEEE Transactions on Emerging Topics in Computing **9**(3), 1330–1343 (2021). https://doi.org/10.1109/TETC.2020.3014636

[4] Cabrero-Holgueras, J., Pastrana, S.: SoK: Privacy-Preserving Computation Techniques for Deep Learning. Proceedings on Privacy Enhancing Technologies (2021), https://petsymposium.org/popets/2021/popets-2021-0064.php

[5] Chabanne, H., de Wargny, A., Milgram, J., Morel, C., Prouff, E.: Privacy-preserving classification on deep neural network. Cryptology ePrint Archive, Paper 2017/035 (2017), https://eprint.iacr.org/2017/035

[6] Chen, H., Hussain, S.U., Boemer, F., Stapf, E., Sadeghi, A.R., Koushanfar, F., Cammarota, R.: Developing Privacy-preserving AI Systems: The Lessons learned. In: 2020 57th ACM/IEEE Design Automation Conference (DAC). pp. 1–4 (Jul 2020). https://doi.org/10.1109/DAC18072.2020.9218662, iSSN: 0738-100X

[7] Dauphin, Y.N., Pascanu, R., Gulcehre, C., Cho, K., Ganguli, S., Bengio, Y.: Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. In: Proceedings of the 27th International Conference on Neural Information Processing Systems. vol. 2, p. 2933–2941. MIT Press, Cambridge, MA, USA (2014)

[8] De, S., Berrada, L., Hayes, J., Smith, S.L., Balle, B.: Unlocking High-Accuracy Differentially Private Image Classification through Scale (Jun 2022). https://doi.org/10.48550/arXiv.2204.13650

[9] Dinh, L., Pascanu, R., Bengio, S., Bengio, Y.: Sharp minima can generalize for deep nets. In: Precup, D., Teh, Y.W. (eds.) Proceedings of the 34th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 70, pp. 1019–1028. PMLR (06–11 Aug 2017), https://proceedings.mlr.press/v70/dinh17b.html

[10] Dwork, C.: Differential privacy. In: Bugliesi, M., Preneel, B., Sassone, V., Wegener, I. (eds.) Automata, Languages and Programming. pp. 1–12. Springer Berlin Heidelberg (2006)

[11] Dziugaite, G.K., Roy, D.M.: Data-dependent pac-bayes priors via differential privacy. In: Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 31. Curran Associates, Inc. (2018), https://proceedings.neurips.cc/paper/2018/file/9a0ee0a9e7a42d2d69b8f86b3a0756b1-Paper.pdf

[12] Gilad-Bachrach, R., Dowlin, N., Laine, K., Lauter, K., Naehrig, M., Wernsing, J.: Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy. In: Balcan, M.F., Weinberger, K.Q. (eds.) Proceedings of The 33rd International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 48, pp. 201–210. PMLR, New York, New York, USA (20–22 Jun 2016), https://proceedings.mlr.press/v48/gilad-bachrach16.html

[13] Google: Tensorflow privacy. GitHub Repository: https://github.com/tensorflow/privacy (2018)

[14] Ibarrondo, A., Önen, M.: Fhe-compatible batch normalization for privacy preserving deep learn-

ing. In: Garcia-Alfaro, J., Herrera-Joancomartí, J., Livraga, G., Rios, R. (eds.) Data Privacy Management, Cryptocurrencies and Blockchain Technology. pp. 389–404. Springer International Publishing (2018)
[15] Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: Bach, F., Blei, D. (eds.) Proceedings of the 32nd International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 37, pp. 448–456. PMLR, Lille, France (07–09 Jul 2015), https://proceedings.mlr.press/v37/ioffe15.html
[16] Kaissis, G., Ziller, A., Passerat-Palmbach, J., Ryffel, T., Usynin, D., Trask, A., Lima, I., Mancuso, J., Jungmann, F., Steinborn, M.M., Saleh, A., Makowski, M., Rueckert, D., Braren, R.: End-to-end privacy preserving deep learning on multi-institutional medical imaging. Nature Machine Intelligence **3**(6), 473–484 (Jun 2021). https://doi.org/10.1038/s42256-021-00337-8
[17] Keskar, N.S., Mudigere, D., Nocedal, J., Smelyanskiy, M., Tang, P.T.P.: On large-batch training for deep learning: Generalization gap and sharp minima. In: International Conference on Learning Representations (2017)
[18] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: Bengio, Y., LeCun, Y. (eds.) 3rd International Conference on Learning Representations (2015)
[19] Krizhevsky, A.: Learning multiple layers of features from tiny images. Tech. rep. (2009)
[20] Kurakin, A., Song, S., Chien, S., Geambasu, R., Terzis, A., Thakurta, A.: Toward Training at ImageNet Scale with Differential Privacy (Feb 2022). https://doi.org/10.48550/arXiv.2201.12328
[21] Ledent, A., Mustafa, W., Lei, Y., Kloft, M.: Norm-based generalisation bounds for deep multi-class convolutional neural networks. Proceedings of the AAAI Conference on Artificial Intelligence **35**(9), 8279–8287 (May 2021). https://doi.org/10.1609/aaai.v35i9.17007
[22] Li, H., Xu, Z., Taylor, G., Studer, C., Goldstein, T.: Visualizing the loss landscape of neural nets. In: Neural Information Processing Systems (2018)
[23] Lotfi, S., Finzi, M.A., Kapoor, S., Potapczynski, A., Goldblum, M., Wilson, A.G.: PAC-bayes compression bounds so tight that they can explain generalization. In: Oh, A.H., Agarwal, A., Belgrave, D., Cho, K. (eds.) Advances in Neural Information Processing Systems (2022)
[24] McMahan, H.B., Ramage, D., Talwar, K., Zhang, L.: Learning differentially private recurrent language models. In: International Conference on Learning Representations (2017)
[25] Mironov, I.: Rényi differential privacy. In: 2017 IEEE 30th Computer Security Foundations Symposium (CSF). pp. 263–275 (2017). https://doi.org/10.1109/CSF.2017.11
[26] Nandakumar, K., Ratha, N., Pankanti, S., Halevi, S.: Towards deep neural network training on encrypted data. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 40–48 (2019). https://doi.org/10.1109/CVPRW.2019.00011
[27] Polyak, B.T., Juditsky, A.B.: Acceleration of stochastic approximation by averaging. SIAM Journal on Control and Optimization **30**(4), 838–855 (1992). https://doi.org/10.1137/0330046
[28] Qiao, S., Wang, H., Liu, C., Shen, W., Yuille, A.: Micro-batch training with batch-channel normalization and weight standardization (2019). https://doi.org/10.48550/ARXIV.1903.10520
[29] Remerscheid, N.W., Ziller, A., Rueckert, D., Kaissis, G.: Smoothnets: Optimizing cnn architecture design for differentially private deep learning. arXiv preprint arXiv:2205.04095 (2022)
[30] Rivest, R.L., Adleman, L., Dertouzos, M.L.: On data banks and privacy homomorphisms. Foundations of Secure Computation, Academia Press pp. 169–179 (1978)
[31] Sander, T., Stock, P., Sablayrolles, A.: Tan without a burn: Scaling laws of dp-sgd. arXiv preprint arXiv:2210.03403 (2022)
[32] Torkzadehmahani, R., Nasirigerdeh, R., Blumenthal, D.B., Kacprowski, T., List, M., Matschinske, J., Spaeth, J., Wenke, N.K., Baumbach, J.: Privacy-Preserving Artificial Intelligence Techniques in Biomedicine. Methods of Information in Medicine **61**, e12–e27 (Jan 2022).

https://doi.org/10.1055/s-0041-1740630
[33] Wingarz, T., Gomez-Barrero, M., Busch, C., Fischer, M.: Privacy-Preserving Convolutional Neural Networks Using Homomorphic Encryption. In: 2022 International Workshop on Biometrics and Forensics (IWBF). pp. 1–6 (Apr 2022). https://doi.org/10.1109/IWBF55382.2022.9794535
[34] Wu, Y., He, K.: Group normalization. International Journal of Computer Vision **128**, 742–755 (2018)
[35] Xiao, H., Rasul, K., Vollgraf, R.: Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. CoRR **abs/1708.07747** (2017), http://arxiv.org/abs/1708.07747
[36] Xiong, A., Nguyen, M., So, A., Chen, T.: Privacy Preserving Inference with Convolutional Neural Network Ensemble. In: 2020 IEEE 39th International Performance Computing and Communications Conference (IPCCC). pp. 1–6 (Nov 2020). https://doi.org/10.1109/IPCCC50635.2020.9391544, iSSN: 2374-9628
[37] Zagoruyko, S., Komodakis, N.: Wide residual networks. In: British Machine Vision Conference (2016)

Digitalisation in voluntary work in Germany — A qualitative study of IT acceptance criteria

Simone Dogu[1], Stefanie Regier[1], Ingo Stengel[1] and Nathan Clarke[2]

[1]*Faculty of Computer Science and Business Information Systems, Karlsruhe University of Applied Sciences, Germany*
[2]*Faculty of Science and Engineering, University of Plymouth, UK*

Abstract
Digitalisation is a central social trend of the 21st century. Digital information and communication technologies are gaining importance in almost all public and private areas of life. The possibilities of digitalisation are also increasingly being used in the area of volunteering, such as the appointment communication of an exercise instructor in a sports club via email. Whether and how volunteers can use digital technologies for their voluntary work depends, among other things, on whether they have access to the Internet and whether the corresponding infrastructure is available to them. In addition to the technical requirements, the intention to use IT in volunteer work also plays an important role. This study is intended to help to learn more about possible acceptance criteria for the use of digital tools in volunteer work.

Keywords
Acceptance criteria, IT, volunteering, expert interviews, digitalisation

1. Introduction

Volunteering is an important pillar of our modern society. In 2019, 28.8 million (39.7 percent) people in Germany aged 14 and over were involved in voluntary work. Between 1999 and 2019, the proportion of volunteers has increased. Whether taking on voluntary positions in local councils, participating in citizens' initiatives, supporting local sports clubs or environmental protection: the tasks and activities are very diverse [3]. The special feature of voluntary work is that it is based on voluntariness and thus differs from full-time work, for example in different resources. Volunteering has changed over the last twenty years. On one hand, the proportion of committed people who spend a lot of time and take on leadership functions in the commitment has decreased. Instead, more and more volunteers carry out their activities in an informally organised framework, which is usually accompanied by flat hierarchical structures and requires fewer leadership and board positions [6]. Along with this, volunteering is currently increasingly shaped by the topic of digitalisation, which also opens up new opportunities. Digital information and communication technologies are gaining importance in almost all public and private areas of life. The possibilities of digitalisation are also being used in the field of volunteering. In many cases, digitalisation of volunteering is about supporting voluntary activities that continue

CERC 2023: Collaborative European Research Conference, June 09–10, 2023, Barcelona, Spain
simone.dogu@h-ka.de (S. Dogu); stefanie.regier@h-ka.de (S. Regier); ingo.stengel@h-ka.de (I. Stengel); nathan.clarke@plymouth.ac.uk (N. Clarke)
0000-0002-3595-3800 (N. Clarke)

ISBN 978-3-96187-021-9
https://doi.org/10.48444/h_docs-pub-518

to take place 'analogue', such as organising a local regulars' table via email or rescheduling training at the sports club at short notice via social network groups. Likewise, new forms of volunteering can also be subsumed under the digitalisation of engagement, such as organising an online regulars' table [7] [8]. In the meantime, more than half of the volunteers use the Internet in the context of their voluntary activities. The proportion of volunteers who use the internet as part of their voluntary work only increased between 2004 and 2009, but has remained fairly stable since then. Currently, the internet is only used by slightly more than 50 percent. This also means that a large proportion of volunteers (more than 40 per cent) do not use the Internet for their activities, for example because it is not relevant for carrying out certain tasks [8]. Digital technologies thus play a major role in the voluntary activities of many volunteers, but this is by no means the case for all people who are active as volunteers [22]. To confirm the central assumption of the present study that attitudinal and behavioural dimensions play a crucial role in volunteers' decision-making on the use of digital tools, this research topic currently lacks detailed research.

The overall research design aims to provide insights into possible IT acceptance criteria and is based on a mixed-methods approach. It includes a qualitative as well as quantitative study. The subject of this paper is the design, implementation and results of the qualitative study. The aim of this is to learn more about the acceptance criteria of volunteers in the context of their activities, based on existing research, and to create a sufficiently well-founded and meaningful theoretical basis for the subsequent quantitative study.

2. Related research and research questions

In the following, an overview of the relevant state of research in the area of attitudinal and behavioural dimensions in the use of digital tools by volunteers will be given.

In the field of volunteering and NGOs, there is a large number of country-specific studies, most of which are descriptive. In Germany, there is one longitudinal study: the German Volunteer Survey is a representative survey on volunteering in Germany. It is conducted every five years since 1999 and targeting people aged 14 and older. It is the most comprehensive and detailed quantitative survey on civic engagement in Germany. Voluntary activities and willingness to get involved are surveyed in telephone interviews and can be presented by population groups and parts of the country (FWS 2019: N=27,759). In addition, changes in the forms and contexts of volunteering can be traced. In addition, those who are engaged and those who are not or no longer engaged can be described. In addition to motives and motivations as well as reasons for termination and obstacles to volunteering, such as lack of time [1], this survey also provides insights into volunteers' Internet use [22]. In addition, there are studies on the importance of volunteering in Germany, also against the backdrop of societal change [11], as well as on differences in terms of gender, age groups, schooling, and employment [20]. The social areas of volunteering were also examined [10], with most volunteering taking place in the areas of sports and exercise, culture and music, and the social sector.

In their review study, Robinson et al [18] developed a typology of projects as well as insights into the use and expectations of corresponding project tools for use in civil society projects. Lau et al [12] investigated the motivations of volunteers to get involved in and support social

projects.

As early as 2007, Zhang and Gutierrez [23] used the Theory of Planned Behaviour (TPB) to investigate various factors influencing IT acceptance (e.g. Perceived Personal Usefulness, Peer Influence, Self-Efficacy, User Intention to Use IT) in the environment of nonprofit organisations.

Taking into account the increasing complexity, Technology Acceptance Model (TAM) was expanded to include supplementary factors. Saura et al [19] investigated the use of digital tools to recruit new volunteers using an extended TAM. The results of this research show that communication strategies are of utmost importance on digital channels for NGOs and platforms that rely on volunteers. Thus, influencing factors such as Perceived Usefulness, Perceived ease of Use, Trust and their importance were demonstrated.

Table 1 presents relevant study results in the research area of volunteering. The studies are both qualitative, quantitative in nature and also mixed-methods approaches. Most of them focus on specific aspects of volunteering, such as motives and motivations or areas of volunteering. The sample sizes of the studies shown vary. Some samples are very small (e.g. Zhang and Gutierrez 2007 [23]), so no representativeness can be guaranteed. Another challenge is the country reference of the studies and the transferability of the results, as there are country-specific differences in their volunteer structures. Currently, there are a large number of studies in various countries that primarily deal with sociodemographic information, motives and motivations of volunteers. Regarding to relevant IT acceptance criteria, there are many studies that deal with individual influencing factors like innovativeness.

A holistic model to explain IT acceptance in volunteering is currently still missing. This study aims to contribute to closing this research gap. In consequence of the theoretical as well as practical importance of attitude formation in respective technology markets, the study's object of cognition will focus on the IT acceptance of digital platforms in voluntary work. Thus, the overall aim of this study is to identify and explain key factors that lead to either acceptance or rejection of using digital tools by volunteers. Building on this interest and considering the stated insufficiencies and knowledge gaps in diffusion and information technology research, this study will be guided by the following fundamental research questions:

- What main components are involved in the process of attitude formation toward using new technologies in volunteering?
- What factors determine the acceptance or resistance decision toward digital tools in volunteering?
- In terms of causality, how can these factors be suitably contextualised and transferred into an explanatory model of IT acceptance in volunteering?

3. Method

In order to learn more about the attitudinal and behavioural dimensions of volunteers' IT acceptance criteria when using digital tools, a mixed-methods approach was chosen for the overall study. This method is suitable because currently there is a lack of empirical-inductive research on this topic. Since this paper presents the qualitative study, the focus is on the selection of qualitative data collection and analysis. According to Mack et al [14], the most

Table 1
Overview of Relevant Research

Author(s), Year	Title	Research and Evaluation Methods	Results
[1] Arriagada, C. & Karnick, N., 2022	Motive für freiwilliges Engagement, Beendigungsgründe, Hinderungsgründe und Engagementbereitschaft	N=27.759, Interviews, Multivariate Analysis (e.g. Factor Analysis)	The study shows general information about motives and reasons for termination of volunteers.
[11] Kausmann, C., Kelle, N., Simonson, J., Tesch-Römer, C., 2022	Freiwilliges Engagement — Bedeutung für Gesellschaft und Politik	N=27.759, Interviews, Descriptive Analysis (e.g. Frequencies)	The study examines differences and inequalities in volunteering, development of volunteering against the background of societal change.
[10] Kausmann, C. & Hagen, C., 2022	Gesellschaftliche Bereiche des freiwilligen Engagements	N=27.759, Interviews, Descriptive Analysis (e.g. Frequencies)	The study shows the TOP 3 areas of volunteerism in Germany: Sports and Exercise, Culture and Music, Social Sector.
[12] Lau et al., 2019	Volunteer motivation, social problem solving, self-efficacy, and mental health: a structural equation model approach	N=1.530, Online survey, Structural Equation Model (SEM)	The study shows that volunteer motivation provides a way to enhance social problem-solving ability and self-efficacy.
[18] Robinson, J.A. et al., 2021	Meeting volunteer expectations — a review of volunteer motivations in citizen science and best practices for their retention through implementation of functional features in CS tools	Review Article	The study creates and summarizes a typology of projects and insights into the use and expectations of appropriate project tools.
[19] Saura, JR. et al., 2020	What Drives Volunteers to Accept a Digital Platform That Supports NGO Projects?	N=245, Online survey, Structural Equation Model (SEM)	The study shows that digital platforms attract motivated potential young volunteers. IT acceptance criteria such as perceived usefulness, perceived ease of use, trust have influence.
[20] Simonson, J. et al., 2022	Unterschiede und Ungleichheiten im freiwilligen Engagement	N=27.759, Interviews, Descriptive Analysis	Examination of differences in terms of gender, age groups, school education, employment, and comparison of East and West Germany, among other things.
[22] Tesch-Römer, C. & Huxhold, O., 2021	Nutzung des Internets für die freiwillige Tätigkeit	N=27.759, Interviews, Descriptive Analysis	The study shows findings on Internet use in volunteer work.
[23] Zhang, W., & Gutierrez, O., 2007	Information Technology Acceptance in the Social Services Sector Context: An Exploration.	N=61, Online survey, Structural Equation Model (SEM)	The study demonstrates relevant IT acceptance criteria such as perceived personal usefulness, peer influence, self-efficacy, user intention to use.

common qualitative methods are participant observation, in-depth interviews and focus groups. Since the present study aims to collect attitudinal factors, direct observations may not be feasible. In addition, focus groups have the fundamental disadvantage that their results depend on group dynamics, possibly leading to biased results [2]. In-depth interviews, on the other hand, make it possible to collect the "complex body of knowledge" [4] of people about the topic under investigation. Since personal individual interviews are very well suited to elicit salient behavioural, normative and efficacy beliefs [15], this qualitative research method fits the

aforementioned goals of this work.

A special form of the semi-structured interview is the expert interview, which refers neither to very open nor to rigidly structured question-answer schemes [4]. Characteristic for the expert interview is the specific focus on the context of the content. According to Gläser and Laudel [5], the selection of experts is therefore consequently oriented towards

- the selection of certain organisations,
- the reputation and position of relevant actors or persons,
- the possibilities to influence persons in relevant decisions or actions.

Furthermore, practical research factors, such as the accessibility and willingness of the interview partners, also play a decisive role in the selection of experts [5]. The evaluation is carried out by means of a qualitative content analysis. It aims to transform collected text material into meaningful findings for the present study [17]. According to Hsieh and Shannon [9], these can be roughly divided into three methods: conventional, directed and summary analysis. Summary content analysis aims to discover the underlying meanings of expressions by analysing the occurrence of a particular word or content in the collected data. Given the objectives of the upstream exploratory research phase, summary content analysis is conducted in this paper, resulting in formed categories that will be considered in the further course of the study in model development.

This research shall build a well-founded and meaningful theoretical basis for the formation of hypotheses. Therefore, a mixed-methods approach was chosen for this study. Initially, a qualitative approach (expert interviews) was chosen with a small sample size of expert interviews to achieve more depth into the research topic, followed by a quantitative approach (survey) with a bigger sample size.

4. Analysis and results

In total, six volunteers working in different functions from different organisations were selected and interviewed as part of the qualitative study. All interviewees had many years of experience in different volunteer roles and organisations and are involved in topics such as digitalisation as well as the execution of different activities. The results of the qualitative study provide a variety of insights into the acceptance criteria for the use of digital tools in volunteering. Overall, several theoretical constructs were derived from the qualitative study that describe the personal experiences and attitudes of the selected respondents towards the use of digital tools in their volunteering activities. These concepts, together with the literature analysis, serve as the basis for the theory-based derivation of the model by providing important aspects and determinants of acceptance behaviour. In the following, the different focal points from the interviews are elaborated and presented.

Five of the interviewees addressed the importance of digitalisation and the associated changes in everyday life during the interviews. Four of the interviewees stated that fears of innovations or new developments have an influence on the use of digital tools in voluntary work and can therefore prevent it. The fact that early adopters/young people have it easier here was mentioned by three of the respondents. The fact that the age of the volunteer is decisive in

the use of digital tools in the context of the activity was explicitly mentioned by four of the respondents.

According to all six interviewees, openness to new things is highly relevant in relation to voluntary work and digitalisation. The openness to try out new tools in the context of voluntary work is reflected in statements such as "The openness is partly there, but partly there is also the attitude: in the old way, the non-digital way, it has worked well so far. So a bit of the new as the enemy." (I1) and "Sometimes you just have to try it out." (I4)

The topic of IT security also played an important role in the expert interviews conducted. Thus, five of the six interviewees addressed the topic of "data protection". However, not only concerns and caution were expressed. In addition to statements about a rather cautious behaviour in dealing with digital tools, there were also statements such as "I personally have a rather uninformed attitude towards this" (I1).

Topics such as "system reliability", "legal certainty" and "protection of minors" were only mentioned by one respondent each, but again underline the importance of the security aspect in the use of digital tools in voluntary work in general.

How important the topics of user-friendliness and accessibility are for voluntary use is shown by statements such as "Lack of usability is a reason for non-use for me". (I1) or "I also see accessibility as important." (I5). Five of the respondents addressed these issues during the interviews, which indicates that this aspect can also be relevant for an acceptance model.

Perceived usefulness was addressed by four of the interviewees. They see added value mainly in efficiency enhancement (4), better communication (4) and central data storage (4). If the use is perceived as added value, the acceptance of the use of digital tools increases. This becomes clear through statements such as "Keeping in touch and communicating with each other is important, so not only making appointments, but also spontaneous exchanges, e.g. in case of problems or if something needs to be organised, I find important". (I6)

The topic of support for volunteers also proved to be of great importance, i.e. if corresponding support offers were available, they would be more likely to use digital tools in the context of their voluntary work. There are two subcategories: "technical support" and "help with use". The subcategory "technical support" refers to support for purely technical problems, which two of the respondents stated in their interviews. The subcategory "help with use", on the other hand, describes support or assistance with the use of digital tools, such as conducting a video conference. This was mentioned by five of the respondents. This is supported by statements such as "And maybe because there is no one to teach us professionally. If I'm the most IT-savvy person on the board with my limited IT knowledge, that's just a statement." (I1) and "Training is definitely important." (I4).

Equipment emerged as another important point. All six interviewees mentioned in their interviews the importance of appropriate software to perform their jobs. Two of the six interviewees addressed the importance of appropriate hardware ("I just need a computer with appropriate prerequisites." (I3))

5. Conclusion

Digitalisation is a central social trend of the 21st century. Digital information and communication technologies are gaining importance in almost all public and private areas of life, including voluntary work. In order to be able to carry out tasks and activities, the internet and various digital tools are used, among other things. In order to learn more about the attitudinal and behavioural dimensions, a mixed-methods approach was chosen for the study. Based on extensive literature research, an interview guideline was developed for the qualitative part and six semi-structured interviews with experts were conducted. The experts have extensive knowledge and experience in their respective organisations through many years of voluntary work(s).

The results obtained through the interviews provide important insights into acceptance factors in the use of digital tools in volunteering. Through a qualitative content analysis, several categories were identified as relevant to volunteering. The main topics mentioned by at least half of the experts interviewed include following topics which should be taken into account for further research:

- Openness to new things (e.g. [21])
- Data protection concerns (e.g. [13])
- Suitable software
- Support / help with application
- Fears of novelty / innovation (e.g. [13] [16])
- Age

These need to be theoretically substantiated in the further course of the study and existing models in acceptance research (such as TAM) need to be modified so that further relevant insights into attitude and behavioural dimensions can be gained within the framework of a quantitative study. In the future, these can help to create appropriate offers for volunteers that support them in carrying out their activities and tasks.

References

[1] Arriagada, C., Karnick, N.: Motive für freiwilliges engagement, beendigungsgründe, hinderungsgründe und engagementbereitschaft. In: Simonson, J., Kelle, N., Kausmann, C., Tesch-Römer, C. (eds.) Freiwilliges Engagement in Deutschland – Der Deutsche Freiwilligensurvey 2019. Deutsches Zentrum für Altersfragen, Berlin, Germany (2021)

[2] Churchill, G.A., Iacobucci, D.: Marketing research: methodological foundations, 10th edition. SouthWestern Cengage Learning (2010)

[3] Deutscher Bundestag: Dritter Engagementbericht. Zukunft Zivilgesellschaft Junges Engagement im digitalen Zeitalter und Stellungnahme der Bundesregierung (Drucksache 19/19320). Deutscher Bundestag, Berlin, Germany (2020)

[4] Flick, U.: Qualitative Sozialforschung. Eine Einführung, 6th edition. Rowohlt, Reinbek / Hamburg, Germany (2007)

[5] Gläser, J., Laudel, G.: Experteninterviews und qualitative Inhaltsanalyse. Als Instrument rekonstruierender Untersuchungen. Verlag für Sozialwissenschaften, Wiesbaden, Germany (2010)
[6] Hagen, C., Simonson, J.: Inhaltliche ausgestaltung und leitungsfunktionen im freiwilligen engagement. In: Simonson, J., Vogel, C., Tesch-Römer, C. (eds.) Freiwilliges Engagement in Deutschland – Der Deutsche Freiwilligensurvey 2014. Springer VS, Wiesbaden, Germany (2017)
[7] Heinze, R.G., Beckmann, F., Schönauer, A.L.: Die digitalisierung des engagements: Zwischen hype und disruptivem wandel. In: Heinze, R.G., Kurtenbach, S., Überbacker, J. (eds.) Digitalisierung und Nachbarschaft. Erosion des Zusammenlebens oder neue Vergemeinschaftung? Nomos Verlag, Baden-Baden, Germany (2019)
[8] Hinz, U., Wegener, N., From, M.W.: Digitales Bürgerschaftliches Engagement. FOKUS, Berlin, Germany (2014), https://www.oeffentliche-it.de/documents/10181/14412/Digitales+B%C3%BCrgerschaftliches+Engagement
[9] Hsieh, H.F., Shannon, S.: Three approaches to qualitative content analysis. In: Qualitative health research, Volume 15 (0) (2005). https://doi.org/10.1177/1049732305276687
[10] Kausmann, C., Hagen, C.: Gesellschaftliche bereiche des freiwilligen engagement. In: Simonson, J., Kelle, N., Kausmann, C., Tesch-Römer, C. (eds.) Freiwilliges Engagement in Deutschland – Der Deutsche Freiwilligensurvey 2019. Deutsches Zentrum für Altersfragen, Berlin, Germany (2021)
[11] Kausmann, C., Kelle, N., Simonson, J., Tesch-Römer, C.: Freiwilliges engagement – bedeutung für gesellschaft und politik. In: Simonson, J., Kelle, N., Kausmann, C., Tesch-Römer, C. (eds.) Freiwilliges Engagement in Deutschland – Der Deutsche Freiwilligensurvey 2019. Deutsches Zentrum für Altersfragen, Berlin, Germany (2021)
[12] Lau, Y., Fang, L., Cheng, L.J., Kwong, H.K.D.: Volunteer motivation, social problem solving,self-efficacy, and mental health: a structural equation model approach. In: Educational Psychology, Volume 39. Routledge (2019). https://doi.org/10.1080/01443410.2018.1514102
[13] Li, X., Liu, X., Motiwalla, L.: Valuing personal data with privacy consideration. In: Decision Sciences, Volume 52 (2020). https://doi.org/10.1111/deci.12442
[14] Mack, N., et al: Qualitative research methods: a data collectors field guide. Family Health International, North Carolina (2005)
[15] Montano, D., Kasprzyk, D., Glanz, K., Rimer, B., Viswanath, K.: Theory of reasoned action, theory of planned behavior, and the integrated behavior model. In: Health behavior and health education: Theory, research, and practice (2008)
[16] Nader, L., de Campos, J.G.F.: Five reasons to fear innovation, SUR 23 (2016), https://sur.conectas.org/en/five-reasons-fear-innovation
[17] Patton, M.Q.: Qualitative Evaluation and Research Methods. SAGE Publications Inc, Newbury Park, US (1990)
[18] Robinson, J.A., Kocman, D., Speyer, O., Gerasopoulos, E.: Meeting volunteer expectations – a review of volunteer motivations in citizen science and best practices for their retention through implementation of functional features in cs tools. In: Journal of Environmental Planning and Management, Volume 64, Number 12. Routledge (2021). https://doi.org/10.1080/09640568.2020.1853507

[19] Saura, J.R., Palos-Sanchez, P., Velicia-Martin, F.: What drives volunteers to accept a digital platform that supports ngo projects? In: Frontiers in Psychology, Volume 11 (2020). https://doi.org/10.3389/fpsyg.2020.00429

[20] Simonson, J., Kelle, N., Kausmann, C., Tesch-Römer, C.: Unterschiede und ungleichheiten im freiwilligen engagement. In: Simonson, J., Kelle, N., Kausmann, C., Tesch-Römer, C. (eds.) Freiwilliges Engagement in Deutschland – Der Deutsche Freiwilligensurvey 2019. Deutsches Zentrum für Altersfragen, Berlin, Germany (2021)

[21] Svendsen, G., Johnsen, J.A., Almas-Sorensen, L., Vitterso, J.: Personality and technology acceptance: The influence of personality factors on the core constructs of the technology acceptance model. In: Behaviour and Information Technology (2011). https://doi.org/10.1080/0144929X.2011.553740

[22] Tesch-Römer, C., Huxhold, O.: Nutzung des internets für die freiwillige tätigkeit. In: Simonson, J., Kelle, N., Kausmann, C., Tesch-Römer, C. (eds.) Freiwilliges Engagement in Deutschland – Der Deutsche Freiwilligensurvey 2019. Deutsches Zentrum für Altersfragen, Berlin, Germany (2021)

[23] Zhang, W., Gutierrez, O.: Information technology acceptance in the social services sector context: An exploration. In: Social Work, Volume 52, Number 3 (2007). https://doi.org/10.1093/sw/52.3.221

Chapter 2

New initiatives and laws in the digital market

Fostering Legal Compliance through Legal Design: a Digital Services Act case-study

Davide Audrito[1], Andrea Filippo Ferraris[2] and Emilio Sulis[3]

[1]Legal Studies Department, University of Bologna, Via Zamboni 27, Bologna, 40126, Italy
[2]Law Department, University of Turin, Lungo Dora Siena 100/A, Torino, 10154, Italy
[2]Computer Science Department, University of Turin, Corso Svizzera 185, Torino, 10149, Italy

Abstract

Legal Design represents a well-rooted research area at the intersection between law, design and computer-science, which has an increasingly relevant role in the clarification and accessibility of legal text sources. This preliminary research work illustrates that legal design fosters the ability of economic operators, including private citizens, public administrations and companies, to abide by complex legislative requirements arising from multi-level legal orders. To achieve this purpose, we rely on the Business Process Model and Notation (BPMN) standard to model a core provision of the so-called Digital Services Act (DSA).

Keywords

Legal Design, Legal Compliance, Digital Services Act, Business Process Modeling and Notation (BPMN)

1. Introduction

Legal design applies human-centered design principles to make legal systems and services more user-friendly [10]. It involves using methods, tools, and visual representations to improve the communication of legal information and interactions with the legal system.

Consistency between legal knowledge and graphical representations is crucial in legal design. Legal language, often complex and redundant [3], needs to be addressed to ensure that design models align with the principles of normativity [13] and explicability [8].

Normativity emphasizes that legal norms are established by legitimate authorities through a valid process. Laws target *personas*, individuals or entities, who can choose to comply with or violate the rules based on self-determination [11] [14]. Our research work focuses on modeling the provisions of the Digital Services Act (DSA) (Regulation (EU) 2022/2065) (see Section 3) using the BPMN standard. This helps PISs, acting as personas, understand and comply with the new obligations.

Legal design methods, including BPMN, facilitate the transmission of legal information while respecting the principle of normativity and avoiding confusion for non-experts.

CERC 2023: Collaborative European Research Conference, September 09–10, 2023, Barcelona, Spain
davide.audrito2@unibo.it (D. Audrito); andreafilippo.ferraris@unito.it (A. F. Ferraris); emilio.sulis@unito.it (E. Sulis)
https://www.unibo.it/sitoweb/davide.audrito2 (D. Audrito); http://www.di.unito.it/~sulis/ (E. Sulis)
0000-0002-9239-5358 (D. Audrito); 0009-0004-7487-5560 (A. F. Ferraris); 0000-0003-1746-3733 (E. Sulis)

ISBN 978-3-96187-021-9
https://doi.org/10.48444/h_docs-pub-518

2. Relevant Work

Legal informatics is engaged with Artificial Intelligence techniques [18], network analysis [16, 17], ontologies [1], and related research areas. The legal design methodology has been rising over the last decades, both on the side of theoretical studies [15, 6] and in the framework of applied works and projects [5, 7], so that nowadays it can be considered an autonomous discipline.

2.1. Legal Design

Legal design principles and visualization techniques offer businesses a valuable means of understanding and navigating complex legal obligations[19, 9]. An exemplar of the transformative potential of legal design is Candy Chang's collaboration with The Street Vendor Project and the Center for Urban Pedagogy, demonstrated through their guide "Vendor Power" [20]. This guide effectively translates intricate legal rules into clear and accessible diagrams while providing information in multiple languages, enabling street vendors to gain a comprehensive understanding of their rights and facilitating compliance with applicable regulations. Moreover, the guide's incorporation of personal stories and recommendations for policy reform underscores the significance of advocacy and community engagement in legal design endeavors[2].

2.2. Business Process Model and Notation (BPMN) for law

BPMN serves as an effective tool for modeling legal knowledge and processes[4], enabling graphical representation of business workflows and identification of improvement areas[12]. Legal professionals employ BPMN to visually depict legal processes, including contract review and approval. However, relying solely on BPMN may not sufficiently enhance the accessibility of legal compliance. To address this, the integration of legal design principles with BPMN offers visually intuitive representations of legal rules and requirements, thereby facilitating businesses' adherence to regulations.

3. Methodology

The Digital Services Act (DSA) establishes a single market for digital services by enshrining clear responsibilities and accountability for providers of intermediary services (PISs), including social media, e-commerce platforms, very large online platforms (VLOPs) and very large online search engines (VLOSEs).

In order to improve the understandability and clarity of the obligations enshrined in the DSA and addressed to PISs, we have adopted the following legal design methodological framework:

1. Legal analysis of the text source based on the theory of law and, where applicable, the philosophy of law and related disciplines.
2. Application of the Business Process Model and Notation (BPMN) standard for the formalization of legal knowledge. This method allows to improve the visualization and

explainability of legal processes, while ensuring the machine-readability of the output file, i.e. in XML format. The BPMN standard was implemented through the editor BPMN.iO[1].

3. Interpretation of the results, possibly measured empirically and with different methods, including indexes and questionnaires.

4. Results

In this section, we show the preliminary results of our work through the modelling of the normative contents enshrined in Article 14 DSA, which is well-suited to our purpose, given the occurrence of heterogeneous obligations.

4.1. Legal analysis of the text source

The Digital Services Act (DSA), adopted on November 16, 2022, represents a significant milestone in Information Society Services liability. It is a progressive update to the regulatory framework established by the European Union directive 31/2000, also known as the e-commerce directive. This self-executing regulation has direct and binding authority over EU individuals, companies, and member states in relation to their respective domains.

Article 14 of the DSA focuses on the requirements imposed on providers of intermediary services (PISs) when formulating the terms and conditions of their services. While the DSA contains various provisions, Article 14 stands out as it specifically addresses the criteria that apply to PISs in their contractual arrangements. Therefore, this article is a relevant benchmark for the application of legal design using the Business Process Model and Notation (BPMN) methodology, considering its wording and implications.

Indeed, Article 14 DSA entirely relies on the use of “shall”, which is the deontic operator for obligations. To show the complexity of the provision, the first paragraph thereof is quoted below:

> Terms and Conditions
>
> 1. Providers of intermediary services shall include information on any restrictions that they impose in relation to the use of their service in respect of information provided by the recipients of the service, in their terms and conditions. That information shall include information on any policies, procedures, measures and tools used for the purpose of content moderation, including algorithmic decision-making and human review, as well as the rules of procedure of their internal complaint handling system. It shall be set out in clear, plain, intelligible, user-friendly and unambiguous language, and shall be publicly available in an easily accessible and machine-readable format. [...]

The text contains lengthy sentences with consecutive direct objects, such as policies, procedures, measures, and tools. We focus on requirements that can be met with a "standard solution" for autonomous compliance. For instance, Article 14(1) of the DSA states that providers of

[1]https://demo.bpmn.io

intermediary services must include recipient-imposed restrictions in their terms and conditions. By incorporating the necessary information, PISs fulfill this requirement. In contrast, Article 14(4) of the DSA mandates diligent, objective, and proportionate behavior for PISs, which falls outside our scope.

4.2. Application of the BPMN standard

The use of the BPMN standard simplifies the legal text and lays the foundations for a questionnaire to be replied by PISs for testing-related purposes. Figure 1 portrays the BPMN representation of Article 14 DSA.

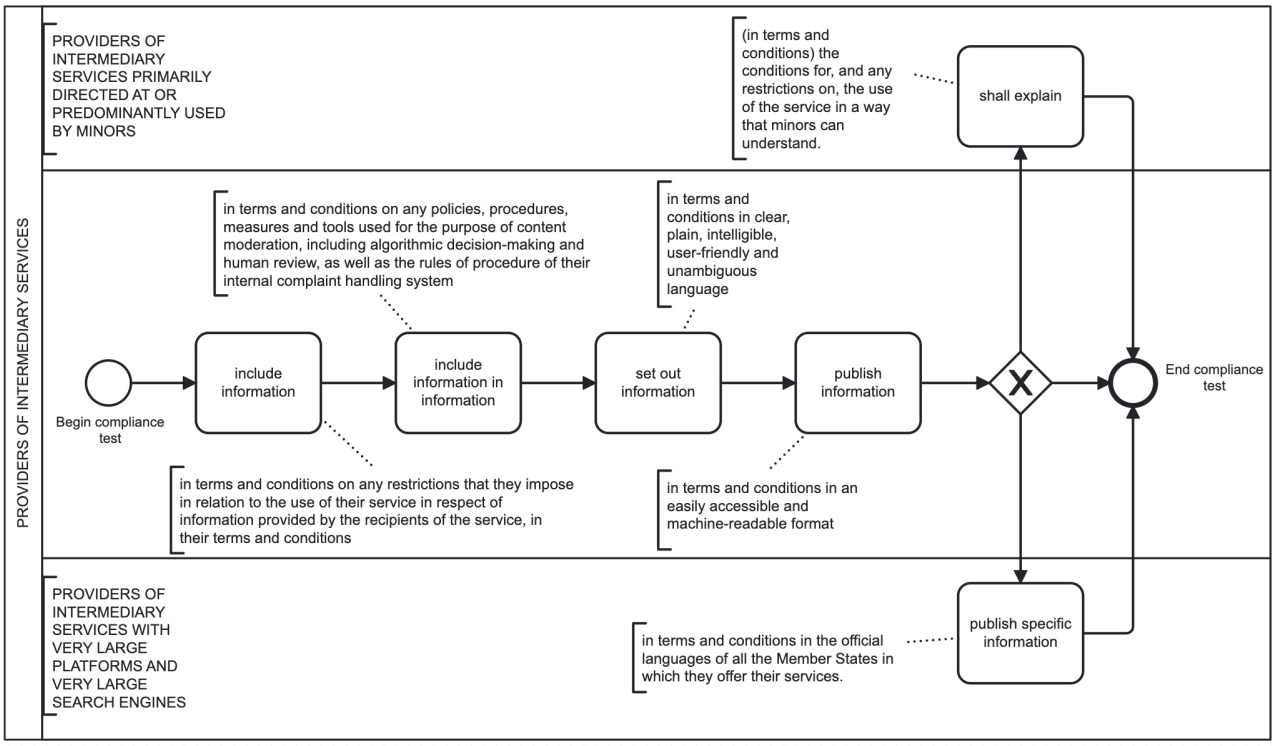


Figure 1: The BPMN representation of Article 14 DSA

The whole flowchart is contained in a “pool”, which represents participants in a process, i.e. providers of intermediary services. In the pool, there are two horizontal “lanes” above and below respectively, which serve to organize and categorize activities. We use lanes as sub-pools for activities of two categories of PISs that are affected by *ad hoc* obligations enshrined in Article 14(3) and (6) DSA. These categories are PISs primarily directed at or predominantly used by minors and PISs with very large platforms and very large search engines.

The process begins with a “start event”, i.e. circle on the left, under the label of “Begin compliance process” and ends with an “end event”, i.e. circle on the right side, under the label “End compliance test”. Starting from the start event, solid lines with solid arrowheads are used to show the “sequence flow” of activities performed in the process. In turn, “activities” are indicated by rounded-corner rectangles and labelled by following the formula “verb + direct object”. The forth activity from the left, i.e. publish information, is connected to an "exclusive

gateway" represented by a diamond with a cross in the middle. Indeed, depending on the nature of the PISs category addressed by the obligations, activities follow one of three alternative flows towards either “shall explain” or “publish specific information” or “end event” respectively.

For our purposes, each activity, i.e. rounded-corner rectangle, is connected to an “annotation”, which is one of the “artifacts” of the BPMN standard and provides additional text information for the reader of the BPMN diagram.

Figure 2 shows annotations of the activities “include information” and “include in information”. The former activity is complemented through the description “in terms and conditions on any restrictions that they impose in relation to the use of their service in respect of information provided by the recipients of the service”.

Annotations enrich activities with essential details needed to comply with the requirements and they enable the deconstruction and modelling of legal concepts. This function also allows to showcase questions for compliance-check purposes. For instance, the activity “set out information” could be integrated by the annotation “Did you set out in your terms and conditions information in clear, plain, intelligible, user-friendly and unambiguous language?”.

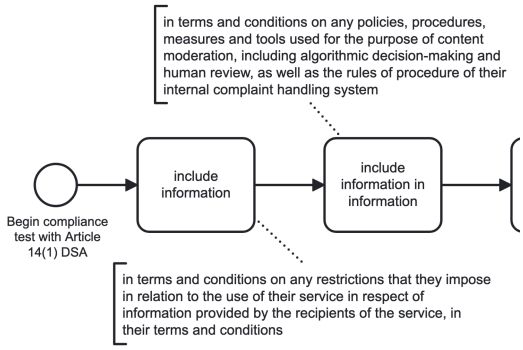


Figure 2: Examples of BPMN annotations contributing to model legal concepts

5. Conclusions and Future Work

This research work consists of a first-stage application of the Business Process Model and Notation (BPMN) standard as a legal design tool adopted for legal compliance purposes. In the future, we first intend to set up a method to test the present preliminary results, possibly through the use of indexes and a questionnaire. This will be possible with text reformulation and and graphic representations, in light of the best practices adopted in the legal design literature, including bullet points, icons, pictures, and symbols.

References

[1] Audrito, D., Sulis, E., Humphreys, L., Di Caro, L.: Analogical lightweight ontology of eu criminal procedural rights in judicial cooperation. Artificial Intelligence and Law pp. 1–24 (2022)

[2] Berger-Walliser, G., Barton, T.D., Haapio, H.: From visualization to legal design: a collaborative and creative process. Am. Bus. LJ **54**, 347 (2017)
[3] Butt, P.: Legalese versus plain language. Amicus Curiae **35**, 28 (2001)
[4] Capuzzimati, F., Violato, A., Baldoni, M., Boella, G., et al.: Business process management for legal domains: Supporting execution and management of preliminary injunctions. In: JURIX. pp. 149–152 (2015)
[5] Curtotti, Michael, H.H.P.S.: Legal design patterns for privacy. In: Schweighofer E. et al. (Eds.) Co-operation. Proceedings of the 18th International Legal Informatics Symposium IRIS 2015. p. 455–462 (2015)
[6] Ducato, R., Strowel, A.: Legal design perspectives. Theoretical and practical insights from the field (2021)
[7] Flood, M., Goodenough, O.: Contract as automaton: The computational representation of financial agreements. OFR Working Paper 15-04 (03 2015)
[8] Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., et al.: Ai4people—an ethical framework for a good ai society: opportunities, risks, principles, and recommendations. Minds and machines **28**, 689–707 (2018)
[9] Haapio, H., Hagan, M., Palmirani, M., Rossi, A.: Legal design patterns for privacy. In: Data Protection/LegalTech Proceedings of the 21st International Legal Informatics Symposium IRIS. pp. 445–450 (2018)
[10] Hagan, M.: Law By Design. (2017)
[11] Hart, H.L.A., Raz, J., Green, L.: The concept of law. oxford university press (2012)
[12] Kocbek Bule, M., Jost, G., Hericko, M., Polančič, G.: Business process model and notation: The current state of affairs. Computer Science and Information Systems **12**, 509–539 (06 2015)
[13] Palmirani, M.: A smart legal order for the digital era: A hybrid ai and dialogic model. Ragion pratica (2), 633–655 (2022)
[14] Rawls, J.: The justification of civil disobedience. Arguing about law pp. 244–253 (2013)
[15] Saha, D., Mandal, A., Pal, S.: User interface design issues for easy and efficient human computer interaction: An explanatory approach. International Journal of Computer Sciences and Engineering **3**, 127–135 (01 2015)
[16] Sartor, G., Santin, P., Audrito, D., Sulis, E., Caro, L.D.: Automated extraction and representation of citation network: A CJEU case-study. In: Guizzardi, R.S.S., Neumayr, B. (eds.) CMLS, EmpER, and JUSMOD 2022. LNCS, vol. 13650, pp. 102–111. Springer (2022)
[17] Sulis, E., Humphreys, L., Vernero, F., Amantea, I.A., Audrito, D., Caro, L.D.: Exploiting co-occurrence networks for classification of implicit inter-relationships in legal texts. Inf. Syst. **106**, 101821 (2022)
[18] Sulis, E., Humphreys, L.B., Audrito, D., Caro, L.D.: Exploiting textual similarity techniques in harmonization of laws. In: et al., S.B. (ed.) AIxIA 2021. LNCS, vol. 13196, pp. 185–197. Springer (2021)
[19] Waldman, A.E.: Privacy, notice, and design. STANFORD TECHNOLOGY LAW REVIEW **21**, 129ss (2018)
[20] Yankovskiy, R.: Legal design: New thinking and new challenges **2019**, 76 (05 2019)

Success Factors of a Digital Training Offer: Insights Based on the Online Course "Digital Expert in 5 Modules" for Specialists and Managers from SMEs

Marwin Bayer[1], Melanie Jakubowski[1] and Tanja Kranawetleitner[1]

[1]*Institute for Digital Transformation (IDT), University of Applied Sciences Neu-Ulm (HNU)*

Abstract

This paper aims to identify possible success factors of a digital training offer. To achieve this, an online course focusing on digital competencies for specialists and managers in small and medium-sized enterprises (SMEs) offered by Neu-Ulm University of Applied Sciences serves as an example. The feedback of participants are analyzed on the basis of two studies. The results show that preliminary success factors for the development and organization of such an online offer can be clustered within the four dimensions "acquisition", "content", "didactics" and "interaction". The article provides initial insights not only for further research but also for online training providers and SMEs who want to improve their offers or are looking for suitable solutions for their workforce.

Keywords

Online Course, Success Factors, Further Training, Digital Competencies

1. Introduction

The pandemic has brought to light various weaknesses of small and medium-sized enterprises (SMEs), including in relation to digitalisation (see [10]). For these companies, the main challenge is that their resources — compared to large corporations — are not sufficient to realize the opportunities of digital transformation and to initiate investments in a targeted manner (see [3]). In order to remain competitive in the field of digitalisation, companies must constantly review and pursue the further training of their employees (see [21]). In response to this challenge, the project "Digital Competencies@Bavarian-Swabia" provides hands-on training in current technologies and transformative methods for digital transformation these days. The program targets professionals and managers in SMEs. Its curriculum focuses on practical and application-oriented learning, with video lectures and online tasks available for remote and flexible learning. By conveying state of the art knowledge about digital transformation methods and new technologies, the online course has the aim to overcome existing digitalisation

CERC 2023: Collaborative European Research Conference, June 09–10, 2023, Barcelona, Spain
marwin.bayer@hnu.de (M. Bayer); melanie.jakubowski@hnu.de (M. Jakubowski); tanja.kranawetleitner@hnu.de (T. Kranawetleitner)
https://www.hnu.de/marwin-bayer (M. Bayer); https://www.hnu.de/melanie-jakubowski (M. Jakubowski); https://www.hnu.de/tanja-kranawetleitner (T. Kranawetleitner)
0009-0006-9357-9737 (M. Bayer); 0009-0005-4104-6640 (M. Jakubowski); 0000-0002-3696-4261 (T. Kranawetleitner)

ISBN 978-3-96187-021-9
https://doi.org/10.48444/h_docs-pub-518

deficiencies. Additionally, the program enables efficient remote work using virtual collaboration tools and online platforms. The following paper focuses on answering the question of which possible success factors based on the results and experiences of the online offer "Digital Expert in Five Modules" can be generally derived for a continuing education online course. In this context, success factors are "elements, determinants or conditions [...] that have a crucial influence on the success or failure of corporate activities" ([12], p. 176). In order to gain an insight into the presented digital continuing education offer, the online course is described in more detail with its thematic focal points as well as its design and structure. Beforehand, the opportunities and challenges of training programmes for accelerating the digital transformation in SMEs are explained in general terms. The results of two small studies conducted during the first year of implementation (March 2022 to March 2023) serve as the basis for an initial identification of possible success factors. The focus lies on the four success dimensions into which the success factors can be categorized based on the data collected from course participants and the experiences of course developers. The paper is completed by a critical discussion of the methods and results presented as well as points of departure for further research.

2. Training Programmes as Opportunity and Challenge to Accelerate Digital Transformation in SMEs

Digital transformation describes the "fundamental change of the entire corporate world through the establishment of new technologies based on the internet with fundamental effects on the entire society" ([19], p. 9). This change is a major challenge for companies. In recent years, the way businesses operate has been rapidly changing with the rise of new technologies. Those who fail to adapt may risk being left behind in the competitive landscape (see [20]). SMEs in particular should not miss the boat on digital transformation (see [24]). The acronym SME is a collective term for companies that do not exceed a defined size in terms of the number of employees, the annual turnover or the balance sheet in total. In this context, the definition of the European Commission for SMEs is applied: SMEs “include enterprises that employ fewer than 250 persons and have an annual turnover not exceeding EUR 50 million or whose annual balance sheet total does not exceed EUR 43 million" ([4], p. 10).

A key aspect of accelerating digital transformation in companies is the workforce. If employees have the essential skills and knowledge in the area of digitalisation, they can make a significant contribution to the success of the company (see [23]). These skills can be summarized under the term “digital literacy”: "Digital literacy encompasses the safe, critical and responsible use of and engagement with digital technologies for education, training, work and participation in society. It covers information and data literacy, communication and collaboration, media literacy, digital content creation (...), security (...), literacy, problem solving and critical thinking" ([5], p. 9). The development and expansion of employees’ skills in the field of digital literacy can be achieved, for example, through workshops and training programmes. While for a long time only classroom training was predominantly conducted, digital training is becoming increasingly popular (see [18]). A key driver for this is the Corona crisis, which has accelerated the digital transformation in adult and continuing education. Therefore the challenge of a digital culture is to meet expectations for media-supported offers. Raising awareness of the digitalisation process enables

the evaluation of different future scenarios (see [22]). Even though digitalisation is increasingly changing people's lives, access to and use of digital media depends on personal and structural characteristics such as level of education, geographical location or social status. Clear origin- and gender-specific differences in the use of digital media in the non-institutional sector as well as in the existing digital competencies could be identified. Without ensuring that all educated people have access to digital media and appropriate competence development according to their individual needs, these inequalities become educational and societal challenges (see [2]). Accordingly, companies will also have to take care of preparing their own workforce for the digital future. However, a study by KfW from 2020 shows that at least one third of SMEs lack digital skills (see [15]). One of the most significant challenges to accelerate the digital literacy of their employees is the lack of resources. Small enterprises in particular tend to have limited budgets and therefore often do not have the financial resources to invest in training programmes (see [24]). To overcome these challenges, they need to develop creative solutions that leverage technology and other resources. For example, online training platforms can be an effective way to deliver programmes to employees in a convenient manner. In addition, SMEs can collaborate with external institutions to provide training offers at reduced costs (see [6]).

3. Online Course "Digital Expert in 5 Modules"

In response to the negative effects of the COVID-19 pandemic, the European Social Fund has launched the REACT-EU[1] funding line. As part of funding measure 19: Vocational Qualification - Knowledge Transfer from Universities to Companies, the Institute for Digital Transformation at Neu-Ulm University of Applied Sciences developed and implements a training programme. The following presents the offer in detail, which serves as the basis for the subsequent analysis.

3.1. Target Group and Main Goals

Launched at the beginning of 2022, the project aims to create and implement an online course entitled "Digital Expert in Five Modules", comprising five modules on various digital transformation topics. The offer is primarily addressed to specialists and managers from SMEs, but it is open to all interested parties — such as digitalisation officers and employees from large companies. All participants will be provided with knowledge and skills related to digital technologies and transformation methods that will enable them to bring new perspectives to their organizations. Specifically, they will be empowered to launch and implement innovative digital initiatives in their organizations. The programme offers participants the opportunity to further develop their skills related to digital transformation, including the ability to design digital business models, recognize the potential of disruptive technologies, and foster an open culture of innovation.

3.2. Structure

The project's funding runs for two years (2022-2023). The entire five modules were activated step by step. On the one hand, the sequencing enables the subsequent modules to be adapted to the

[1] https://www.esf.bayern.de/esf-foerderung/react-eu

wishes of the participants if required; on the other hand, participants for whom closer support is important have the option of completing the programme according to plan. In addition, it is possible to start the training at any time and the order of the modules is also at the decision of the participants. The aim of this format is to create a time- and location-independent offer that can be tailored individually by participants to their respective needs. However, this flexibility also carries the risk that they will be overwhelmed by this freedom and that the lack of exchange with other course participants and instructors will have a negative effect on motivation (see [18]). To counteract this, the online course is accompanied by voluntary add-on offers that give participants the opportunity to get in touch with the organizers, instructors and other participants. Supplemental offers include online kickoffs for each of the five modules, as well as additional online sessions tailored to the content. In addition, a closed LinkedIn group serves as a platform for networking, while regular posts provide further information and encourage knowledge sharing. Participants learn about the potential and underlying concepts and are introduced to examples from SMEs. They also share experiences, network with each other and receive inspiration for practical implementation.

3.3. Content

Each module is divided into a theoretical part, which is prepared by professors of Neu-Ulm University of Applied Sciences. These chapters are complemented by practical experts from the respective field in the form of interviews or short articles. Each module is concluded with an online final test consisting of ten questions. After completing the course successfully, participants receive a certificate of participation. Furthermore, gamification incentives are provided through supplementary content. Depending on the performance in the final test, additional learning materials such as studies and articles on the corresponding topic area are unlocked. Figure 1 shows the modules with content focus and start date.

4. Methods and Results of the Research on the Online Course

Qualitative research aims to understand human behavior, while quantitative research analyzes and statistically evaluates the relationships and structures of different facts (see [14]). When a research question cannot be adequately answered by using only quantitative or only qualitative data, mixed methods research offers a compelling option. By mixing them, the strengths of both research methods can be leveraged and deeper and more meaningful findings can be obtained, allowing for the formation of more robust conclusions (see [11]). The goal of this mix of methods is to gain insights on multiple levels. This paper intends to identify initial possible success factors of an online course for continuing education. For this purpose, a mixture of qualitative and quantitative research methodology is used.

At the beginning, the composition of the course participants is presented. The description is based on the data of the course participants in terms of demographic data, company size, module choice and participation in the final quiz. The data was extracted on 27th of February 2023 and thus corresponds to a snapshot. It consists of merging data from the online platform data and the participant database. The first study includes the qualitative analysis of semi-structured interviews. The purpose is to gain deeper insights into the expectations, wishes and experiences

Module	Topics Covered	Start Date
1) Holistic Digitalization	*Digital transformation basics, Digital maturity level, Digital strategy, Implementation, Best practices, Roadmap with tools, Instruments for practical digital transformation*	*25/04/2022*
2) New Work and Digital Leadership	*Various tools for collaboration, Evaluation of tools regarding suitability for professional life*	*21/06/2022*
3) Digital Business Models	*Business model basics, Business model navigator, Identifying and implementing new digital business models*	*04/10/2022*
4) Disruptive Technologies	*Disruptive technologies (Mixed Reality, 3D Printing, Artificial Intelligence, Smart Data, Internet of Things), Examples of their application in business practice*	*29/11/2022*
5) Sustainability	*Technological measures for increasing energy efficiency and reducing CO2 emissions, Benefits of relevant technological measures primarily related to electromobility*	*07/02/2023*

Figure 1: Module Contents (own illustration)

of individual course participants. This study was conducted at the beginning of the online course, after completion of Module 1 from August to September of 2022. The second study presented in this paper consists of an evaluation to assess the course modules. The aim is to receive continuous feedback on the content, structure and didactic concept of the various modules. The surveys will be released to the course participants on the platform used after finishing the respective module. To ensure the highest possible quality, only fully completed questionnaires are evaluated. The evaluation of Module 5 is not yet available, as this part of the course has not yet been completed and the course itself is still active (until the end of 2023). The following presents the results of the different studies.

4.1. Sample Description of the Course Participants (Sample)

Of the 82 total course participants, just under one-third (30.5%, 25) were female and nearly two-thirds (69.5%, 57) were male. In terms of company size, just about half (48.8%, 40) of the participants came from small and medium-sized enterprises (SMEs), while the other half (51.2%, 42) belonged to large companies or other types of organizations such as universities. The majority of participants (91.5%, 75) were enrolled in Module 1, with slightly fewer participants found in Modules 2 through 5: 72 (87.8%), 73 (89.0%), 69 (84.1%), and 67 (81.7%), respectively. Looking at the participation rate of the final tests after each module, a similar picture emerges: While 25.3% (19) course participants took the quiz in Module 1, participation decreased steadily with the exception of Module 3 (21.9%, 16) (Module 2 final quiz 15.3%, 11; Module 4 final quiz 11.6%, 8).

4.2. Qualitative Evaluation of the Participant Interviews (Study 1)

The interviews were conducted with five participants of Module 1. As with the evaluations of the individual modules, a similar demographic distribution is evident. Four male participants and one female participant took part in the interviews. The interviews followed a semi-structured interview approach. This approach enables a uniform query of the interviewees on the given topic, but still leaves enough room for individual response options (see [9]). The interview questions were designed to investigate various dimensions related to digital competencies. They have a classic questionnaire structure. An initial question asked for the personal definition of digital competencies. The other questions inquired the participants' personal experience of acquiring digital skills and transferring them into professional practice. The transcribed interviews were evaluated using qualitative content analysis according to Mayring (see [17]). It was important for the study that the interviewees had the same level of knowledge and experience (see [9]). In this case, the criterion was the completion of Module 1, as the other modules had not yet been unlocked at that time. In the analysis of the responses, relevant statements were grouped into categories (see [17]). According to this consolidation into categories, such as "Definition of digital competencies", "Influencing factors on companies" or "Impact on knowledge transfer", the following key statements could be noted. The respondents described digital competencies as the ability to use technology devices and software tools, as well as having soft skills such as openness to digitalisation. The ways to acquire digital competencies include social media, online courses, training, seminars, consulting, and open labs. The majority of the respondents considered online courses with some level of trainer support and peer interaction helpful, but highlighted the need to avoid content overload. The high level of knowledge transfer was frequently mentioned by the interviewees, as the following quote from an interview illustrates: "[...] that was really totally practice-oriented and I really did this practice model project and we are now in the process of implementing it technically. So of course the development is still taking a bit [...] a trade fair, a big one, and then we actually want to offer the product.". The online course was also positively evaluated for its division into modules and the opportunity for practical application. SME-specific issues relating to the cost of training and the different needs of SMEs and companies were mentioned as a positive aspect, too. From this study, it can be concluded that the analyzed online course offers potentials in the transfer of knowledge of digital competencies, but that there are also certain challenges to be overcome for longer-term success.

4.3. Quantitative Evaluation of the Module Contents (Study 2)

The evaluation after each module contains a total of 14 identical questions. In addition, there is an opportunity at the end of the questionnaire to express wishes, praise and criticism in free text. The survey regarding Module 1 on the topic of holistic digitalisation (n=9) revealed that the most important sources for initial contact (How did you hear about the training programme?) of the training programme were online, in particular LinkedIn (44.4%, 4). Professional development was cited as the main reason for enrolling in a course (77.8%, 7). Participants expressed a desire for more practical application examples and flexibility in module times.

In contrast to Module 1, the main sources of initial contact in the evaluation of Module 2

"New Work and Digital Leadership" (n=5) were LinkedIn (40.0%, 2) and other channels (40.0%, 2). Participants mentioned professional development (80.0%, 4) and personal interest (60.0%, 3) as reasons for enrollment. They expressed a desire for more best practice examples and concrete information about funding opportunities and company contacts.

In the survey on Module 3 "Digital Business Models" (n=12), LinkedIn (25.0%, 3) and Word-of-Mouth (41.7%, 5) were named as the main sources for initial contact. Professional development (75.0%, 9) and personal interest (83.3%, 10) were the reasons for registration. The option for free text for e.g. suggestions for improvement was not used.

Initial awareness through Word-of-Mouth (80.0%, 4) was named by respondents in the evaluation to Module 4 on the topic of sustainability (n=5). Reasons for enrollment included professional development (100.0%, 5), personal interest (60.0%, 3), and the university's reputation as a professional provider (40.0%, 2). Respondents suggested offering more face-to-face webinars and less video material, on-site events, less technical content, and more opportunities for questions and feedback.

In summary, the module evaluations show that sources such as LinkedIn and Word-of-Mouth were the most common touchpoints for first awareness. Professional development and personal interest were the primary reasons for enrollment. Suggestions for improvement included, on the one hand, the need for more practical application examples, less technical content, flexibility in module times, and specific information about funding opportunities and company contacts. On the other hand, respondents wanted more in-person webinars, on-site events, and more opportunities for questions and feedback.

5. Success Factors of a Digital Further Training Offer

The previous explanations, in particular the results of the two studies and a sample description just mentioned, lead to the question referred to at the beginning: "Based on the results and experiences of the online offer "Digital Expert in Five Modules", which possible success factors can be derived in general for a continuing education online course?" To answer this question, the possible success factors can be divided into four success dimensions: Acquisition, content, didactics and interaction (see Fig. 2).

5.1. Acquisition

A crucial dimension is the acquisition of participants. Without them, no course can be conducted. Although, according to the project application, the training is aimed at the workforce of SMEs and is also primarily advertised as such, the SME criterion is not exclusive. The online course is open to all interested parties. Thus, employees from large companies, students, and self-employed persons could and can sign up for the course. The openness to all interested participants contributes to a higher reach, which in turn leads, among other things, to more SMEs being addressed. In general, the possibility of participation for all interested parties (see sample) contributes to a higher number in total. A paper concerning three studies on the accessibility of online courses also makes clear that easy access is one of the strengths of this format (see [8]). The reputation of the course provider was another important aspect in attracting course participants. According to the participants, the reputation of Neu-Ulm University of Applied

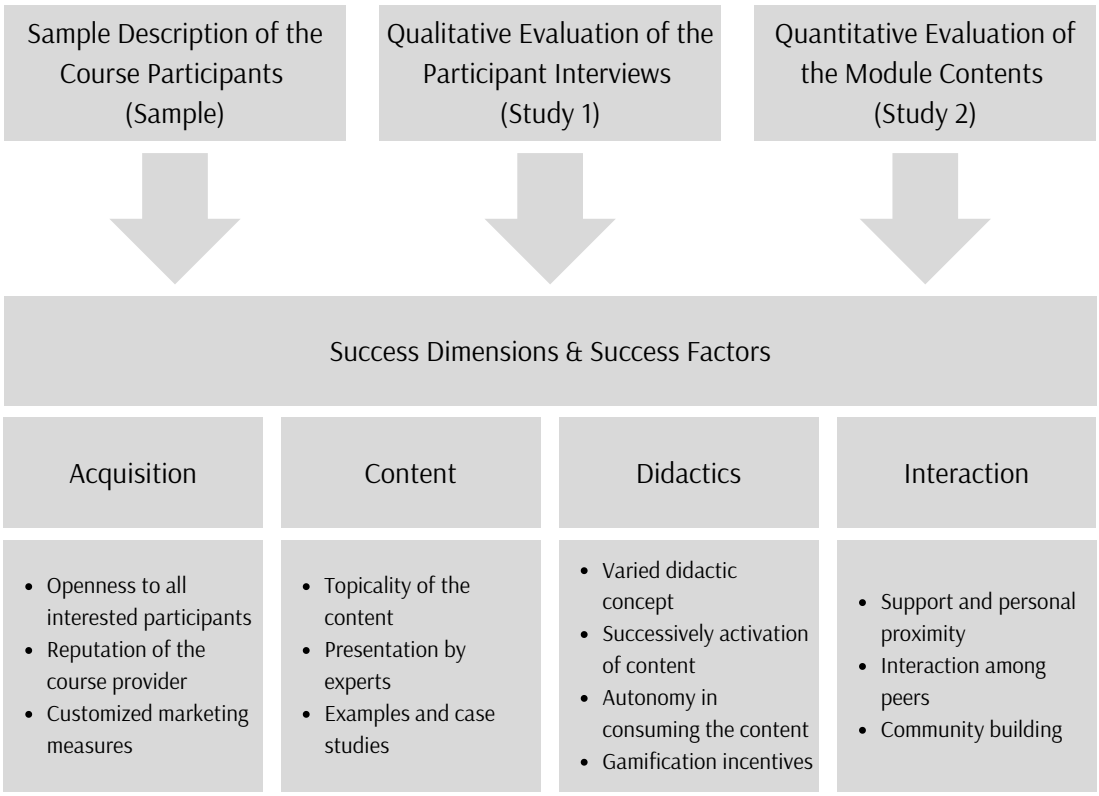


Figure 2: Success Dimensions and Respective Success Factors (own illustration)

Sciences combined with its quality as an educational institution were one of the main reasons for their enrollment in the course (see study 2). In addition, customized marketing measures for target groups are also essential to reach potential participants. In this case, marketing is primarily done through social media. The social media channels of Neu-Ulm University of Applied Sciences and the Institute for Digital Transformation serve as contact points for advertising the online course. In particular, the professional digital network LinkedIn was named as one of the main sources for initial contact with the online offer (see study 2). The range of the mentioned social media channels have played a key role in reaching many interested parties and later course participants.

5.2. Content

The course content represents another dimension of success. On the one hand, the topicality of the content is essential in this context (see [16]). Particularly in a dynamic field such as digital transformation, it is important to keep the content up to date in order to provide participants with a profitable training experience. On the other hand, it has proved beneficial for the content to be presented by experts in both theory and practice. In the online course "Digital Expert in Five Modules", professors from Neu-Ulm University of Applied Sciences lay the primarily theoretical framework for the respective modules. In order to make the theoretical content more accessible to practitioners, the course content is supplemented by examples and case studies from experts in business practice. This practical relevance with a concrete application of the

theory was positively emphasized by the participants: They make the theoretical constructs more tangible and applicable in their own working lives (see study 1). In some cases, the participants would like to see even more practical examples of application (see study 2). It is this mixture of theory transfer by professors and practical transfer by company practitioners with current insights that leads to a high relevance for the course participants. This last success factor can be summarized as the project team's conclusion up to now based on individual feedback from the course participants.

5.3. Didactics

In addition to the course content, the didactic aspect should not be ignored as a further dimension of success. Participants advocated for more face-to-face webinars and less video material, the organization of on-site events, and less technical content and more opportunities for questions and feedback (see study 2): It can therefore be concluded that a varied didactic concept is a relevant factor to activate participants. A mix of asynchronous videos and synchronous online sessions, for example, could be helpful, too. Furthermore, too much content leads to high frustration and quick saturation, which is confirmed by responses from course participants (see study 1). One way to tackle this is to successively activate content. This ensures that participants are not overwhelmed by complex topics or demotivated by a large number of issues. Another didactic aspect is the autonomy in consuming the content. The format of an online course makes it possible to work on the content independently of time and place. Participants positively highlight the format of an self-organized online course (see study 1). A comparable study which focused on students of a master-level distance learning program as a target group also emphasized the importance of self-organized learning (see [7]).

The high number of registrations from various companies can also be seen as confirmation of the flexible offer (see sample). Creating gamification incentives is another success factor for an online course. In the course described in this paper, for example, additional content can be unlocked if the participants score high or very high on the respective final test. This is meant to motivate the participants to deal intensively with the content and to successfully complete the test following the respective module. Feedback from one-on-one meetings and individual feedback from course participants to the project team confirm the benefits of gamification incentives. An acceptance survey among company employees also showed that there is acceptance for gamification elements in e-learning across all age groups (see [13]).

5.4. Interaction

The last success dimension that can be derived from the data collected and previous experiences is interaction. The main focus here is on support and personal proximity — on the one hand to the lecturers and on the other to the project team. Instructors and the project team should be available to participants for questions and individual support. Modules in which interaction with instructors was more strongly encouraged had a higher participation rate in the final quiz, which generally indicates higher engagement (see sample). In addition, personal support from an instructor was positively highlighted by participants (see study 1). Feedback during the voluntary online sessions as well as via email shows that participants perceive and appreciate a

timely response to questions, difficulties, and suggestions from the project team as beneficial. However, not only the contact with the trainers or the project team, but also the interaction among peers represents a success factor. A certain amount of interaction with other course participants creates a sense of community, which can promote motivation. Thus, interaction with other course participants was mentioned positively (see study 1). The importance of providing participants the opportunity to exchange with each other and to network is also pointed out in an article about courses at a distance learning university (see [1]). The aforementioned sense of community toward community building can be listed as another success factor. Outside of the online course environment, participants should also be given the opportunity to network with each other. A group in a professional network such as LinkedIn can be helpful for this, as was demonstrated in the online course "Digital Expert in Five Modules." Due to the target group in the work context, LinkedIn is a suitable choice, as it also makes professional networking beyond the training simple. The online course uses the network as a platform for providing up-to-date additional information on the modules, which participants can use to exchange experiences with one another.

6. Conclusion and Further Work

In summary, the answer to the question "Based on the results and experiences of the online offer "Digital Expert in Five Modules", which possible success factors can be derived in general for a continuing education online course?" is divided into four categories. As just described, openness to all interested parties, the reputation of the course provider and marketing measures belong to the dimension of participant acquisition. With regard to the content dimension, topicality, the presentation of experts from theory and practice, and examples as well as case studies are promising. However, content is also dependent on didactics. In this dimension, a variation of the didactic concept, step-by-step release and flexibility in the use of the content have been derived as possible success factors. Gamification incentives also fall under this success dimension. Support and proximity to the lecturers plus the project team, interaction with peers, and community building among the participants are preliminary success factors in the fourth dimension interaction.

These four dimensions of success are mainly based on two studies conducted during the project “Digital Competencies@Bavarian-Swabia”. It should be mentioned at this point that scientific support is not a requirement of the founding provider and is not a work package in the project plan. All data was collected in addition to the implementation and organization of the online course "Digital Expert in Five Modules". The interviews conducted (study 1) after Module 1 were taken from a study paper that dealt with the value creation of an online course offer for knowledge transfer in the field of digital competencies shortly after the start of the project. The evaluations after each module (study 2) primarily serve or have served as feedback from the participants for the conception of the further modules. In this way, the project team would like to respond flexibly to the wishes and suggestions for further development. Participation in the evaluations is voluntary and also dependent on when the participants complete the respective module. This can also explain the low response rate. Some success factors primarily relate to verbal feedback from participants in the voluntary online sessions as well as E-Mails. These

were neither empirically documented nor scientifically evaluated.

In general, this paper focuses on the results of the two studies and a sample description that are relevant to answering the question about possible success factors. Other data that do not play a role in answering the question were not mentioned in the results section. Furthermore, it is important to add that when the interviews were conducted (study 1), only Module 1 had been completed. For a more comprehensive overview, interviews should be repeated again after the completion of each further module. In addition, all possible success factors relate to the structure and organization of an online offer in general. SME-specific elements could not be identified. The target group orientation took place primarily in the creation of the content, where attention was paid to an SME focus. However, since the offer was and still is open to all interested parties, SME affiliation plays a subordinate role.

In order to understand the complex interactions between these factors and their influence on success in different contexts, further research is needed, especially with a larger number of participants. Due to the small number of participants, the listed dimensions of success only give an initial indication of possible success factors for online offers and must first be confirmed in a larger context. In addition, they must continue to prove themselves in real-life scenarios. The compilation does not claim to be exhaustive and needs to be adapted to different types of courses.

Acknowledgments

The project “Digitalkompetenzen@Bayerisch-Schwaben” (Digital Competencies@Bavarian-Swabia) is funded under the funding line REACT-EU in the funding action 19: “Vocational qualification — knowledge transfer from universities to enterprises” of the European Social Fund.

References

[1] Adamus, T., Marks, A., Sperl, A.: Distanz mit digitalen tools überbrücken — methoden aus der fernlehre. In: Pfannstiel, M., Steinhoff, P. (eds.) E-Learning im digitalen Zeitalter: Lösungen, Systeme, Anwendungen. Springer Fachmedien, Wiesbaden (2022). https://doi.org/10.1007/978-3-658-36113-6_6

[2] Autorengruppe Bildungsberichterstattung: Bildung in Deutschland 2020. Autorengruppe Bildungsberichterstattung (2020), https://www.bildungsbericht.de/de/bildungsberichte-seit-2006/bildungsbericht-2020/pdf-dateien-2020/bildungsbericht-2020-barrierefrei.pdf

[3] Bundesministerium für Wirtschaft und Energie (BMWi): Digitalisierung in Deutschland — Lehren aus der Corona-Krise. Gutachten des Wissenschaftlichen Beirats beim Bundesministerium für Wirtschaft und Energie (BMWi). Bundesministerium für Wirtschaft und Energie (BMWi) (2021), https://www.bmwk.de/Redaktion/DE/Publikationen/Ministerium/Veroeffentlichung-Wis-senschaftlicher-Beirat/gutachten-digitalisierung-in-deutschland.pdf

[4] Europäische Kommission, Generaldirektion Binnenmarkt, I.U.u.K.: Benutzerleitfaden zur Definition von KMU (2020), https://op.europa.eu/de/publication-detail/-/publication/79c0ce87-f4dc-11e6-8a35-01aa75ed71a1/language-de
[5] Europäischer Rat: Empfehlung des Rates vom 22. Mai 2018 zu Schlüsselkompetenzen für lebenslanges Lernen. Europäischer Rat (2018), https://eur-lex.europa.eu/legal-con-tent/DE/TXT/PDF/?uri=CELEX:32018H0604(01)&from=SV
[6] Franz, J., Robak, S.: Best-practice-beispiele für digitale weiterbildungsangebote. In: e. V., H.V. (ed.) Hessische Blätter für Volksbildung (HBV) (2020). https://doi.org/10.3278/HBV2003W008
[7] Halbritter, K.: Ein didaktischer ermöglichungsrahmen für projektbasiertes lernen im fernstudium. In: Hattula, C., Hilgers-Sekowsky, J., Schuster, G. (eds.) Praxisorientierte Hochschullehre: Insights in innovative sowie digitale Lehrkonzepte und Kooperationen mit der Wirtschaft. Springer Fachmedien, Wiesbaden (2021). https://doi.org/10.1007/978-3-658-32393-6_26
[8] Iniesto, F., McAndrew, P., Minocha, S., Coughlan, T.: Accessibility in moocs. In: Rienties, B., Hampel, R., Scanlon, E., Whitelock, D. (eds.) Open World Learning: Research, Innovation and the Challenges of High-Quality Education. Routledge, Milton Park (2022). https://doi.org/10.4324/9781003177098-11
[9] Kaiser, R.: Qualitative Experteninterviews. Konzeptionelle Grundlagen und praktische Durchführung. Springer VS, Wiesbaden (2021)
[10] Kay, R.: Mittelständische Unternehmen in der Covid-19-Pandemie — Betroffenheit von und Umgang mit der Krise (2022), https://www.ifm-bonn.org/fileadmin/data/redaktion/publikationen/ifm_materialien/dokumente/IfM-Materialien-295_2022.pdf
[11] Kelle, U.: Die Integration qualitativer und quantitativer Methoden in der empirischen Sozialforschung: theoretische Grundlagen und methodologische Konzepte. VS Verlag für Sozialwissenschaften (2008)
[12] Kreilkamp, E.: Strategisches Management und Marketing: Markt-und Wettbewerbsanalyse, Strategische Frühaufklärung, Portfolio-Management. Walter de Gruyter, Berlin (2010)
[13] Kretschmar, R., Herzog, A., Pfaff, D., Hedfeld, P.: Nur was für junge Gamer? Wie Gamification das E-Learning in Unternehmen auffrischen kann. IT-Daily, Published Online (2021), https://www.it-daily.net/it-management/digitalisierung/28564-wie-gamification-das-e-learning-in-unternehmen-auffrischen-kann
[14] Kuckartz, U.: Mixed Methods: Methodologie, Forschungsdesigns und Analyseverfahren. Springer VS (2014). https://doi.org/https://doi.org/10.1007/978-3-531-93267-5
[15] Leifels, A.: Mangel an Digitalkompetenzen bremst Digitalisierung des Mittelstands — Ausweg Weiterbildung? KfW Research (2020), https://www.kfw.de/PDF/Download-Center/Konzernthemen/Research/PDF-Dokumente-Fokus-Volkswirtschaft/Fokus-2020/Fokus-Nr.-277-Februar-2020-Digitalkompetenzen.pdf
[16] Lüpkes, N., Rommel, I.: Unternehmensspezifische digitalisierungsbedarfe und institutionelle bildungsentscheidungen — exploration von begründungsszenarien für kleine und mittlere unternehmen (kmu). In: Robak, S., Kühn, C., Heidemann, L., Asche, E. (eds.) Digitalisierung und Weiterbildung: Beiträge zu erwachsenenpädagogischen Forschungs- und Entwicklungsfeldern. Barbara Budrich, Opladen (2022)
[17] Mayring, P.: Qualitative Inhaltsanalyse. Grundlagen und Techniken. Beltz, Weinheim

(2015)
[18] Nürnberg, V.: Digital Learning Experience. Betriebliche Weiterbildung durch Blended Learning zukunftsfähig gestalten. Haufe Group, Freiburg (2021)
[19] PwC: Digitale Transformation — der größte Wandel seit der industriellen Revolution. PwC, Frankfurt (2013)
[20] Schallmo, D., Lang, K., Hasler, D., Ehmig-Klassen, K., Williams, C.: An approach for a digital maturity model for smes based on their requirements. In: Schallmo, D., Tidd, J. (eds.) Digitalization: Approaches, Case Studies, and Tools for Strategy, Transformation and Implementation. Springer International Publishing, Cham (2021)
[21] Schnelle, J., Schöpper, H., Kersten, W.: Corona: Katalysator für Digitalisierung und Transparenz? (2021). https://doi.org/10.30844/I40M_21-1_S27-31
[22] Schäfer, E.: Die digitale Transformation in der Erwachsenen- und Weiterbildung als Herausforderung für die Organisationsentwicklung: Gelingensbedingungen und Gestaltungsoptionen. MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung (2022). https://doi.org/https://doi.org/10.21240/mpaed/50/2022.12.12.X
[23] Wintermann, O.: Perspektivische Auswirkungen der Corona-Pandemie auf die Wirtschaft und die Art des Arbeitens (2020). https://doi.org/10.1007/s10273-020-2733-0
[24] Zimmermann, V.: KfW-Digitalisierungsbericht Mittelstand 2018. Digitalisierung erfasst breite Teile des Mittelstands — Digitalisierungsausgaben bleiben niedrig (2019), https://www.kfw.de/PDF/Download-Center/Konzernthemen/Research/PDF-Dokumente-Digitalisierungsbericht-Mittelstand/KfW-Digitalisierungsbericht-2018.pdf

Reference Model of Virtual Shopping Personal Assistant

Olga Cherednichenko and Ľuboš Cibák

Bratislava University of Economics and Management, Bratislava, Slovak Republic

Abstract

In modern times, e-commerce has become an integral part of our daily lives. However, with the abundance of e-commerce websites, customers often face difficulty in finding the right product quickly and easily. This leads to spending significant amounts of time comparing descriptions and photos across different platforms and verifying whether items are identical or similar before making a purchase decision. To overcome this challenge, the use of shopbots has become increasingly common in e-commerce. These virtual shopping assistants can perform various functions beyond just comparing prices, such as comparing product features, user reviews, delivery options, and warranty information. The central theme of this paper is the potential advantages of conversational chatbots in enhancing the online shopping experience and offering novel opportunities for e-commerce. We introduce a reference model for a virtual shopping personal assistant that incorporates four crucial elements: a conversational unit, a task identification, data searching and exploration component, and a recommendation model. Our approach enables natural conversation to extract user preferences and obstacles, compare products, group similar items, and provide personalized recommendations. This all-inclusive approach forms the foundation of our proposed reference model.

Keywords

e-commerce, chatbot, recommender system, item matching

1. Introduction

E-commerce has become a vital part of modern life. The market is flooded with numerous e-commerce stores that offer a wide variety of products across different categories and brands, ranging from food, electronics, shoes, clothing, and more, sourced from various manufacturers. However, the practice of multiple e-shops selling the same real-world products has had an impact on both customers and sellers. Customers now spend a significant amount of time searching for the right product, comparing descriptions and photos across different platforms, and deciding if items are identical or similar before making a purchase decision. Similarly, sellers devote a considerable amount of time analyzing the e-commerce market, evaluating demand and supply, and examining the price policies of competitors to remain competitive and maintain their market share.

Researchers have observed that the proliferation of websites and information on the web is making it difficult for users to quickly and easily find the information they need. To address this, the use of shopbots or robots to assist shoppers has become increasingly common on e-commerce sites and general-purpose web portals. Over time, shopbot technology has advanced significantly, with current shopbots able to perform a range of functions beyond just comparing prices. For example, they can compare product features, user reviews, delivery options, and warranty information.

A chatbot is a computer program that mimics human conversation through text or voice interactions with users [1]. Chatbots can provide several advantages for e-commerce businesses, including:

CERC 2023: Collaborative European Research Conference, June 09–10, 2023, Barcelona, Spain
olga.cherednichenko@ vsemba.sk (Olga Cherednichenko); lubos.cibak@vsemba.sk (Ľuboš Cibák)
0000-0002-9391-5220 (Olga Cherednichenko; 0000-0003-3881-7924 (Ľuboš Cibák)

ISBN 978-3-96187-021-9 https://doi.org/10.48444/h_docs-pub-518

1. 24/7 availability: Chatbots can operate round the clock, providing customers with access to support and information outside regular business hours. This ensures that customers can get their questions answered and issues resolved even if they reach out at odd hours.
2. Faster response times: Chatbots can instantly provide answers to common queries, freeing up support staff to tackle more complex issues. This can result in faster response times and reduced wait times for customers, which can improve their overall experience with the brand.
3. Personalization: Chatbots can use data about the customer's previous purchases, browsing history, and preferences to offer personalized product recommendations and promotions. This can help businesses to build a stronger connection with customers and increase sales.
4. Scalability: Chatbots can handle multiple conversations simultaneously, allowing businesses to handle a large volume of customer queries and support requests without hiring additional staff. This can help businesses to scale their customer support operations more efficiently.
5. Cost-effectiveness: Chatbots can provide cost-effective support and reduce the need for businesses to hire additional support staff. This can help businesses to save money while providing excellent customer service.

The use of chatbots can help e-commerce businesses to provide better customer support, increase sales, and reduce costs. By automating common tasks, chatbots can help businesses to focus on more critical aspects of their operations, such as product development and marketing.

Chatbots are typically designed for specific tasks or providing information to users. They operate under pre-defined rules, following scripts or decision trees to respond to user inputs [1, 2]. Some chatbots enable users to perform open-ended data analysis tasks by combining commands through nested conversations. Several platforms offer advanced capabilities in this regard. IBM Watson Assistant [3] allows users to create complex conversational flows and merge various commands to achieve data science tasks. The Dialogflow platform [4] features a natural language understanding system that enables users to build conversation flows and nested dialogues for handling intricate data science tasks. Rasa [5] is an open-source conversational AI framework that empowers developers to build AI assistants with advanced NLP capabilities and flexible dialogue management. The Botpress platform [6] provides a workflow builder, which allows users to design chatbots with nested conversations, custom scripts, and integrations with third-party services. The Wit.ai platform [7] provides an NLP system that enables developers to create intelligent chatbots, voice assistants, and other conversational AI applications with complex dialogue flows. These are just a few examples of chatbots that enable users to perform open-ended data exploration tasks by combining commands through nested conversations.

This research focuses on the potential benefits of conversational chatbots in improving the online shopping experience and presenting new opportunities for e-commerce. We propose a reference model for a virtual shopping personal assistant that comprises four essential components: a conversational unit, a task identification, data searching and exploration component, and a recommendation model. Our approach allows for a human-like dialogue, which can extract user preferences and obstacles, compare items, group similar items, and create personalized recommendations. This comprehensive approach serves as the basis for the proposed reference model.

The paper is structured as follows: the next section explains the concept of a shopping personal assistant and outlines why chatbots are ideal for this purpose. The research questions are then introduced, followed by a review of the current state-of-the-art. The fourth section provides a brief overview of the methods employed, while the fifth section presents the virtual assistant reference model. A discussion section is included to provide additional insights, and we conclude by summarizing our results.

2. The concept of a shopping personal assistant

During online shopping, the process of looking for items on the web typically involves using a search engine or visiting an e-commerce website and browsing through various product categories or using filters to narrow down the options. The user may input a specific search query or use general keywords related to the item they are looking for. The search results may display various items with their prices, descriptions, and images. The user can then click on the item to view more details or add it to their cart

for purchase. Additionally, the use of shopbots or virtual shopping assistants can help automate the search process and provide personalized recommendations based on user preferences and previous search history.

A shopping personal assistant is a type of virtual assistant that helps consumers with their shopping tasks. It is a computer program or system that uses natural language interactions to engage with users and provide them with relevant information about products, services, prices, and more. The shopping personal assistant can assist with a variety of tasks, such as searching for products, comparing prices, checking availability, making recommendations, and processing transactions. It can also learn from the user's preferences and behavior to provide personalized assistance and improve the overall shopping experience.

Some shopping personal assistant or shopbots can even make purchases on behalf of the user, with the user's consent, of course. Shopbots can also be integrated with voice assistants, such as Amazon's Alexa or Google Assistant, to allow for voice commands and hands-free shopping.

A search assistant is any tool that helps users find information more easily and efficiently on the internet. This can include search engines like Google or Bing, which use algorithms to crawl and index web pages and provide relevant results for user queries. It can also include browser extensions or plugins that help users search for specific types of information, such as images or videos, or that provide additional information about search results, such as ratings or reviews.

Figure 1 illustrates the general process of online product search and purchase, which can be classified into three distinct scenarios based on the level of complexity for the buyer and the degree of uncertainty in decision-making. The first scenario involves a basic search where the buyer already knows the specific product and the e-commerce site from which to make the purchase. In the second scenario, the task expands to include not only the product search but also the selection of the trading platform or offer. The third scenario is more intricate, where the buyer is uncertain about the exact product to purchase, for example, when searching for a laptop without a specific model in mind. This scenario requires the buyer to compare various products based on their features, read reviews, and explore different offers, which demands significant time investment and particular skills.

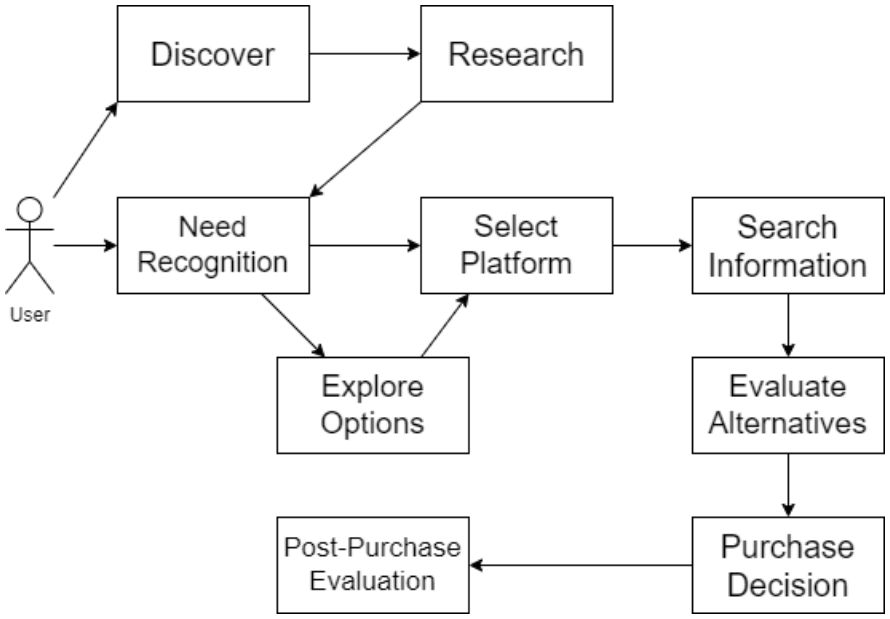


Figure 1: The general process of online product search and purchase

To address this challenge, we propose the concept of a virtual personal assistant. The primary purpose of this assistant is to provide support to buyers in all scenarios, by offering assistance with searching, comparing, filtering, matching, and ranking of products. Additionally, the personal shopping assistant can suggest relevant items and facilitate their purchase.

Thus, the following research questions are raised.

RQ1. Can a virtual personal assistant using a conversational interface help simplify the buying process for novice buyers?

RQ2. What features of a virtual shopping personal assistant are most helpful in assisting buyers who have limited knowledge of the product they want to purchase?

RQ3. How can virtual shopping personal assistants be designed to better support the decision-making process of novice buyers?

3. The state-of-the-art

E-commerce represents a significant revenue stream in the digital economy, and as such, there is a particular interest in implementing chatbots to enhance user adoption of online shopping, thereby building trust in the process as a routine part of daily life. To investigate the role of chatbots and their implementation in e-commerce, we relied on a systematic literature review documented in [8]. This review aimed to identify the current state of the art regarding chatbot implementation and adoption in the e-commerce domain.

Chatbots can be classified into different categories based on their functionalities and the type of collaboration they facilitate [2, 9, 11]. They can provide insights, recommendations, or predictions based on the available data. Chatbots can also send notifications and alerts to users triggered by predefined actions, such as changes in data or anomalies in key metrics.

Although a chatbot is a type of conversational agent (CA), not all CAs are chatbots. CA is a broader term that includes any computer program or system that can engage in natural language interactions with users [10, 12]. CAs can be rule-based or use machine learning (ML) and natural language processing (NLP) techniques to comprehend and respond to user inputs.

Examples of CAs include Apple's Siri [13], Amazon's Alexa [14], and Google Assistant [15]. These virtual assistants use NLP to help users perform tasks, answer questions, and execute actions. Additionally, various chatbot platforms are available, such as Dialogflow [16] and Microsoft Bot Framework [17], which enable developers to create their own CAs.

Drawing upon a review of 233,085 papers, the authors of [8] observed that despite the widespread interest in chatbot integration in e-commerce, only 81 papers met the evaluation criteria for inclusion, such as a relevant abstract, clear methodology presentation, full-text availability, relevance, and use of English. The findings from [8] identified four areas where chatbots are prevalent, with e-commerce being the most predominant at 41%, followed by bank management at 28%. The research indicates that "chatbot" and "artificial intelligence" are the two keywords with the highest co-occurrence in the selected papers. The use of the Python programming language is prevalent in developing chatbots for e-commerce [8]. Consequently, we can conclude that while the topic is not novel, it is still cutting-edge, with numerous successful chatbot and conversational agent implementations demonstrating their potential. Various tools and language models are available to implement the personal shopping assistant. This paper focuses on a domain-specific framework that outlines the component parts of a virtual shopping personal assistant.

4. Methods and materials

Conversational agents (CAs) are computer programs designed to engage in natural conversations with human users. They can be categorized as either chatbots for informal chatting or task-oriented agents for providing users with specific information related to a task [2, 9]. CAs may use text-based input and output or more complex modalities such as speech. The handling of dialogue is a crucial component of any CA, which can range from simple template-matching systems to complex deep neural networks (DNNs).

While chatbots and task-oriented agents are the primary categories for classifying CAs, other aspects such as input and output modalities, applicability, and the agent's ability to self-learn could also be considered [18]. Chatbots can be classified into three main groups based on their response generation: pattern-based, information-retrieval, and generative. Task-oriented agents must provide precise and consistent answers to fact-based conversations on specific topics. The pipeline model, shown in Fig. 2, depicts the various stages of processing as separate components. The three basic components are natural language understanding (NLU), cognitive processing, and natural language generation (NLG). NLU involves identifying precisely what the user said or meant, while cognitive processing executes the

dialogue policy to construct the knowledge required for the response using a knowledge base. NLG involves formulating the answer as a sentence.

Overall, CAs have numerous applications and offer a wide range of possibilities for dialogue representation and handling. While chatbots and task-oriented agents are the primary classification categories, other taxonomical aspects could also be considered when classifying CAs. The pipeline model provides a useful framework for understanding the various components involved in CA dialogue handling.

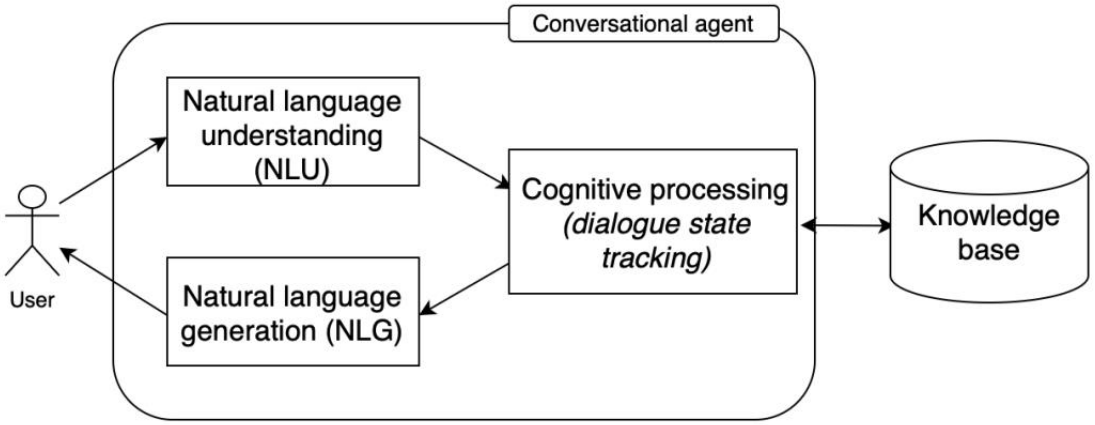


Figure 2: The basic pipeline for conversational agent (adopted from [19])

Reference models serve various purposes, including facilitating the creation of models for objects and their interrelationships. By breaking down complex problem spaces into basic concepts, reference models can be used to compare two different solutions to a problem, allowing for a detailed discussion of the various component parts of each solution in relation to one another. Thus, in this research we try to find out what the most significant components are and how they interact each other in order to support personal needs of online buyers during the complex scenario of online shopping.

5. Reference model development

Based on the findings of the literature review presented in [8], there are four main areas in which Chatbots are most commonly utilized, with e-Commerce being the most predominant. This is likely due to the vast number of users who visit e-Commerce websites daily. As customers often require quick and efficient solutions to their queries in order to facilitate their purchasing decisions, Chatbots have become an increasingly popular tool to meet these needs. Through the use of natural language chat interfaces, users are able to receive rapid and personalized assistance, minimizing the time required to identify and address their concerns. By doing so, e-Commerce companies can better cater to the specific needs and preferences of their customers, thereby streamlining the purchasing process and enhancing overall customer satisfaction.

To answering the first research question we can emphasize that virtual personal assistant using a conversational interface has the potential to simplify the buying process for novice buyers. By providing a natural language interface, virtual assistants can assist users in discovering, selecting and purchasing items in an intuitive and conversational manner. This can help novice buyers navigate the often-complex buying process, providing them with guidance and support as needed. Additionally, virtual personal assistants can learn from user interactions, becoming better at predicting and meeting user needs over time. The use of conversational interfaces in e-commerce has already been shown to improve customer satisfaction and loyalty, suggesting that virtual assistants could be an effective tool for simplifying the buying process.

A virtual shopping personal assistant can be particularly useful for buyers with limited product knowledge by providing guidance and support throughout the buying process. Some of the most helpful features of a virtual shopping personal assistant in this regard might include:

- Product recommendations: The assistant can suggest products based on the buyer's preferences and needs, helping them to make informed decisions.

- Education: The assistant can provide information about product features, specifications, and benefits, helping the buyer to understand the product they are considering.
- Question answering: The assistant can answer any questions the buyer may have about the product, clarifying any uncertainties or confusion they may have.
- Comparison: The assistant can provide side-by-side comparisons of products, helping the buyer to choose the best option for their needs.
- Personalization: The assistant can take into account the buyer's previous purchase history, preferences, and behavior to offer tailored recommendations.

Eventually, a virtual shopping personal assistant can offer a range of features to assist buyers, simplifying the buying process and helping to ensure that they make informed purchasing decisions.

Based on the scenarios presented in Figure 1, there are several important features that must be implemented in virtual shopping assistant software in order to assist novice buyers effectively. These features include:

- The ability to perform various e-commerce website commands such as filtering, querying, selecting, and setting parameters.
- An information retrieval module is necessary to find relevant information, explore options, and research user needs.
- Item matching is necessary to compare different offers and proposals.
- A searching algorithm and search engine are required to search the internet for information related to the user's needs.
- Personalization based on user preferences, history, and behavior, which can enhance the relevance and effectiveness of the recommendations provided by the assistant.
- Machine learning algorithms to continuously learn from user interactions and improve the assistant's ability to provide personalized recommendations.
- Language understanding and text generation are both necessary in order to effectively communicate with the user.

In addition to the features mentioned above, there are several other important features that should be implemented in a virtual shopping assistant software, such as:

- Natural language processing (NLP) to understand and interpret user queries and commands in a conversational manner.
- Integration with multiple e-commerce platforms and marketplaces to provide a wider range of options and availability to users.
- The ability to seamlessly integrate with payment gateways can help streamline the checkout process and provide a more convenient shopping experience for users.
- Secure payment processing and data protection features to ensure the safety of user information and transactions.
- The ability to search for products based on images or pictures can be a valuable feature for users who may have difficulty describing the product they are looking for in words.
- A virtual shopping assistant should be available to users at any time of the day or night, in order to provide timely assistance and support.

Overall, the implementation of these features can greatly enhance the functionality and usability of a virtual shopping assistant software, making it a valuable tool for both novice and experienced buyers.

In order to generalize the discussed features of a virtual personal shopping assistant, we propose the reference model depicted in Figure 3. Fig. 3 shows the data flows during the interaction between the customer and the software system of the virtual assistant. The entities defined in the framework reflect the main functional tasks, the solution of which is possible independently. We assume that for the implementation of a virtual assistant, it is possible to use various approaches that have proven themselves in a certain area, for example, for generating a dialogue in natural language, for information retrieval, and for generating recommendations. The proposed reference model allows us to evaluate the possibilities of integration or the prospects for breaking down new software solutions for each of the selected components in the context of achieving the overall goal of creating a virtual assistant. The proposed reference model outlines four essential components of a virtual shopping personal assistant:

- Natural language conversation processing: This component is responsible for processing the dialogue between the user and the virtual assistant. The virtual assistant must accurately identify the

user's intentions and execute the appropriate dialogue policy to construct a response. It then generates a response in natural language that is easily understandable by the user.

- Task identification: This component involves identifying the services that best match the user's needs based on natural language processing and semantic analysis of the input text. It also extracts relevant keywords for information retrieval, behavior annotation, and matching.
- Data searching and exploration: This component detects potential items related to the user's needs, extracts and compares them, groups them, and evaluates them. It comprises a set of services that interact with each other and use knowledge bases or the web to meet the user's needs.
- Recommender model: This component is responsible for personalizing the shopping experience based on the user's preferences, shopping history, and behavior. It employs machine learning and artificial intelligence algorithms to learn from data exploration and user interactions.

Thus, the reference model provides a comprehensive framework for developing a virtual shopping personal assistant software by identifying the main modules necessary for facilitating the software creation.

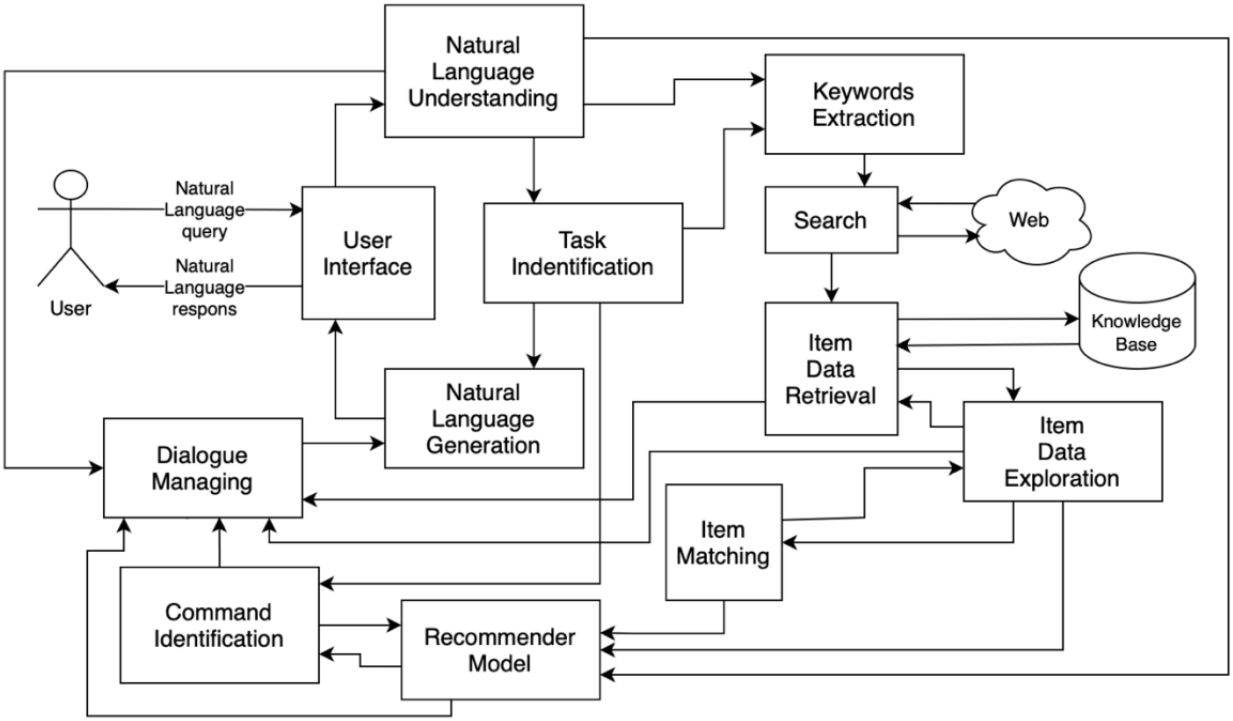


Figure 3: The reference model of virtual shopping personal assistant

6. Discussion and Conclusion

The increasing popularity of e-commerce has revolutionized the business world and is expected to continue doing so for years to come. As a part of this trend, many companies are adopting conversational agents (CAs) in their sales and customer service. Research in this area often focuses on enhancing customer satisfaction. For instance, a study conducted by [20] examined customer service CAs for luxury brands in the context of customer satisfaction. Similarly, the authors of [21] explored how sentiment analysis could be used to evaluate the customer experience and found that automated sentiment analysis can serve as a substitute for direct customer feedback. One key takeaway from research in this area is that customers typically prefer systems that offer quick and efficient solutions to their problems. Another important design principle is to use a tiered approach.

E-commerce has become an essential part of our daily lives, but the vast number of e-commerce websites makes it challenging for customers to find the right product quickly and easily. This results in spending a significant amount of time comparing descriptions and photos across various platforms and verifying whether items are identical or similar before making a purchase decision. To overcome this challenge, the virtual shopping assistants have become increasingly popular in e-commerce. They can

perform a range of functions beyond just comparing prices, including comparing product features, user reviews, delivery options, and warranty information. Our proposed reference model for a virtual shopping personal assistant consists of four key components: a conversational unit, a task identification module, a data searching and exploration component, and a recommendation model. The conversational unit enables natural language dialogue with the user, allowing the assistant to extract preferences and identify obstacles. The task identification module determines the user's needs and selects the appropriate service to fulfill those needs. The data searching and exploration component finds and compares relevant items and groups similar ones together, while the recommendation model generates personalized recommendations based on the user's shopping history and behavior. With this approach, our virtual shopping personal assistant can simulate a human-like conversation and provide an enhanced shopping experience for the user.

Reference models have multiple applications, such as simplifying the creation of object and relationship models for engineers and developers. Additionally, they can be used to establish clear roles and responsibilities, which can result in high-quality outcomes. By deconstructing complex problem spaces into fundamental concepts, reference models enable a comparative analysis of two distinct problem-solving approaches. This approach facilitates a detailed discussion of the individual components of each solution, with regard to their relationships to one another. The suggested reference model outlined in this study defines the software components responsible for assisting buyers in discovering, selecting, and purchasing items. It may serve as a basis for evaluating the suitability of each candidate solution in fulfilling user needs.

7. Acknowledgements

The research study depicted in this paper is funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V01-00078.

8. References

[1] D. Avalverde, A Brief History of Chatbots. Perception, Control, Cognition, 2019. URL: https://pcc.cs.byu.edu/2018/03/26/a-brief-history-of-chatbots/.
[2] J. Masche, N.-T. Le, A review of technologies for conversational systems, in: International Conference on Computer Science, Applied Mathematics and Applications, Springer, 2017, pp. 212–225. doi:10.1007/978-3-319-61911-8_19.
[3] Watson Assistant. URL: https://www.ibm.com/products/watson-assistant.
[4] Dialogflow. URL: https://cloud.google.com/dialogflow.
[5] Conversational AI Platform. URL: https://rasa.com.
[6] Botpress. URL: https://botpress.com.
[7] Wit. URL: https://wit.ai.
[8] J. Gamboa-Cruzado et al., Use of chatbots in e-commerce: a comprehensive systematic review, Journal of Theoretical and Applied Information Technology 101(4) (2023) 1172–1183. URL: http://www.jatit.org/volumes/Vol101No4/3Vol101No4.pdf.
[9] N. Svenningsson, M. Faraon, Artificial Intelligence in Conversational Agents: A Study of Factors Related to Perceived Humanness in Chatbots, In: Proceedings of the 2019 2nd Artificial Intelligence and Cloud Computing Conference (AICCC 2019), Association for Computing Machinery, New York, NY, USA, 2020, pp. 151–161. doi:10.1145/3375959.3375973.
[10] M. Akhtar, J. Neidhardt, H. Werthner, The Potential of Chatbots: Analysis of Chatbot Conversations, In: 2019 IEEE 21st Conference on Business Informatics (CBI), IEEE, 2019, pp. 397–404. doi:10.1109/CBI.2019.00052.
[11] G. Caldarini, S. Jaf, K. McGarry, A Literature Survey of Recent Advances in Chatbots, Information 13(1) (2022) 41. doi:10.3390/info13010041.
[12] R. Bavaresco, D. Silveira, E. Reis, J. Barbosa, R. Righi, C. Costa, R. Antunes, M. Gomes, C. Gatti, M. Vanzin, S. C. Junior, E. Silva, C. Moreira, Conversational agents in business: A systematic literature review and future research directions, Computer Science Review 36 (2020) 100239. doi:10.1016/j.cosrev.2020.100239.

[13] Siri – Apple. URL: https://www.apple.com/siri/.
[14] Amazon Alexa Voice AI. URL: https://developer.amazon.com/en-US/alexa.
[15] Google Assistant. URL: https://assistant.google.com/.
[16] Dialogflow. URL: https://cloud.google.com/dialogflow.
[17] Microsoft Bot Framework. URL: https://dev.botframework.com/.
[18] S. Diederich, A. B. Brendel, L. M. Kolbe, Towards a taxonomy of platforms for conversational agent design, In: 14th International Conference on Wirtschaftsinformatik (WI2019), 2019, pp. 1100–1114. URL: https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1247&context=wi2019
[19] M. Wahde, M. Virgolin, Conversational Agents: Theory and Applications, 2022. URL: https://arxiv.org/pdf/2202.03164.pdf.
[20] M. Chung, E. Ko, H. Joung, S. J. Kim, Chatbot e-service and customer satisfaction regarding luxury brands, Journal of Business Research 117 (2020) 587–595. doi:10.1016/j.jbusres.2018.10.004.
[21] J. Feine, S. Morana, U. Gnewuch, Measuring service encounter satisfaction with customer service chatbots using sentiment analysis, In: 14 Internationale Tagung Wirtschaftsinformatik (WI2019), 2019, pp. 1115–1129. URL: https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1248&context=wi2019

The story of experiences — the evolution of branding

Ania A Drzewiecka[1]

[1]*Heriot-Watt University, United Kingdom*

Abstract

Purpose This study aspires to contribute to the literature on the evolutionary stages of the concept of branding through the acknowledgement of its connection with sensory (tangible) and philosophical (intangible) experiences. The concept of branding is planned to be understood from a two-fold perspective on experiences, palpable and non-physical. Design/methodology/approach In this qualitative research, a wide set of sources on the concept of branding lied the grounds for a historical methodology focused on providing a thorough insight into branding and its relatedness to tangible and intangible experiences. Findings This paper offers comprehensive views on the historical developments associated with the concept of branding. Through this investigation, a plethora of phenomena on the tangible and intangible character of branding has been unearthed confirming the complexity of branding and the experiences it evokes. This study demonstrates that understanding the concept of branding with its multi-faceted nature could offer a valuable source of support toward bettering a cognizance of the future of a brand and its reputation in a marketplace. Originality/value This investigation employs the approach of brand experience as sensations, feelings, cognitions and behavioural responses introduced by Brakus, Schmitt and Zarantonello (2009) and builds upon their call for more research on experiences. This exploration is original in its approach as it considers a two-fold character of experience and its presence in branding. The investigation's novel character is embraced by distinguishing those two types of experiences against branding and simultaneously integrating both, tangible and intangible, occurrences as a stream of intrinsic characteristics of the essence of branding.

Keywords

Branding, Branding evolution, Brand research, Brand experiences

1. Introduction

The concept of branding has been vastly covered in the literature through diverse approaches to its understanding. From emphasizing the complexity of brands and their impact on consumers' perceptions (De Chernatony, 2010), voicing the controversial character of defining branding in marketing (Kapferer, 2012), agreeing that branding can be defined as a differentiating factor between competitive parties (Aaker, 2009; Van Zyl, 2011; Du Toit & Erdis, 2013) and the distinguishing characteristic can be a name, design, sign, symbol or "a combination of these" (Committee on Definitions, 1960, p.8) as pointed by the American Marketing Association, to seeing branding as something richer and far mor than a name and logo (Aaker, 2014), as an image that is evoked in consumers' minds (De Chernatony and Riley, 1998), a promise of value to consumers (Kapferer, 2012), a totality of everything that people can "think, feel, suspect, imagine, believe, wish and say about a brand" (Middleton, 2011, p. 108), and finally defining a brand as "the definition of your organisation" (Jones and Bonevac, 2013, p.117-8).

CERC 2023: Collaborative European Research Conference, June 09–10, 2023, Barcelona, Spain

a.drzewiecka@hw.ac.uk (A. Drzewiecka)

ISBN 978-3-96187-021-9
https://doi.org/10.48444/h_docs-pub-518

This investigation adapts the approach of brand experience as sensations, feelings, cognitions and behavioural responses introduced by Brakus, Schmitt and Zarantonello (2009) and builds upon their call for more research on experiences. An improved understanding of brand experiences generates vast opportunities for further inquiry. According to Schmitt (2009) experiences profoundly focus on "one of the most important aspects of our lives: the seemingly superficial world of brands" (p.419). This paper's general theme is the evolution of brand experiences and realising the value of the distinction of experiences that accompany branding. The primary aim of this research paper is to discuss the evolutionary phases of brand experiences. The principal objectives that support the achievement of this investigation outcome are focused on distinguishing the development stages of branding, understanding the concept of brand experiences, and studying, both, tangible and intangible types of experiences evoked by brands. Finally, a simple framework for integrating the tangible and intangible brand experiences and supporting strategic management of branding is presented. This model serves as a supportive tool for executives and managers involved in or/and responsible for the direction of activities and programs related to branding in their organisations. The paper concludes by identifying the significance of this examined field to academic and non-academic grounds.

This qualitative research is designed in accordance with a historical research focus that encourages collating a wide spectrum of sources to enrich the desired outcomes of the study. Witkowski and Jones (2006, p.76) emphasize that "collecting different sources, both within and across categories, is highly desirable". This historiography of the concept of branding poses some important questions and attempts to respond by analysing systematically collected evidence from various sources and historical studies including primary sources (archives), secondary sources (other scholars' works), running records (case studies notes), and artefacts (artworks). This historical research focused on brand experiences and their evolution is built on the historical fundamentals of an array of sources to enhance the overall credibility of the study.

In this qualitative research, a wide set of sources on the concept of branding lied the grounds for a historical methodology focused on providing a thorough insight into branding and its relatedness to tangible and intangible experiences. The selected methodological approach for this study, empirical historiography, aspires to deliver factual data free of judgements or interpretations to allow an objective set of historical events to emerge rather than a contextual analysis or explanation enriched with opinions and suggestions. According to Elton (1967), the "historical method is no more than a recognised and tested way of extracting from what the past has left and the true facts and events of that past" (Danto, 2008, p.12). The descriptive history model adopted by this study aims to outline neutral grounds for social scientists that can be further used in academic and non-academic settings to investigate the area of the evolution of brand experiences and the concept of branding.

2. Evolution of branding

With many attempts to define branding one important question about the origins of branding emerges. Herman stated that "branding, as a concept, is older than the modern theory" (2003, p.71), providing a solid ground for the value of seeking the historical events that shaped the modern concept of branding. Tracing the history and the evolution of branding many scholars

Figure 1: A row of ham image found in Pompeii

drew attention to the Norse word "brandr" which was used in marking cattle (Hart and Murphy, 1998; Keller, 2008; Riezebos, 2003) meaning to burn a symbol or a mark on a skin surface that would identify livestock's owners (Khan & Mufti, 2007; Maurya & Mishra, 2012; Roper & Parker, 2006). This identifier served the purpose of emphasizing the ownership of livestock and also distinguishing them from others present in marketplaces. The names of families were used for several purposes, firstly, as a brand, secondly, helped to associate the livestock, and thirdly as a mark of quality (Sheth & Parvatiyar, 1995). The branding of livestock dates back to 2000 BC (Dranove and Jin, 2010), there is some earlier evidence of branding in relation to the origins of products. This Ancient Norse concept of symbolising ownership (around 350AD) has shaped the modern understanding of branding as an identifier, a mark, or a symbol. Approximately 600 years later the meaning of branding evolved and spread to a burning piece of wood (950AD). Another 300 years added some context of a tool or a factor that can burn a piece of wood.

Many would argue that branding beginnings can reach as far as the human species initiated some basic forms of communication and information exchange. Possibly the term depicting the meaning of a symbol, or an item was not even thought of when the various ways of promoting goods amongst our ancient ancestors were practised. Some primitive forms of communicating messages about who was offering what and where were discovered in anthropological studies of the Greeks and Romans. Room (1998) discussing the historical origins of branding pointed the very early forms of advertising were related to a personal level such as a name of an individual was equally critical as an item that was being offered (Hart and Murphy, 1998). This phenomenon can find its mirroring practice of naming conventions of stores based on the owners of those establishments. In the earlier days of ancient Rome first commercial exchanges, signs or images were the media to visually present an offering to the public. Fig. 1 presents a sign found in the ruins of Pompeii displaying several ham pieces offered by a local butcher (Hart and Murphy, 1998).

While Dranove and Jin (2010) state that the branding of cattle dates back to 2000 BC, there are indications that branding practices indicating products' origin of production date back even further when branded livestock in 2700 BC was used as an identifier for stolen cattle by Ancient Egyptians (Khan and Mufti, 2007). The main purpose of marking and placing some pictorial symbols on items, products, and other objects as well as livestock was to differentiate and trademark as a form of ownership but also quality and guarantee (Blackett, 1998; Farquhar,

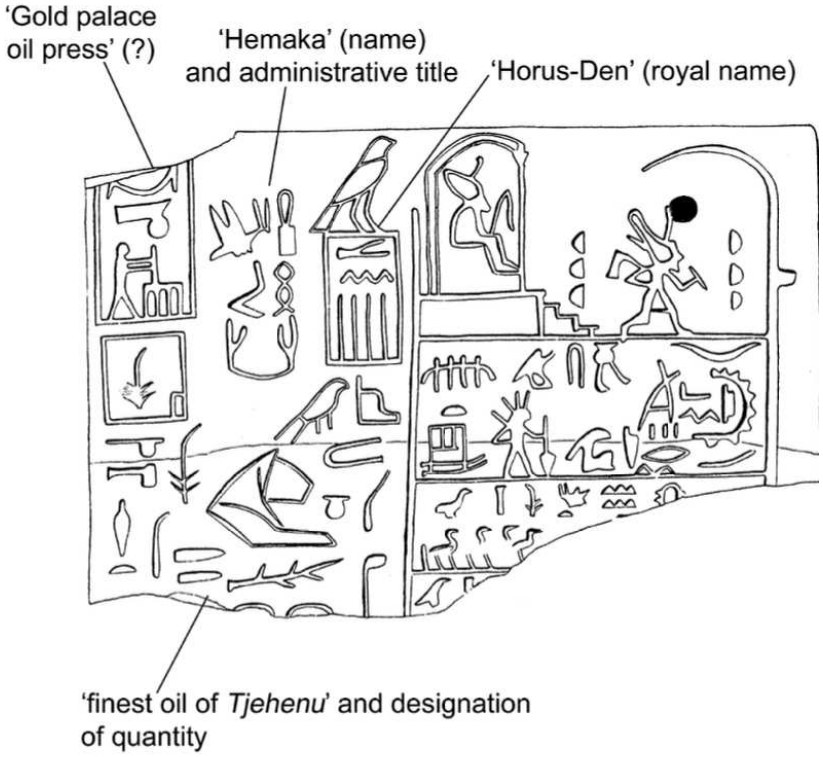


Figure 2: An ancient Egyptian commodity label (after Petrie 1900, pl. 15/6 found in Wengrow, 2008)

1990; Khan and Mufti, 2007). The abundance of evidence of branding practices in the forms of pots and clay figurines marked by a potter found in ancient Greece and Rome over the years provides solid grounds for identifying the early visual forms of branding (Khan and Mufti, 2007).

Some other forms of corroboration that some branding origins can be discovered in even earlier periods such as 7000 BC in the Mesopotamia region, 3500 BC in the Middles East, and 3000 BC in Egypt were unearthed by Wengrow (2008), Eskhards and Bengtsson (2010). The evidence of early branding practices encompasses some sealing practices (Wengrow, 2008) that could have acted as indicators of quality and origins (Eckhardt and Bengtsson, 2010) as well as marks of ownership (Yang, Sonmez, & Li, 2012). The early discoveries' traditions of the use of visual symbols to mark ownership, represent value and guarantee as well as quality are prominent antecedents of modern branding and its heritage associations with the ancient civilisations' practices. In Fig. 2 which depicts an ancient Egyptian commodity oil label (3000 BC) and Fig. 3 illustrating a modern Australian commodity wine label, there are some evident similarities between labelling practices in relation to the presence of quality, origins and the core messages (Wengrow, 2008).

Over 5000 years later and the essence of branding remain close to the ancient commodity practices.

The noticeable similarities between ancient commodity label practices and modern branding are the marks of early brand activities (Wengrow, 2008). Some other branding- related practices discovered between 2700 BC and 2000 BC have revealed that identification and differentiation were the main aims of stamping pottery in China (2700 BC) (Eckhardt & Bengtsson, 2009) as well as craftsmen sealing of containers and other items in modern-day India (the Indus Valley)

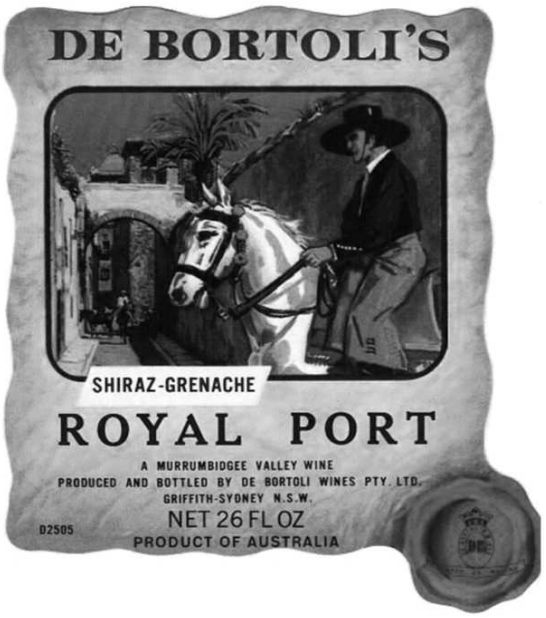


Figure 3: A modern Australian commodity label (courtesy of De Bortoli's Ltd. Found in Wengrow (2008))

(2250-2000 BC) (Yang et al. 2012). Moore and Reid (2008) argued that those seals made by craftsmen were the first signs of brand imagery, oftentimes presenting vivid pictorial displays of animals or gods, and used as trademarks in shops (Reddi, 2009).

Further development stages of branding discovered through findings from the period 2000 BC — 1500 BC, known in history as the middle bronze era, illustrate the nurturing of traditions of marking products to present origins and ensure quality. In Shang China, items were crest marked by a king (wang) to carry a Zu family identification, considered the initial form of primitive branding (Moore & Reid, 2008). Between 1500 BC and AD 500 across the Mediterranean regions the evolution of branding took the form of large ceramic containers (amphorae) to initiate the beginnings of consumer packaging (Grace, 1979; Twede, 2002; Lawall, 2021). Those popularly used by the ancient Greeks and Romans commercial transport large vessels were used for shipping wine and oil (Twede, 2002) and gave the foundations for successful design practices for packaging used by brands. The amphoras' design objectives were clear and functional, they had to be economical to manufacture and fit for the purpose of shipping (Twede, 2002). According to Twede (2002), today's marketing campaigns are directly influenced by the traditions of the use of the trademark shape of amphoras and its mark on coins. Holleran (2012) reminded of product identification and differentiation as the essence of the practice of marking those ancient vessels.

Another period in the evolutionary stages of branding that engraved its importance in historical findings belongs to the Song Dynasty in China (AD 960 — AD 1127) where the first complete brand emerged, the White Rabbit brand of a needle manufacturer (Eckhardt & Bengtsson, 2009). This Chinese brand, the White Rabbit, initiated practices that guide today's modern branding such as logo print on paper packaging displaying the producers' details, production specifications, usability and discount (Muller, 2017). The brand name took its inspiration from a Chinese legend that considers a white rabbit as an essence of feminine

ideologies (Lai, 1994; Masako, 1995). Some other contributions of the Song Dynasty period to the advancement of branding are mass advertising (Starcevic, 2015) and block-printing as an initiator of mass communication (Landa, 2005). Between the period of the Song Dynasty and the Industrial Revolution, a couple of practices such as printing labels for products by Chinese manufacturers in the 14th century and the design of the modern print press in the 15th century created a solid communication ground enabling organisations, shops, craft practices to share their products offering to customers.

The new evolutionary stage was initiated by the Industrial Revolution during the 18th century which is considered to be the predecessor of modern branding (Roper & Parker, 2006). During the Industrial Revolution mass production and mass communication replaced the individuality of producers and their products, and the focus was placed on productivity (Varey, 2011). During the 19th century, organisations realised the power of competitiveness and the element of personalisation emerged as a differentiator but also relationship-building support between brands and their customers (Muller, 2017). Those market dynamics and organisations' movements towards the customers to connect initiated the thinking around brands and their ability to evoke feelings at the start of the 1900s (Klein, 2000). As the competition started rising to its power, organisations began to outrun their rivals in ways to adapt to the demands of customers and convince the market about the superiority of their offering over their competitors.

In the 20th century, the popularity of marketing and communication directed towards raising engagement between organisations and customers or potential customers started seeing branding as part of corporate identity (Suchman, 2007). According to Daffey and Abratt (2002) between the 50s and 80s (20th century) the shift from corporate image and personality to corporate brand management can be observed (Muller, 2007). The concept of branding has gained its value as a central element that conveys originality, quality and differentiation as part of a corporate body offering. Together with the goals of authenticity and uniqueness for brands, the threat of replaceability transpires, therefore trademarks as ways to guard brand identities have surfaced and remain relevant in modern branding (Duguid, 2009; Muller, 2007; Petty, 2013).

The presence of branding or brands as separate phenomena was insignificant or rather non-existing before the 19th century. The main reason was related to the practice of selling goods mainly in bulk and lack of distinction between the superiority or a "better quality" or "better value" of one store over another as in the majority of towns there were single stores offering sacks or barrels of products (Bastos & Levy, 2012). The late 19th and early 20th centuries saw the rise of naming, labelling ad packaging conventions as a way to add some extra value to offered products, an example of which include "producers such as Folger (1872), Kraft (1903), and Vlasic (1942) showed pride in their brands by putting their names on their coffee, cheese, and pickles, respectively" (Bastos & Levy, 2012, p. 354). Moore and Reid (2008) pointed out that the end of the 19th and early 20th centuries had a prominent impact on the evolution of branding as a result of the appearance of media "[...] this is largely a phenomenon that could have only occurred starting at the end of the nineteenth century and into the twentieth century, due to the media (TV, radio, print advertising, e-marketing, etc.)" (p. 429).

The Second World War did not bring much progress, nor vitality to civilisation, but it served as a competitive ground for producers operating to serve customers' demands and supply the war's needs. The rising levels of produced goods and growing appetites for purchasing amongst customers in the 40s and the 50s (20th century) were described as a period of "Customer

Revolution" (Bastos & Levy, 2012). During that time many brands came to competitive battles for the superiority pedestal for such categories as coffee, hamburger, soft drinks and more.

Since the end of the Second World War and the rising level of competition between products, organisations, and countries, there was a need for a "greater awareness of the social and psychological nature of products — whether brands, media, companies, institutional figures, services, industries, or ideas" (Gardner and Levy, 1955, p. 34) that would support customers in making conscious purchasing decisions. The importance of brand names and the symbolism associated with presentation conventions has emerged as a practice aiding customer choice. The symbolic character of products offered by brands has risen in its importance to customer behaviour and selection dynamics observed around buying activities because oftentimes customers' choices are driven by the symbolic meaning of products in addition to their primary use (Gardner and Levy, 1955). The modern era has embraced the value of personality in relation to branding and many characteristics have been associated with brand personality, amongst which, it enables to market a brand in various cultures (Plummer, 1985), it allows self-expression of consumers (Belk, 1988), it impacts consumer preference and use (Biel, 1993), it acts as a differentiator in a category of products (Halliday, 1996), it can be captured through five basic dimensions: excitement, competence, sincerity, sophistication, and ruggedness (Aaker, 1997).

The review of the historical literature on the subject of branding unearthed some common themes that serve as evolution stages of the concept of a brand. Moore and Reid (2008) pointed out the main development stages of brand characteristics that can be categorised based on their focus on information and image. Table 1 presents an insight into the evolution of brand characteristics through the time periods of the eras of early bronze, middle bronze, late bronze, the iron age revolution, the iron age and the modern era.

It is noticeable that the informational focus on origin and quality was the branding driving force during those early transactional periods in history when products were marked to display the location they were made and often ownership. On the contrary, image-focused branding related to power, value and personality was the core objective for later historical transformational periods focused on enriching the buyer and the choice made to own a product representing a brand. The modern branding evolution displays hybrid characteristics of informational and transformational value that support communicating cultural meaning (McCracken, 1986) that relates to, both, information and image (Moore & Reid, 2008).

3. Concept of brand experiences

The idea of branding and its presence in academic literature is relatively modern, it dates back to the 1950s of the last century when works on names and the symbolic value of products started emerging (Gardner & Levy, 1955). However, when studying people's relationships with goods, symbols, pictorial signs, trademarks, identifiers of origins and quality, since the early period in history when goods were made for exchange, the concepts of ownership and differentiation were not alien. The essence of trademarks, those early distinguishing signs, was to assure customers of the value of products represented by them and emphasise their origins. Those positive associations started arising.

On the contrary, history delivers numerous examples of the primitive forms of branding,

Table 1
Brand characteristics in the ancient and modern worlds (adapted from Moore and Reid (2008))

	Brand Characteristics				
Period	**Information : Origin**	**Information : Quality**	**Image : Power**	**Image : Value**	**Image : Personality**
Early Bronze IV 2250 - 2000 BCE The Indus Valley	X	X			
The Middle Bronze Age 2000 - 1500 BCE Shang China	X	X			
The Late Bronze Age 1500 - 1000 BCE Cyprus	X	X		X	
The Iron Age Revolution 1000 - 500 BCE Tyre	X	X	X	X	
The Iron Age 825 - 336 BCE Greece	X	X	X	X	
Mdern	X	X	X	X	X

marking, that convey the negative connotations of the experience of branding. Some salient exemplars include the Second World War and the practice exercised by Nazis to brand people with numbers who were later sent to concentration camps, and the period of the trans-Atlantic slave trade when the Africans were marked by the branding irons for ownership display purposes (Keefer, 2019). The act of branding studied from the perspective of its archaeological and anthropological history provides vast evidence that its negative connotation is rooted in the act of stigmatization of the criminal and dehumanizing. Keefer (2019) stated that "branding is one of the most charged symbols of the evils of slavery, and branding irons are displayed from that period as a testament to the inhumanity of the slave trade" (p. 660). Those permanently placed "country marks" on enslaved individuals in the Americas were used for identification and categorisation (Gomez, 1998). Therefore, the very early experiences of branding concern body modification practices representing violent acts executed against the will of a branded individual. Warmer (2016) pointed out the relatedness of the commodity and branding when branding was used to commodify individuals in the trans-Atlantic slave period.

The historical literature provides a rich source of evidence that the early practices of branding were not only exercised on humans and animals but also container ships or food products, meaning that the branding experience was purposeful to act as some form of protection and guarantee (Keefer, 2019). The act of branding of living flesh was practiced far earlier than the period of slavery and the Second World War, some examples were found in the law code of Hammurabi in 1754 BC and Pharaonic Egypt when slaves had inscribed their owners' names on their arms (Handcock, 1920). The act of branding and associated experiences maintain negative attachments from the earliest practices of trade until the early modern period. This

Figure 4: A drawing after sketch in Register of Liberated Africans 1808-1812. Sierra Leone Public Archives, Fourah Bay College, Freetown, Sierra Leone (Found in Keefer, 2019)

Figure 5: A drawing after sketch in sketch in RLA 1808-1812, SLPA (Found in Keefer, 2019)

form of marking evokes emotions supporting the sense of authenticity and ownership from the perspective of the owner, however, from the standpoint of those branded, in the event of studying the experiences of living individuals, it could be seen as dehumanizing and commodifying. Figures 4 and 5 present some drawing practices as acts of branding the living flesh of humans and animals to indicate ownership and were also used in register books to report enslaved individuals and brands as owners.

The aspiration of this study is not to provide an exhaustive historical dataset on the concept of branding and brand experiences but to pinpoint some of the significant stages in the evolution of branding hoping to stimulate curiosity for future research and interpretation. Therefore, the discussion around the negative connotations associated with branding is important to retain impartiality in relation to scholars who value more and dedicate their attention to positive connotations of branding and those whose interest focuses on the negativity around branding. Since the Industrial Revolution, the importance of branding and advertising to support the dissemination of information about goods and products has grown. The positive experience of branding and promotion of items reached its peak stimulating demands for products and new customers, the mass communication efficiently served the need of manufacturers and producers to attract potential buyers. The appearance of the first advertising agencies in England was noted in the 1800s, they were working with sellers on approaches to reach customers who were not necessarily fond of reading newspapers, the first banners, poles and branded items such as umbrellas were launched, and the significance of an image and visual communication became apparent (Landa, 2005). The brand's name, label and packaging were dedicated to stimulating the demand and transforming a product line created by a brand into an object of desire. Landa (2005, p. xxiii) suggested that the rise of a "brand word" was a sign of the twentieth century in

industrialized countries where corporations treated the brand identities created for their goods as sets of identifications standards to evoke happiness in consumers. The unified look achieved by consistent visual communication was the essential purpose of organisations which were aiming for having a personality engraved within their corporate style.

Branding can be described as "the entire development process of creating a brand, brand name, brand identity, and in some cases, brand advertising" (Landa, 2005, p. 9), a name, term, design, symbol that identifies goods (The American Marketing Association), and a strategic company objective (Kapferer, 2008; Keller, 208). From the symbolic representation of beliefs, values, and personality to an element that creates an emotional attachment with consumers, branding has been transformed (Beig & Nika, 2019) and its meaning has deepened. The challenging dynamics of global markets have impacted the approach organisations undertake to attract potential clients. Creating memorable experiences is the focus of brand offerings (Beig & Nika, 2019), oftentimes infused with hedonic incidents that leave a long-lasting emotional impact on customers' mental images created through contact with brand offerings. Contentment, pleasure and emotional attachment are the desirable feelings that brands focus on evoking (Hirshman & Holbrook, 1982; Dhar & Wertenbroch, 2000). Beig and Nika (2019) highlighted the importance of focusing on the perceived hedonic value of a brand that stresses the urgency of "primary process thinking in accord with the pleasure principle" (Holbrook & Hirshman, 1982) that indicates pleasurable experiences associated with a brand. Sensory, emotional, cognitive, behavioural and relational values are often replacing functional values (Schmitt, 1999) when considering experiences as the notion of connection or building blocks of a relationship between a brand and a consumer, which is critical to bonding experience (Fournier, 1998).

The concept of brand experience is considered essential to capturing the ethos of branding (Schmitt, 2009). With expanding demands of a global consumer who carefully selects items and brands for further engagement, the expectations rise to a level of a unique kind of delight that engages the senses and delivers excitement and authentic experiences. Brand experiences are "subjective, internal responses (sensations, feelings and cognitions) as well as behavioral response evoked by brand-related stimuli that are part of a brand's design and identity, packaging, communications and environment" (Schmitt, 2009, p. 417). Improved offerings, personalised products and widely available items are the results of utilising customer experiences evoked by brands to deliver value (Addis & Holbrook, 2001; Prahalad & Ramaswamy, 2004). From the purely hedonic purpose of fun, fantasies and feelings (Holbrook & Hirschman, 1982) to rational and emotional attributes (Schmitt & Rogers, 2008), experiences form a unique relationship between a consumer and a brand that can be of a different character, aesthetic, escapist, entertainment and educational (Pine & Gilmore, 1999). Brands can evoke various experiences, according to Schmitt (1999, 2003), sense experiences (sensory perception), feel experiences (affect and emotions), think experiences (creative and cognitive), act experiences (physical behaviour, actions, lifestyles), relate experiences (connection with a group or culture) (Keller & Lehmann, 2006). Branding can be perceived as a method of creating customers value through experiences (Vargo & Lusch, 2008; Lee & Jeong, 2014).

4. Tangible and intangible brand experiences: natural science, philosophy, psychology, and semantics

Brand experiences can be viewed as customers' sensations, feelings, cognitions, and behavioral responses evoked by brand-related stimuli (a brand's design and identity, packaging, communications, and environments) (Brakus et al., 2009).

Whether there is a specific need that arises and requires to be satisfied (physiological, survival) or a desire related to wants and self-actualization needs (Maslow, 1943, 1954), humans continuously look for satisfaction experiences that would meet deficits whether physiological and survival or growth. Connecting this discussion about brand experiences to Maslow's hierarchy of needs (1943, 1954, 1970a, 1970b) unravels the primary and secondary motives behind seeking fulfilment of needs through experiences. This observation creates a path towards bridging tangible experiences generated by brands to fulfil the primary needs and wants of customers based on biological requirements and survival as well as safety and comfort, and their intangible counterparts that serve the needs of self-actualisation, belonging, esteem, as well as cognitive, aesthetic and transcendence needs.

Maslow pointed out that humans are motivated by a hierarchy of needs (1943, 1954, 1970a, 1970b), therefore it is apparent that to fulfil needs and wants on various levels individuals will be looking for satisfaction. Experiences that lead to addressing various needs deficits can occur directly and are often led by conscious decisions and calculated choices, but some experiences are evoked indirectly mainly by media of promotion, advertising, or marketing communication (Brakus et al., 2009).

Customers are often exposed to diverse attributes and features of goods that provide stimuli aiming to fulfil individual needs. In addition to utilitarian features, items offered by brands are enriched with specific brand-related stimuli including colours (Bellizzzi & Hite, 1992; Meyers-Levy & Peracchio, 1995), shapes (Veryzer & Hutchinson, 1998), typefaces (Mandel & Johnson, 2002), slogans (Keller, 1987) that generate "subjective, internal consumer responses" (Brakus et al., 2009, p. 53), brand experiences. Tangible brand experiences can be described as emotions or thoughts or preferences evoked or caused by tangible aspects of a brand or branding such as brand image, brand logo, brand colours, brand slogans, and marketing communication. Brand intangibles refer to those aspects of a brand or branding that is non-physical, intangible, and do not represent specific, concrete characteristics (Levy, 1999; Keller & Lehmann, 2006).

To deepen the understanding of experiences evoked by brands, it is important to review germane works from other fields to further the interdisciplinary and multi- dimensional character of human experience. The value of comprehension of different levels of experiences is sought to be better understood by marketing and management professionals to support their practices but also it aspires to lay foundations for better-informed multidisciplinary research on branding and the experiences it evokes.

From a philosophical viewpoint, experience is a kind of cognition that requires understanding (Kant, 1929), it comes from perception and memory, and it initiates art and science according to Aristotle (Gregoric, 2006), and is a subjective mental phenomenon stated Descartes in the 17th century (Britannica, 2023). The Kantian view of experiences as the knowledge that forms the human understanding of the world is based on a priori experiences. Dewey, the 19th/20th-

century philosopher argued with the Kantian perception of experiences, and in addition to the intellectual experience favoured by Kant, Dewey proposed experiences evoked by feelings, perception (senses) and actions (Brakus et al., 2009). From a psychological perspective, experiences are often connected to pleasure. Dube and LeBel (2003) distinguished pleasure dimensions including intellectual, emotional, social and physical pleasures. "Emotional response, which was part of [the] pleasure category, is at an implicit level. Pleasure, a hierarchical concept, is considered at various levels" (Bapat & Thanigan, 2016, p. 1359).

Jantzen (2013) considered experience as fundamental for understanding mental life. Some historical definitions of experience crafted within the science of psychology considered psychology as a science of immediate experience and the foundation of the humanities (Wundt, 1896). James (1911) proposed viewing psychology as radical empiricism based on understanding the reality that is created every single day rather than encapsulated in a priori concepts.

The diverse range of experiences as well as the semantic connotations associated with the word "experience" enrich the understanding of the experience phenomenon. Studying the semantics of the word "experience" it is apparent that the variety of meanings encapsulated in the word itself provides a rich foundation for its use and understanding. According to the Oxford University Press dictionary, a spectrum of connotations is listed, and an experience can be "the knowledge and skill that you have gained through doing something for a period of time; the process of gaining this", "the things that have happened to you that influence the way you think and behave", "an event or activity that affects you in some way", "what it is like for somebody to use a service, do an activity, attend an event", "events or knowledge shared by all the members of a particular group in society, that influences the way they think and behave" (2023). Similarly, the Merriam-Webster dictionary distinguishes the following denotations, "direct observation of or participation in events as a basis of knowledge", "practical knowledge, skill, or practice derived from direct observation of or participation in events or in a particular activity", "something personally encountered, undergone, or lived through", "the conscious events that make up an individual life", "the act or process of directly perceiving events or reality" (2023). Experiences cannot be narrowed to only goods or services consumed, they cannot be purchased nor offered by brands for selling, they are evoked by interactions and are subjective, and their mental and physical significance is complex (Jantzen, 2013).

5. Strategic brand management framework

Brand experiences occur during various events and periods in time of consumers associated with a brand. The need for knowledge about how to intensify the customer experience during, after or ahead of direct and indirect interaction with a brand, increases among organisations outrunning each other in ways to better their engagement practices. Some engaging brand experiences can attract but also alter consumers' perceptions (Hoch, 2002). Brands need to understand the foundations of behaviours of their target markets on an individual level, to better satisfy their needs and deliver customer delight. Brand experience relates to cognitive, motivational and affective aspects (Schmitt, 2009). Therefore, to support better management practices of a brand, an integrative framework of strategic brand management based on a 3-phased approach is proposed (Fig. 6). The model seeks to increase cognizance of a consolidative

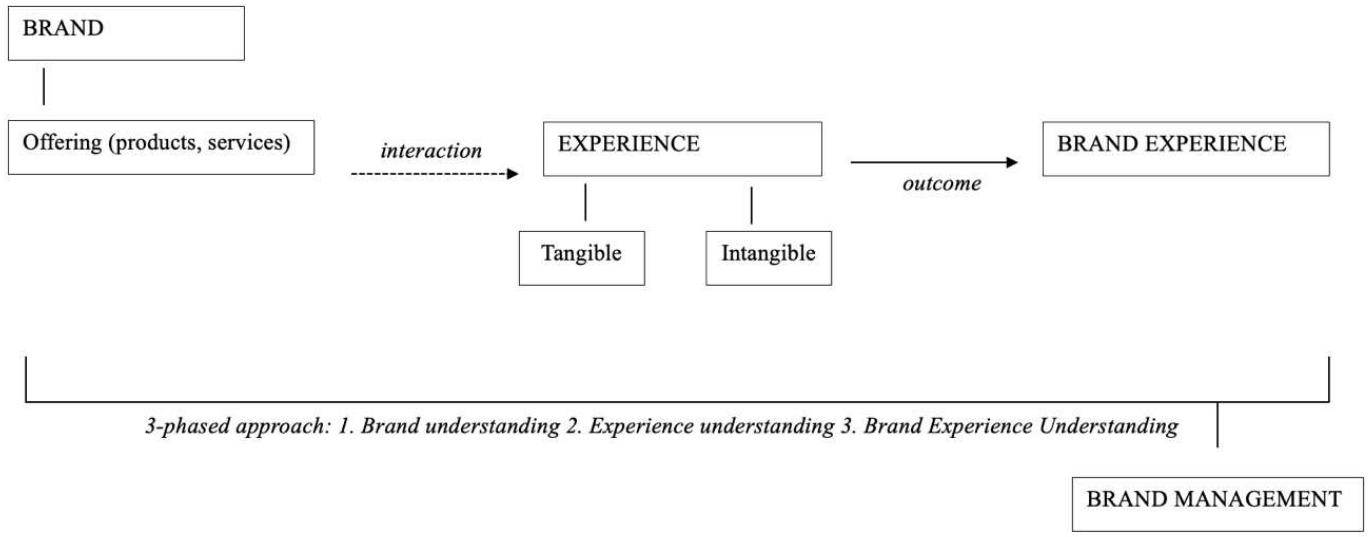


Figure 6: A brand management model (authors own)

nature of a brand and brand management of individuals involved in decisions regarding the presence and future status and behaviour of a brand.

This framework (Fig. 6) for integrating the tangible and intangible brand experiences aims to support strategic management of branding and ignite curiosity in academic discussions towards studying brand-related concepts through the lens of their interconnected and multi-dimensional character.

6. Significance and the future direction

Developing, exploring and authenticating the integrative character of branding serves the purpose of deepening understanding of its components and interrelationships. Since brand experiences arise in a plethora of settings and following Schmitt et al. (2009) conceptualisation of brand experience as "subjective consumer responses that are evoked by specific brand-related experiential attributes" (p.65), different brand dimensions (sensory, affective, intellectual, and behavioural) require adequate understanding and their impact on experiences evoked in customers. The results of this study contribute to a clearer perception of the concept of branding and the two-fold character of experiences and their implications on the complexity of customer experiences and brand management. From a theoretical standpoint, the study proposes a conceptual framework that integrates existing explanations of tangible and intangible experiences, and brand experiences. This model is a step forward towards validating the interconnectivity and interdisciplinary character of branding and brand experiences. The current study contributes to the existing literature by identifying primary areas of attention for improving knowledge and practice related to brand management. Considering the ascending competition in global markets and the significance of brand equity and brand position, the role of brand experiences has gained attention as they act as a knowledge base for managers and executives involved in brand-related practices and decisions. Beyond theoretical contributions, this paper serves as practical guidelines for practitioners across industries. This study's novel

character is embraced by the proposal of studying brand experiences from a two-fold perspective as well as introducing an integrative framework aiding strategic brand management efforts.

A call for further research is directed towards further investigations of the multi-disciplinary features of branding and potential discoveries of its interrelatedness with other yet not analysed disciplines. In addition, a future study may focus on validating the proposed framework (Fig. 6) by employing its principles for various types of brands. It would be also recommended for future research in identifying how customers of brands offering categorise their experiences and whether they are cognizant of tangible and intangible differences while experiencing brands.

References

[1] Aaker, J. L. (1997). Dimensions of brand personality. Journal of Marketing Research, 34 (August), p. 347-356.

[2] Aaker, D.A. (2009). Managing brand equity. Capitalizing on the value of a brand name. The Free Press, London.

[3] Aaker, D. A. (2014). Aaker on branding. 20 principles that drive success. Morgan James Publishing.

[4] Addis, M., & Holbrook, M. B. (2001). On the conceptual link between mass customisation and experiential consumption: an explosion of subjectivity. Journal of Consumer Behaviour: An International Research Review, 1(1), p. 50—66.

[5] Bapat, D., & Thanigan, J. (2016). Exploring relationship among brand experience dimensions, brand evaluation and brand loyalty. Global Business Review, 17(6), p. 1357-1372.

[6] Bastos, W. & Levy, S.J. (2012). A history of the concept of branding: practice and theory. Journal of Historical Research in Marketing, 4(3), p. 347-368.

[7] Beig, F.A. & Nika, F.A. (2019). Brand experience and brand equity. Vision, 23(4), p. 410—417.

[8] Bellizzi, J. A., & Hite, R.E. (1992). Environmental color, consumer feelings, and purchase likelihood. Psychology and Marketing, 9 (5), p. 347-63.

[9] Belk, R. W. (1988). Possessions and the extended self. Journal of Consumer Research, 2 (September), p. 139-168.

[10] Biel, A. (1993). Converting image into equity, in Brand Equity and Advertising, David A. Aaker and Alexander Biel, eds. Hillsdale, NJ: Lawrence Erlbaum Associates. p. 67—82.

[11] Blackett, T. (1998). Trademarks. Basingstoke: Macmillan.

[12] Brakus, J.J., & Schmitt, B.H., & Zarantonello, L. (2009). Brand experience: what is it? How is it measured? Does it affect loyalty? Journal of Marketing, 73(3). Retrieved 2023, April 14 from https://www.researchgate.net/publication/228168877_Brand_experience_What_Is_It_How_Is_It_Measured_Does_It_Affect_Loyalty.

[13] Britannica. (2023). Being, nature and experience. Retrieved 2023, April 20 from https://www.britannica.com/biography/Socrates.

[14] Daffey, A. & Abratt, R. (2002). Corporate branding in a banking environment. Corporate Communications: An International Journal, 7(2), p. 87-91.

[15] Danto, E.A. (2008). Historical research. Oxford University Press. Retrieved 2023, April 17

[16] de Chernatony, L. (2010). Creating powerful brands. Routledge: London, 4th edition. Retrieved 2023, April 14 from https://www.taylorfrancis.com/books/mono/10.4324/9781856178501/creating-powerful-brands-leslie-de-chernatony.
[17] de Chernatony, L., & Riley, F.D. (1998). Modeling the components of brand. European Journal of Marketing, 32(11/12). Retrieved 2023, April 15 from https://www.researchgate.net/publication/38176107_Modeling_the_Components_of_Brand.
[18] Dewey, John. (1922). Human nature and conduct. New York, NY: The Modern Library.
[19] Dewey, J. (1925). Experience and nature (rev. ed.). New York, NY: Dover.
[20] Dhar, R., & Wertenbroch, K. (2000). Consumer choice between hedonic and utilitarian goods. Journal of Marketing Research, 37(1), p. 60—71.
[21] Dubé, L., & LeBel, J. L. (2003). The content and structure of laypeople's concept of pleasure. Cognition and Emotion, 17(2), p. 263—295.
[22] Duguid, P. (2009). French connections: The international propagation of trademarks in the nineteenth century. Enterprise and Society, 10(01), p. 3-37.
[23] Dranove, D. & Jin, G. Z. (2010). Quality disclosure and certification: theory and practice. NBER Working Paper No. w15644. National Bureau of Economic Research.
[24] Eckhardt, G.M. & Bengtsson, A. (2010). A brief history of branding in China. Journal of Macromarketing, 30(3), p. 210- 221.
[25] Elton, G.R. (1967). The practice of history. Fontana Paperbacks. Retrieved 2023, April 17 from http://seas3.elte.hu/coursematerial/LojkoMiklos/Elton,_The_Practice_of_History_47_pp.pdf.
[26] Farquhar, P. H. (1990). Managing Brand Equity. Journal of Advertising Research, 30(4).
[27] Fournier, S. (1998). Consumers and their brands: Developing relationship theory in consumer research. Journal of Consumer Research, 24(4), p. 343—373.
[28] Gomez, M. A. (1998). Exchanging our country marks: the transformation of African identities in the Colonial and Antebellum South. Chapel Hill: University of North Carolina Press, p. 39—40, 42, 97—98.
[29] Gardner, B.B., & Levy, S.J. (1955). The product and the brand. Harvard Business Review, March-April, p. 33-9.
[30] Grace, V. (1979). Amphoras and the ancient wine trade (Vol. 16). ASCSA.
[31] Gregoric, P. (2006). Aristotle's notion of experience. Archiv fur Geschichte der Philosophie, 88(1). Retrieved 2023, April 19 from https://www.researchgate.net/publication/249941681_Aristotle%27s_Notion_of_Experience.
[32] Halliday, J. (1996). Chrysler brings out brand personalities with '97 ads. Advertising Age, 67 (September), p.3-5.
[33] Handcock, P. (1920). The Code of Hammurabi (New York: The MacMillan Company, 1920), 22, 35. Hart, S. & Murphy, J. (1998). Brands the new wealth creators. Macmillan, London.
[34] Herman, R. (2003). An exercise in early modern branding. Journal of Marketing Management, 19(7-8), p. 709-727.
[35] Hirschman, E. C., & Holbrook, M. B. (1982). Hedonic consumption: Emerging concepts, methods, and propositions. Journal of Marketing, 46(3), p. 92—101.
[36] Hoch, S.J. (2002). Product experience is seductive. Journal of Consumer Research, 29(3), p. 448—454.

[37] Holbrook, M. B., & Hirschman, E. C. (1982). The experiential aspects of consumption: Consumer fantasies, feelings, and fun. Journal of Consumer Research, 9(2), p. 132— 140.
[38] Holleran, C. (2012). Shopping in Ancient Rome: the retail trade in the late Republic and the principate. Oxford University Press.
[39] James, W. (1911). Some Problems of Philosophy. London: Longmans, Green and Co.
[40] Jantzen, C. (2013). Experiencing and experiences: a psychological framework. Retrieved 2023, April 20 from https://www.researchgate.net/publication/322302480_Experiencing_and_experiences_A_psychological_framework.
[41] Jones, C. & Bonevac, D. (2013). An evolved definition of the term 'brand': Why branding has a branding problem. Journal of Brand Strategy, 2(2).
[42] Kant, I. (19129). Critique of pure reason. London: Macmillan.
[43] Kapferer, J.N. (2012). The new strategic brand management. Advanced insights and strategic thinking. Kogan Page Limited, 5th edition. Retrieved 2023, April 14.
[44] Keefer, K.H.B. (2019). Marked by fire: brands, slavery, and identity. Slavery & Abolition, 40(4), p. 659-681.
[45] Keller, K.L. (1987). Memory factors in advertising: the effects of advertising retrieval cues on brand evaluations. Journal of Consumer Research, 14 (December), p. 316-33.
[46] Keller, K.L., & Lehmann, D.R. (2006). Brand and branding: research findings and future priorities. Marketing Science, 25(6), p. 740-759.
[47] Khan, S.U., & Mufti, O. (2007). The hot history and cold future of brands. Journal of Managerial Sciences, 1(1).
[48] Lai, W. (1994). Recent PRC Scholarship on Chinese Myths. Asian Folklore Studies, 53(1):151-161.
[49] Landa, R. (2005). Designing brand experience: creating powerful integrated brand solutions. Clifton Park: Cengage Learning.
[50] Lawall, M.L. (2021). Aegean transport amphoras (sixth to first centuries BCE): exploring social tension in a path dependency model. In: Gimatzidis, S., Jung, R. (eds) The Critique of Archaeological Economy. Frontiers in Economic History. Springer, Cham.
[51] Lee, S., & Jeong, M. (2014). Enhancing online brand experiences: an application of congruity theory. International Journal of Hospitality Management, 40, p. 49-58.
[52] Levy, S.J. (1999). Brands, consumers, symbols, and research: Sydney J. Levy on marketing. Sage Publications, Thousand Oaks, CA.
[53] Mandel, N., & Johnson, E.J. (2002). When web pages influence choice: effects of visual primes on experts and novices. Journal of Consumer Research, 29(September), p. 235-45.
[54] Masako, M. (1995). Restoring the" Epic of Hou Yi". Asian folklore studies, p. 239-257.
[55] Maslow, A. H. (1943). A theory of human motivation. Psychological Review, 50(4), 370-96.
[56] Maslow, A. H. (1954). Motivation and personality. New York: Harper and Row.
[57] Maslow, A. H. (1970a). Motivation and personality. New York: Harper & Row.
[58] Maslow, A. H. (1970b). Religions, values, and peak experiences. New York: Penguin. (Original work published 1966).
[59] Maurya, U.K., & Mishra, P. (2012). What is a brand? A Perspective on Brand Meaning. European Journal of Business and Management, 4, p.122-133.

[60] Merriam-Webster Dictionary. (2023). Experience. Retrieved 2023, April 20 from https://www.merriam-webster.com/dictionary/experience.

[61] Meyers-Levy, J., & Peracchio, L.A. (1995). How the use of color in advertising affects attitudes: the influence of processing motivation and cognitive demands. Journal of Consumer Research, 22 (September), p. 121-38.

[62] Middleton, S. (2011). What you need to know about marketing. Chichester: Capstone Publishing.

[63] Moore, K. & Reid, S. (2008). The birth of brand: 4000 years of branding. Business History, 50(4), p.419-432.

[64] Muller, R. (2017). The prominence of branding through history and its relevance to modern brands: a literature review. Global Business and Technology Association.

[65] Oxford University Press Dictionary. (2023). Experience. Retrieved 2023, April 20 from https://www.oxfordlearnersdictionaries.com/definition/english/experience_1?q=experience.

[66] Petty, R.D. (2013). 'Towards a Modern History of Brand Marketing: Where Are We Now? (In Neilson, L.C. ed. (2013). Varieties, Alternatives and Deviations in Marketing History: Proceedings of the 16th Biennial Conference on Historical Analysis and Research in Marketing, Copenhagen, Denmark: CHARM Association).

[67] Pine, B. J., & Gilmore, J. H. (1999). The experience economy. Brighton, MA: Harvard Business Press.

[68] Plummer, J. (1985). Brand personality: a strategic concept for multinational advertising, in American Marketing Association's National Marketing Educators' Conference, Robert F. Lusch, Gary T. Ford, Gary L. Frazier, Roy D. Howell, Charles A. Ingene, Michael Reilly, and Ronald W. Stampfl, eds. NY: Young & Rubicam, p. 1-31.

[69] Prahalad, C. K., & Ramaswamy, V. (2004). Co-creation experiences: The next practice in value creation. Journal of Interactive Marketing, 18(3), p. 5—14.

[70] Reddi, C.N. (2009). Effective public relations and media strategy. PHI Learning Pvt. Ltd.

[71] Riezebos, R. (2003). Brand management: a theoretical and practical approach. 1st edition. Pearson

[72] Roper, S., & Parker, C. (2006). Evolution of branding theory and its relevance to the independent retail sector. Marketing Review, 6(1).

[73] Schmitt, B. (1999). Experiential marketing. New York, NY: The Free Press.

[74] Schmitt, B. (2009). The concept of brand experience. Journal of Brand Management. Retrieved 2023, April 14 from https://www.researchgate.net/publication/263105183_The _concept_of_brand_experience.

[75] Schmitt, B. H., & Rogers, D. L. (2008). Handbook on brand and experience management. Northampton, MA: Edward Elgar Publishing.

[76] Starcevic, S. (2015). The Origin and Historical Development of Branding and Advertising in the Old Civilizations of Africa, Asia and Europe, Marketing, 46(3), p. 179-196.

[77] The American Marketing Association. (2023). Branding. Retrieved 2023, April 18 from https://www.ama.org/topics/branding/page/41/.

[78] Twede, D. (2002). Commercial amphoras: the earliest consumer packages? Journal of Macromarketing, 22(1), p. 98- 108.

[79] Twede, D. (2002). The packaging technology and science of ancient transport amphoras. Packaging Technology and Science, 15, p. 181—195.

[80] Van Zyl, R. (2011). Branding: looks are important! (In Parker, E., ed. Marketing your own business. Northlands: Pan Macmillan, 61-113.)

[81] Varey, R.J. (2011). A sustainable society logic for marketing. Social Business, 1(1), p. 69-83.

[82] Vargo, S.L., & Lusch, R.F. (2008). Service-dominant logic: continuing the evolution. J. Acad. Marketing Sci. 36 (1), p. 1—10.

[83] Veryzer, R.W., & Hutchinson, J.W. (1998). The influence of unity and prototypicality on aesthetic responses to new product designs. Journal of Consumer Research, 24 (March), p. 374-94.

[84] Warmer, J.P. (2016). Royal branding and the techniques of the body, the self, and power in West Cameroon, in Cultures of Commodity Branding, ed. Andrew Bevan and David Wengrow (New York: Routledge).

[85] Wengrow, D. (2008). Prehistories of commodity branding. Current Anthropology, 49(1), p.7-34.

[86] Witkowski, T. H., & Brian Jones, D.G.B. (2016). Historical research in marketing: literature, knowledge, and disciplinary status. Information & Culture, 51(3), summer 2016. Retrieved 2023, April 17.

[87] Wundt, W. (1896). Grundriss der Psychologie (Outline of Psychology). Leipzig: Engelmann.

[88] Yang, D., Sonmez, M. & Li, Q. (2012). Marks and Brands: Conceptual, Operational and Methodological Comparisons. Journal of Intellectual Property Rights, 17, p.315-323.

Chapter 3

Medical, Social and economic sciences : New developments

Towards Robust Named Entity Recognition to Support the Extraction of Emerging Technological Knowledge from Biomedical Literature

Sabrina Lamberth-Cocca[a], Victoria Dimanova[a], Sebastian Bruchhaus[a], Christian Nawroth[a], Paul Mc Kevitt[b] and Matthias Hemmje[a]

[a]*Faculty of Mathematics and Computer Science, University of Hagen, Germany*
[b]*Academy for International Science & Research (AISR), Derry/Londonderry, Northern Ireland*

Abstract

We present evaluation results of a hybrid transformer-based, domain-specific Named Entity Recognition (NER) model we call *TransformNER-med*. Based on the concept of robust NER, we combine a fine-tuned transformer model, adversarial examples, and an additional knowledge source. The evaluation is the basis for an experimental Information Retrieval System (IRS) prototype to support the recognition of Named Entities (NEs) and emerging Named Entities (eNEs) representing In-Vitro Diagnostic (IVD) technologies in the broader context of medical technology. We also introduce for the first time an expert-validated dataset for terms specifically representing IVD technologies. By using a combination of the fine-tuned transformer model SciBERT [5] and an additional knowledge source to filter out frequently misclassified NEs, we increase the F_1 score of our previous IRS prototype [27] by 27 point absolute improvement to 0.66%. Although the additional use of adversarial examples does not lead to any further improvement, we show potential directions for future research.

Keywords

Biomedical Named Entity Recognition, Information Extraction, Transformer-based Large Language Models, Adversarial Machine Learning, Emerging Medical Technology

1. Introduction and Motivation

This work is part of *RecomRatio (Rationalizing Recommendations)* [1] which is a DFG-funded research project that aims to support medical decisions based on knowledge from biomedical publications by developing clinical Virtual Research Environments (VREs). Medical experts need to retrieve information for various reasons, one of which is the continuous monitoring of the state of the art in science and technology for the purpose of Post-Market Surveillance (PMS). Here, we are interested in the PMS of In-Vitro Diagnostic (IVD) devices after being placed on the market. An obligatory step is *Post-Market Performance Follow-up (PMPF)*, leading to the research-intensive and thus time-consuming task of collecting latest relevant information about IVD technologies. The ever-increasing volume of available publications in subject-specific databases such as PubMed [34] leads to an increasing effort for the medical research community to find suitable information and to evaluate its relevance.

CERC 2023: Collaborative European Research Conference, June 09–10, 2023, Barcelona, Spain
sabrina.lamberth-cocca@fernuni-hagen.de (S. Lamberth-Cocca); victoria@chugunov.de (V. Dimanova); sebastian.bruchhaus@fernuni-hagen.de (S. Bruchhaus); christian.nawroth@fernuni-hagen.de (C. Nawroth); p.mckevitt@aisr.org.uk (P. Mc Kevitt); matthias.hemmje@fernuni-hagen.de (M. Hemmje)
https://orcid.org/0000-0002-8092-5219 (S. Lamberth-Cocca); https://orcid.org/0000-0002-7783-2636 (S. Bruchhaus); https://orcid.org/0000-0001-9715-1590 (P. Mc Kevitt); https://orcid.org/0000-0001-8293-2802 (M. Hemmje)
0000-0002-8092-5219 (S. Lamberth-Cocca); 0000-0002-7783-2636 (S. Bruchhaus); 0000-0001-9715-1590 (P. Mc Kevitt); 0000-0001-8293-2802 (M. Hemmje)

ISBN 978-3-96187-021-9
https://doi.org/10.48444/h_docs-pub-518

Our aim is to provide a machine-based Information Retrieval System (IRS) to support the extraction of relevant and particularly new developments regarding IVD technologies, by identifying emerging knowledge in relevant topic fields such as *gene expression analysis*. IVD devices are defined as, "any device which is a reagent, reagent product, kit, instrument, equipment or system, whether used alone or in combination, intended by the manufacturer to be used in vitro for the examination of samples derived from the human body with a view to providing information on the physiological state, state of health or disease, or congenital abnormality thereof" [9] (p. 169/4).

Our approach supports the identification and extraction of acknowledged and newly arising terms representing technologies from biomedical literature based on Named Entity Recognition (NER) and emerging Named Entity Recognition (eNER) [33]. The work presented in this article extends a previous IR prototype at our research lab, called MedTech-eNER-IRS [27], which combines a learning-based NER model trained from scratch and a rule-based model to identify emerging Named Entities (eNEs). We further refined the NER engine to improve the accuracy metrics, while keeping the eNER component. The contributions of our paper are as follows:

- We introduce an IVD technology vocabulary which has been annotated by a Subject Matter Expert (SME) and can be used for fine-tuning and testing machine learning models for NER in this specific knowledge domain.
- We demonstrate the superiority of pretrained transformer-based Large Language Models (LLMs) after further retraining on domain-specific vocabulary (fine-tuning) and by adding a knowledge source excluding frequently misclassified terms by the domain-specific transformer-based model, in comparison to a domain-specific machine learning model built from scratch.
- We present examples of adversarial learning to further improve robustness of domain-specific NER. Although we did not observe any improvement in the case of our particularly chosen strategy, we show potential directions for future research.
- The validated corpus, code and models are available at: https://github.com/VictoriaDimanova/Robust-medical-NER.

In the following Section 2, we give an overview of the state of the art and related work, followed by a description of our approach and its implementation in Section 3. In Section 4, we give evaluation results of various experiments and conclude with recommendations for future research in Section 5.

2. State of the Art and Related Work

Named Entities (NEs) represent real-world objects for which a designation is used, such as a person, a city, a commercial brand. Named Entity Recognition (NER) is a task in biomedical Natural Language Processing (NLP), also known as biomedical Named Entity Recognition (BioNER). BioNER techniques are used to detect entities representing, e.g., genes and diseases [39]. NER modeling methods can be broken down into four categories: rule-based, dictionary-based, machine learning (ML) based and hybrid [10]. ML-based approaches are currently the most prominent in research, including supervised, semi-supervised and unsupervised learning. Hybrid approaches combine some or all of the other three methods mentioned above [39] to combine the strengths of each; mostly ML with dictionaries [8] or with rules [31], but also approaches exist, which combine dictionaries with rules [47].

In the following subsections, we discuss essential related work regarding our approach, which is hybrid in nature, combining transformer-based ML and methods for robust NER with our existing rule- and dictionary-based eNER component.

2.1. Transformer Models

Since the introduction of the transformer architecture by the Google Brain Team in 2017 [44], a variety of transformer models have been developed. The success of transformer models in many NLP tasks such as e.g., NER, text generation, translation, has led to their strong presence in the research community today, representing the state of the art in NLP, and replacing Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) in many areas.

Transformer models are the basis for Large Pretrained Language Models (LLMs), and have been adapted to various domain-specific NLP tasks, including the biomedical field with more than 40 Biomedical Pretrained Language Models (BPLM) today. The first BPLM was introduced 2019, called BioBERT [28], followed by several other BPLMs, such as ClinicalBERT [3], ClinicalXLNet [15], BlueBERT [38], PubMedBERT [13] or ouBioBERT [45].

Three pretraining methods are common in the biomedical area: Mixed-Domain Pretraining (MDPT), Domain-Specific Pretraining (DSPT) and Task-Adaptive Pretraining (TAPT) [22]. In MDPT, a model is pretrained with both general and domain-specific (biomedical) texts - either continuously (first pretraining with general, then domain-specific texts) or simultaneously (combined corpora with a preponderance of domain-specific versus general texts). PubMedBERT is an example for DSPT with a pretraining on PubMed abstracts and PMC (PubMed Central) full texts [13]. TAPT has the advantage over DSPT and MDPT in that less training data is required for a model to learn task-specific domain and task knowledge, which means that the training is less expensive. Fine-tunig methods are used to further increase a model's domain- and task-specific knowledge, and categorized either as intermediate-task fine-tuning based on large similar training datasets, or multi-task fine-tuning. The latter gains domain- and task-specific knowledge from several tasks, requires less resources which are often scarce in the biomedical field, and is less prone to overfitting [23].

2.2. Robust Named Entity Recognition

This paper focuses on improving the performance of a prototypical software system for eNER that this paper's authors introduce in [27]. An important challenge for ML models is robustness [46]. *Robustness* means an alignment of the model with its stakeholders' expectations when presented with unseen data, i.e., data that are not present in the training set. The NER-model's black box nature makes verification unfeasible. In the case of eNER, *robust recognition* means satisfactory performance. The emergent nature of eNER poses some specific problems, as basically the training data set cannot be a fixed representation of all data. A viable strategy to elude this problem is to augment model fine-tuning with alternative means.

Adversarial examples are a technique that third parties employ to exploit weaknesses of ML models and achieve different outcomes than the model's designers had in mind [6]. Some authors have suggested that hardening models against adversarial examples may also improve the overall robustness against random noise in the data [4]. Several libraries exist that introduce typical attack patterns into text data for training.

Another design pattern that may help with robustness is adding a handcrafted, rule-based AI system to the deep-learning NER-model [42]. This component is used to check system output consistency against a set of constraints and potentially alter it accordingly. This additional layer may or may not be ML-based.

2.3. Medical Technology Foresight and Emerging Named Entity Recognition

The early identification of newly developing research trends and technological concepts is what organizations endeavor to achieve in the context of *technology intelligence* or *technology foresight. Emerging technology* encompasses new technologies as a recombination of existing technologies [7], and also so-called *breakthrough technologies* which, "introduce a radically new capability or a drastic performance improvement" [14] (p. 2).

For information seeking experts in the biomedical field, the machine-supported recognition of novel technological knowledge as early as possible is still an unsolved problem. The challenge with *emerging knowledge* is that it, "arises suddenly and unexpectedly and it cannot be planned and predicted", and it often, "resides in humans in the form of tacit knowledge and explicit knowledge" [37] (p. 425). The definitions underpin the dynamic character of *emergence* in the context of technological knowledge that needs to be addressed and managed by an IR system for medical technology foresight.

Emerging Named Entity Recognition (eNER) is a possible approach used to support the extraction of emerging knowledge. Emerging Named Entities (eNEs) are terms in use that are not listed in controlled vocabularies or databases yet, i.e., terms that are not acknowledged yet [33], [32]. This definition is based on the idea to discriminate entities mentioned in a text corpus and match the time of their occurrence by referencing an existing list of same or similar terms, and relating them to other entities (cf. [25]) or metadata reflecting an emerging research trend, which is characterized as a "topic area that is growing in interest and utility over time" [24] (p. 2).

3. Approach and Implementation

The goal of our approach is to increase robustness of ML models in NER tasks in specific knowledge domains. Pretrained transformer-based Large Language Models (LLMs) represent the core of our hybrid strategy which combines further fine-tuning and additional model components. The work presented here is one step of an incremental improvement of the abovementioned MedTech-eNER-IRS, and focuses on a benchmarking of several LLMs on the specific task and corpus, as well as an evaluation of our approach including the selected models.

3.1. User Perspective

According to the User-Centered Design (UCD) methodology [35], we support the user stereotype *laboratory worker* in searching for relevant information and in taking informed decisions based on state of the art and, in particular, emerging technological knowledge. This main use context is extended to the user stereotype *data scientist* for transformer-based pipeline building and NER/eNER model evaluation, and refined to specific use cases accordingly.

Use cases: We reused the use cases as discussed in [27] with slight modifications and extensions. New NER/eNER use cases are: *Provide validation questionnaire* to support the generation of the SME-validated IVD technology vocabulary by the data scientist, *Train transformer models* to support the training of preselected transformer models by the data scientist, *Select transformer model* to support the recognition of NEs by the laboratory worker based on pretrained transformer models.

User interface: A simple graphical user interface (GUI) is used to run and evaluate the NER/eNER. Users can select a task according to their goal (*NER/eNER* or *NER evaluation*). A window opens and asks the user to select a text file to perform the selected task on. After pushing the *Start* button, model inference runs. Finally, the results are presented visually in the form of the text from the selected file with integrated markings of the detected NEs and eNEs. If the NER evaluation is selected,

Table 1
Data: Corpus creation and characteristics.

Search and Extraction Strategy (Top-Down)	Value
(1) Retrieval of Texts:	
Retrieval Source, Interface, Format	PubMed [34], REST API, XML
Filter 1: Keyword	"Gene expression in vitro"
Filter 2: Time Span	Past 10 years
Filter 3: Source Type	Free full text
Filter 4: Text Type	"Clinical trial" or "randomized controlled trial"
Extracted Paragraphs (Keyword)	"Materials and methods" (cf. [11]) ("method")
Number of Available Publications	220 out of 395
Number of Extracted Text Files	186
(2) Text Annotation:	
Method	Semi-automatic (automatic and manual annotation / check)
Reference	NE vocabulary and rule-based eNE engine from [27]
Annotation Tool	LabelStudio
Data Format	JSON
Number of Annotated NE Candidates	6,092
Number of Annotated eNE Candidates	17
(3) SME-based Data Validation:	
Method	Manual classification (definition of IVD technology)
Domain Expert	Laboratory diagnostics professor
Form	Interactive questionnaire
Answering Options	"No IVD term, but relevant"
	"Not relevant"
	"I don't know / context required"

an additional HTML file is generated with the following information: model name, application of knowledge source, evaluation type (exact or partial match), and Precision[1], Recall[2] and F_1 scores[3].

3.2. Data

To generate a domain-specific vocabulary for model fine-tuning and testing, we created a training corpus and conducted a validation with a Subject Matter Expert (SME). Details are summarized in Table 1.

3.3. Model and Pipeline Implementation

Our approach is to combine a transformer model with adversarial examples and a knowledge source for NER which we call *TransformNER-med*, and cobine it with our existing rule-based eNER component. This includes the following steps:

Transformer model selection: Several transformer models were fine-tuned and evaluated based on precision, recall, and F_1 score after data transformation to *.spacy* format and 10 training epochs on 80 annotated PubMed texts and evaluation on 20 texts in the first round (corpus size 100). The training,

[1] *Precision* is the ratio of correctly predicted positive observations to the total number of predicted positive observations.
[2] *Recall* is the ratio of correctly predicted positive observations to all observations in the class.
[3] F_1 score is the weighted average of Precision and Recall.

evaluation and selection process was iterative and led to the second round of training and evaluation on a corpus size of 150. The following transformer models were shortlisted, with the number in the extension referring to the size of our training corpus: RoBERTa100, PubMedBERT100, SciBERT100, RoBERTa150, PubMedBERT150, and SciBERT150. RoBERTa [29] stands for *Robust Optimized BERT Pretraining Approach* and is a transformer model for general NLP tasks. The model was trained to predict hidden text segments in unannotated language examples. RoBERTa modifies important hyperparameters in the BERT model, was trained with more data than BERT and over a longer time period. RoBERTa has the same architecture as BERT, but uses BPE (Byte-Pair Encoding) as a tokenizer. RoBERTa was trained on Book Corpus [16] and English Wikipedia [21], CC-News [17], openwebtext [18], and STORIES [36]. PubMedBERT [19] [13] is a transformer model which was specifically pretrained from scratch for biomedical NLP tasks, based on PubMed abstracts and PMC full texts. SciBERT [2] [5] is a BERT model trained on scientific biomedical and computer science texts.

Adversarial examples: Adversarial texts were created based on 15 (10%) PubMed texts, which were modified by manipulating words and sentences with TextAttack modules *WordNetAugmenter* and *CharSwapAugmenter*. In each text, 5% each of words were replaced by synonyms and typos were simulated by modifying letters. The adversarial texts then were annotated as described above.

Knowledge source: A vocabulary consisting of tokens from the which are unlikely to appear in IVD terms (misclassified terms from the development set) is defined as our knowledge source. The vocabulary has 4 categories: *common use* (products for general laboratory supply that don't fall under the IVD definition), *statistics* (software products and terms related to statistics), *other domain* (terms from laboratory diagnostics which are not explicit IVD terms, such as in-vivo, or that belong to a another domain), and *organisation* (organisation names or parts thereof misclassified as IVD terms).

Rule-based eNER: This module was reused from our previous prototype as described in [27]. We store the *year of publication* in the vocabulary. After a NE is identified by the transformer model, it is matched against this vocabulary to calculate if it is an NE or an eNE.

Base technologies for implementation: Google Colab [12] was used for text corpus creation and transformer model training. Further technologies applied: Python (version 3.7.13), Natural Language Toolkit (NLTK) library [40], Hugging Face transformer library [20], spaCy [43] (version 3.2.0), TextAttack[41] [30] for adversarial examples, LabelStudio [26] for manual text annotation.

Figure 1 shows the overall architecture of MedTech-eNER-IRS with details regarding components of *TransformNER-med* to increase robustness, as well as interfaces to external databases and tools. MedTech databases store information about medical technologies, such as manufacturer databases or technologies approved by the United States of America (U.S.) Food and Drug Administration (FDA), and are described in-depth in [27].

4. Experiments and Evaluation

We conducted several experiments to test various fine-tuned transformer-based models, both without and with trained adversarial examples and/or use of the knowledge source. As evaluation corpus, we applied the SME-validated IVD technology vocabulary introduced in Section 3.2. and Table 1.

Transformer models: Overall, six transformer models were evaluated (BERT, RoBERTa, BioBERT, PubMedBERT, SciBERT, BioELECTRA) based on the development set of 100 texts, out of which the three with highest F_1 score were selected for further experiments and evaluated again based on 150 texts: RoBERTa (100: 0.59%; 150: 0.58%), PubMedBERT (100: 0.64%; 150: 0.60%), and SciBERT (100: 0.57%; 150: 0.61%). This resulted in a training corpus size selection of 100 texts.

Final evaluation: To determine the most robust NER pipeline, we evaluated the following cases:

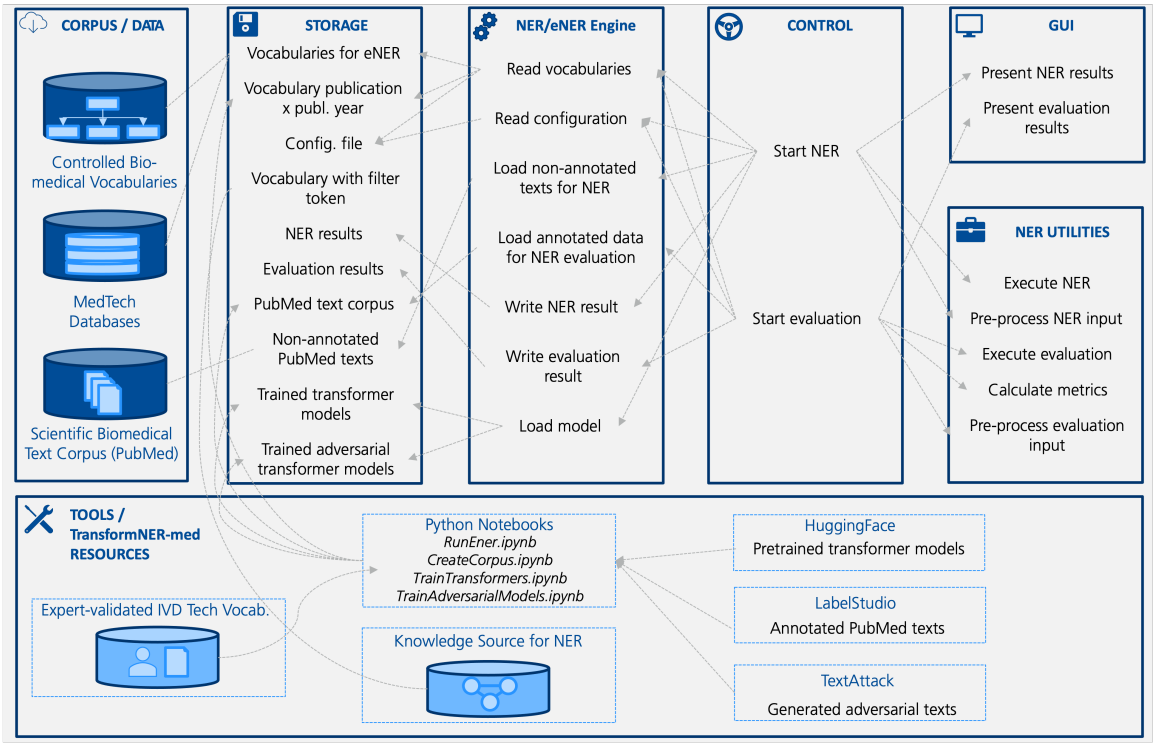


Figure 1: Conceptual MedTech-eNER-IRS architecture with TransformNER-med

Table 2
Evaluation results: Combination of transformer models with adversarial training and knowledge source.

NER Pipeline	Exact Match			Partial Match		
Components	Precision	Recall	F1 score	Precision	Recall	F1 score
RoBERTaAdv + knowledge source	0.52	0.43	0.47	0.68	0.54	0.60
PubMedBERTadv + knowledge source	0.50	**0.56**	**0.53**	0.61	**0.67**	0.64
SciBERTadv + knowledge source	**0.55**	0.51	**0.53**	**0.70**	0.63	**0.66**

(1) Transformer model, (2) Transformer model + adversarial examples, (3) Transformer model + knowledge source, (4) Transformer model + adversarial examples + knowledge source. The use of adversarial examples (adv) led to an increase in performance of the models, with PubMedBERT as an exception: PubMedBERTadv achieved the highest F_1 score of 64% of all three transformer models trained with adversarial examples, but with the same F_1 score as PubMedBERT100.

Regarding combinations of transformer models with adversarial training and knowledge source, SciBERT showed the best result with a partial match F_1 score of 0.66% (SciBERTadv + knowledge source, see Table 2). The use of the knowledge source in combination with fine-tuned transformer models (see Table 3) had the strongest effect on performance improvement. Based on our experiments, we overall recommend the following NER pipelines based on F_1 scores and *partial match* evaluation: PubMedBERT100, PubMedBERT100 + knowledge source, SciBERT150 + knowledge source, and SciBERTadv + knowledge source. With a partial match F_1 score of 0.66%, the following pipelines showed the best possible performance in our experiments: *SciBERT150 + knowledge source*, and *SciBERTadv + knowledge source*.

Table 3
Evaluation results: Combination of transformer models with knowledge source.

NER Pipeline	Exact Match			Partial Match		
Components	Precision	Recall	F1 score	Precision	Recall	F1 score
RoBERTa100 + knowledge source	0.55	0.55	0.55	0.63	0.67	0.65
RoBERTa150 + knowledge source	0.58	0.49	0.53	0.68	0.59	0.63
PubMedBERT100 + knowledge source	**0.62**	0.51	**0.56**	**0.72**	0.60	0.65
PubMedBERT150 + knowledge source	0.57	0.45	0.50	0.71	0.55	0.62
SciBERT100 + knowledge source	0.60	0.48	0.53	0.71	0.56	0.63
SciBERT150 + knowledge source	0.53	**0.56**	0.54	0.65	**0.67**	**0.66**

5. Conclusion and Future Work

In this paper we have evaluated the use of a hybrid machine learning model for NER in the specific domain of IVD technologies, consisting of pretrained, further fine-tuned transformer models, adversarial examples and a knowledge source. We discussed the underlying concepts aiming at increasing the robustness of a domain-specific NER model, including a benchmarking of various LLMs. However, there are a few remaining challenges. In order to further increase the *transformer-based model's* performance, we suggest two strategies: (1) Use more data for fine-tuning, from sources other than PubMed, and (2) Further analysis of tokens that are frequently misclassified by transformer models.

Regarding the use of *adversarial examples*, more experiments need to be conducted, in order to leverage the potential to increase model performance by mitigating overfitting. The use of a *knowledge source* that contains frequently misclassified NEs has proven to be an effective strategy to support the overall model's performance. This method could be further expanded by adding more cases to be excluded to the knowledge source.

We conclude that the NER task in the specific IVD domain alone is difficult and needs further research. The *eNE* component adds further challenges on top, mainly due to too few eNE candidates in our corpus. More IVD terms need to be gathered, in order to increase the scarce data basis. This could be achieved by adding data from sources such as blogs, calls for papers by conferences and publishers, whitepapers or papers explicitly referring to *emerging technology* in the relevant topic field. Furthermore, these data could be combined with metadata to determine, if they are emerging, such as the according entry and editing history in knowledge bases or increasing number of search queries by year reflecting an *emerging trend* (see definition in Section 2.3). Another method for consideration is the creation of synthetic texts through replacement of NEs by eNEs.

References

[1] Rationalizing recommendations (recomratio), university of bielefeld (2015), http://ratio.sc.cit-ec.uni-bielefeld.de/de/projekte/ratiorec/
[2] allenai: scibert, https://github.com/allenai/scibert
[3] Alsentzer, E., Murphy, J.R., Boag, W., Weng, W.H., Jin, D., Naumann, T., McDermott, M.: Publicly available clinical bert embeddings. arXiv preprint arXiv:1904.03323 (2019)
[4] Bělohlávek, P., Plátek, O., Žabokrtský, Z., Straka, M.: Using adversarial examples in natural language processing. In: Proceedings of the Eleventh International Conference on Language

Resources and Evaluation (LREC 2018). European Language Resources Association (ELRA), Miyazaki, Japan (May 2018), https://aclanthology.org/L18-1584
[5] Beltagy, I., Lo, K., Cohan, A.: Scibert: A pretrained language model for scientific text. arXiv preprint arXiv:1903.10676 (2019)
[6] Biggio, B., Roli, F.: Wild patterns: Ten years after the rise of adversarial machine learning. Pattern Recognition **84**, 317–331 (2018). https://doi.org/https://doi.org/10.1016/j.patcog.2018.07.023, https://www.sciencedirect.com/science/article/pii/S0031320318302565
[7] Bruck, P., Réthy, I., Szente, J., Tobochnik, J., Érdi, P.: Recognition of emerging technology trends: class-selective study of citations in the us patent citation network. Scientometrics **107**(3), 1465–1475 (2016)
[8] Campos, D., Matos, S., Oliveira, J.L.: Gimli: open source and high-performance biomedical name recognition. BMC bioinformatics **14**(1), 1–14 (2013)
[9] Council, E.: Council directive 93/42/eec concerning medical devices. Official Journal of The European Communities, Luxembourg (1993)
[10] Eltyeb, S., Salim, N.: Chemical named entities recognition: a review on approaches and applications. Journal of cheminformatics **6**, 1–12 (2014)
[11] Ghasemi, A., Bahadoran, Z., Zadeh-Vakili, A., Montazeri, S.A., Hosseinpanah, F.: The principles of biomedical scientific writing: Materials and methods. International Journal of Endocrinology and Metabolism **17**(1) (2019)
[12] Google: Colaboratory, https://colab.research.google.com
[13] Gu, Y., Tinn, R., Cheng, H., Lucas, M., Usuyama, N., Liu, X., Naumann, T., Gao, J., Poon, H.: Domain-specific language model pretraining for biomedical natural language processing. ACM Transactions on Computing for Healthcare (HEALTH) **3**(1), 1–23 (2021)
[14] Hein, A.M., Brun, J.: A conceptual framework for breakthrough technologies. In: Proceedings of the Design Society: International Conference on Engineering Design. vol. 1, pp. 1333–1342. Cambridge University Press (2019)
[15] Huang, K., Singh, A., Chen, S., Moseley, E.T., Deng, C.y., George, N., Lindvall, C.: Clinical xlnet: modeling sequential clinical notes and predicting prolonged mechanical ventilation. arXiv preprint arXiv:1912.11975 (2019)
[16] HuggingFace: Datasets: bookcorpus, https://huggingface.co/datasets/bookcorpus
[17] HuggingFace: Datasets: cc-news, https://huggingface.co/datasets/cc-news
[18] HuggingFace: Datasets: openwebtext, https://huggingface.co/datasets/openwebtext
[19] HuggingFace: microsoft / biomednlp-pubmedbert-base-uncased-abstract-fulltext, https://huggingface.co/microsoft/BiomedNLP-PubMedBERT-base-uncased-abstract-fulltext/tree/main
[20] huggingface: transformers, https://github.com/huggingface/transformers
[21] HuggingFace: Datasets: Wikipedia (2023), https://huggingface.co/datasets/wikipedia
[22] Kalyan, K.S., Rajasekharan, A., Sangeetha, S.: Ammu: a survey of transformer-based biomedical pretrained language models. Journal of biomedical informatics p. 103982 (2021)
[23] Khan, M.R., Ziyadi, M., AbdelHady, M.: Mt-bioner: Multi-task learning for biomedical named entity recognition using deep bidirectional transformers. arXiv preprint arXiv:2001.08904 (2020)
[24] Kontostathis, A., Galitsky, L.M., Pottenger, W.M., Roy, S., Phelps, D.J.: A survey of emerging trend detection in textual data mining. Survey of text mining pp. 185–224 (2004)
[25] Krenn, M., Zeilinger, A.: Predicting research trends with semantic and neural networks with an application in quantum physics. Proceedings of the National Academy of Sciences **117**(4), 1910–1916 (2020)
[26] LabelStudio: Open source open-source data labeling platform, https://labelstud.io
[27] Lamberth-Cocca., S., Maier., B., Nawroth., C., Mc Kevitt., P., Hemmje., M.: Named entity recog-

nition for the extraction of emerging technological knowledge from medical literature. In: Proceedings of the 14th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management - KEOD,. pp. 101–108. INSTICC, SciTePress (2022)
[28] Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C.H., Kang, J.: Biobert: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics **36**(4), 1234–1240 (2020)
[29] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692 (2019)
[30] Morris, J.X., Lifland, E., Yoo, J.Y., Grigsby, J., Jin, D., Qi, Y.: Textattack: A framework for adversarial attacks, data augmentation, and adversarial training in nlp. arXiv preprint arXiv:2005.05909 (2020)
[31] Naderi, N., Kappler, T., Baker, C.J., Witte, R.: Organismtagger: detection, normalization and grounding of organism entities in biomedical documents. Bioinformatics **27**(19), 2721–2729 (2011)
[32] Nawroth, C., Engel, F., Eljasik-Swoboda, T., Hemmje, M.L.: Towards enabling emerging named entity recognition as a clinical information and argumentation support. In: DATA. pp. 47–55 (2018)
[33] Nawroth, C., Hemmje, I.M., Mc Kevitt, P.: Supporting Information Retrieval of Emerging Knowledge and Argumentation. Ph.D. thesis, University of Hagen, Germany (2021)
[34] NIH: Pubmed (2023), https://pubmed.ncbi.nlm.nih.gov
[35] Norman, D.A., Draper, S.W.: User centered system design: New perspectives on human-computer interaction (1986)
[36] Paperswithcode: Cc-stories, https://paperswithcode.com/dataset/cc-stories
[37] Patel, N.V., Ghoneim, A.: Managing emergent knowledge through deferred action design principles. Journal of Enterprise Information Management (2011)
[38] Peng, Y., Yan, S., Lu, Z.: Transfer learning in biomedical natural language processing: an evaluation of bert and elmo on ten benchmarking datasets. arXiv preprint arXiv:1906.05474 (2019)
[39] Perera, N., Dehmer, M., Emmert-Streib, F.: Named entity recognition and relation detection for biomedical information extraction. Frontiers in cell and developmental biology p. 673 (2020)
[40] Project, N.: Nltk natural language toolkit documentation, https://www.nltk.org
[41] QData: Textattack, https://github.com/QData/TextAttack
[42] Song, Y., Wang, T., Cai, P., Mondal, S.K., Sahoo, J.P.: A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities. ACM Comput. Surv. (feb 2023). https://doi.org/10.1145/3582688, https://doi.org/10.1145/3582688
[43] spaCy: Industrial-strength natural language processing in python, https://spacy.io
[44] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
[45] Wada, S., Takeda, T., Manabe, S., Konishi, S., Kamohara, J., Matsumura, Y.: Pre-training technique to localize medical bert and enhance biomedical bert. arXiv preprint arXiv:2005.07202 (2020)
[46] Wang, X., Wang, H., Yang, D.: Measure and improve robustness in NLP models: A survey. In: Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. pp. 4569–4586. Association for Computational Linguistics, Seattle, United States (Jul 2022). https://doi.org/10.18653/v1/2022.naacl-main.339, https://aclanthology.org/2022.naacl-main.339
[47] Wei, C.H., Kao, H.Y., Lu, Z.: Sr4gn: a species recognition software tool for gene normalization. PloS one **7**(6), e38460 (2012)

SenseCare KM-EP: Combining Gaming and Affective Computing for Improved rehabilitation Outcomes

Hayette Hadjar [1], Binh Vu [2], Paul Mc Kevitt[3] and Matthias Hemmje[4]

[1] *University of Hagen, Faculty of Mathematics and Computer Science, Hagen, Germany*
[2] *SRH University Heidelberg, Heidelberg, Germany*
[3] *Ulster University, Derry/Londonderry, Northern Ireland*
[4] *Research Institute for Telecommunication and Cooperation, Dortmund, Germany*

Abstract

Tele-rehabilitation is a new approach to medical care that has proven promising to meet the demand for accessible and engaging rehabilitation programs for people with a variety of physical disorders. This paper relates to the Sensor Enabled Affective Computing for Enhancing Medical Care (SenseCare) project, along with the SenseCare KM-EP (Knowledge Management-Ecosystem Portal) Affective Computing (AC) platform. The SenseCare KM-EP for of home tele-rehabilitation platform uses games-based web immersive technology. In our experiments, we used a leap motion controller device to interact with the game via hand movement, as well as audiovisual devices to track patient emotions during gameplay. Our proposed prototype combines interactive and web immersive game-based rehabilitation with AC to create a comprehensive system capable of remotely monitoring and supporting patients with hyper-limb rehabilitation needs. We discuss the design and implementation of the prototype, including initial results. This work demonstrates the potential for using technology to enhance traditional rehabilitation practices and improve patient outcomes, and also to be an effective and enjoyable option for people who need rehabilitation. Personalized exercise programs combined with online monitoring tools can help patients perform better and offer caregivers important support. The paper concludes with a discussion of the potential for further research and development in this area.

Keywords

SenseCare KM-EP, Hyper-Limb Rehabilitation, Affective Computing, WebXR Game, Leap Motion Controller

1. Introduction

After a stroke, patients frequently experience arm and hand weakness, but movement can be restored with plenty of exercise [1]. Arm movements may need to be performed up to 2500 times to achieve a person's maximum degree of motor function [2], and gaming systems with controllers and video monitoring are both accessible, affordable options that can be employed in rehabilitation [2].

Affective Computing (AC) is a rapidly expanding interdisciplinary area that studies how feelings influence human-technology relations [3]. It investigates methods for detecting and creating emotions, as well as how they affect system design, execution, and assessment. Emotions play a significant role in how users and games communicate. Developing engaging stories with new and more widely accepted technologies is a challenge for designers working in Extended Reality (XR) environments [4].

CERC 2023: Collaborative European Research Conference, June, 9th 2023, Barcelona, Spain
hayette.hadjar@fernuni-hagen (H. Hadjar); binh.vu@srh.de (B.Vu); p.mckevitt@ulster.ac.uk (P. Mc Kevitt); mhemmje@ftk.de (M.Hemmje);
https:// / (H.Hadjar); http:// (B.Vu)
0000-0001-9540-6473 (H.Hadjar); 0000-0001-8567-1193 (B.Vu); 0000-0001-9715-1590 (P. Mc Kevitt), 0000-0001-8293-2802 (M.Hemmje)

ISBN 978-3-96187-021-9 https://doi.org/10.48444/h_docs-pub-518

Assessing the patient's emotional state could help to define and control the parameters with some cognitive tasks while training. Emotion detection and caregiver input undoubtedly play a vital role in monitoring [5] [6] [7], calibrating, and changing tasks in learning processes.

There is an increasing body of work showing the advantages of gaming and Virtual Reality (VR) apps in the therapy of Neurodegenerative (ND) diseases [8] [9]. Numerous research has focused on the benefits of gamification in increasing motivation for a specific task or its impact on breaking a habit by presenting interesting challenges that are adapted to the user's status [10][11][12].

For remote home healthcare uses, this paper contributes to the SenseCare project [13]. SenseCare was a 48-month project that the European Union funded in 2016 [14], with the FTK Research Center for Tele-communication and Cooperation as partner [15]. The SenseCare platform's core system is referred to as the Knowledge Management Ecosystem Portal (KM-EP) [32].

Given how quickly and frequently human emotions change, it can be challenging to recognize human emotions from audiovisual sensors. As a result, machine learning techniques are frequently used to organize and detect emotions. In the SenseCare KM-EP, three analytical techniques are now available: (1) Support Vector Machines (SVMs) [16], (2) Artificial Neural Networks (ANNs) [17], and (3) Convolutional Neural Networks (CNNs) using TensorFlow [35]. But nevertheless, all of the current studies only allow for the detection of emotions via images, without audio or other sensors.

The SenseCare KP-EP does not include a gaming application with AC. Detecting emotions in gaming is challenging, and there is no research that combines AC to support emotional detection in gaming in the SenseCare KM-EP. Additionally, The SenseCare KM-EP has not been used to conduct experiments providing a gaming environment, and integrating games into the system is a complex task.

Based on our motivation and problem statement, we've developed four research questions. The first query relates to upgrading the SenseCare KM-EP to enable emotion detection through various sensors. The second question examines building a SenseCare KM-EP that supports gameplay, and the third question investigates how gaming platforms can be used for tele-rehabilitation. And finally, the fourth question concerns integrating gaming technologies, such as emotion recognition, with tele-rehabilitation uses.

Our research objectives include developing and implementing a conceptual model and exploring methods for using gaming with AC for tele-rehabilitation, which includes integrating it on the SenseCare KM-EP for physiotherapy apps. We also aim to investigate how to integrate gameplay for tele-rehabilitation with AC in the SenseCare KM-EP. After this, we plan to assess the system's combination with the SenseCare KM-EP, which combines gaming and AC for tele-rehabilitation.

In our previous paper [18], we investigated how VR and Augmented Reality (AR) games are hosted on web browsers, how the SenseCare KM-EP recognizes a patient's emotions during gaming, and how human movements are recognized. In this paper, we expand on our previous research on remote rehabilitation using immersive web games by explaining its UML design. In addition, we investigate the incorporation of a leap motion controller in our WebXR game experiments for hyper-limb rehabilitation situations. Our study also includes monitoring patients' emotional states using audiovisual sensors.

The remainder of this paper is structured as follows. Section 2 explores related work on gaming for rehabilitation studies, and the use of XR (VR, AR, Mixed Reality (MR)) in the therapy of degenerative diseases. Section 3 details the UML diagrams of the proposed system. Section 4 discusses its implementation and demonstration. Section 5 discusses experimental results and implications. And finally, Section 6 concludes the paper and discusses future work.

2. Related Work

SenseCare aims to improve medical treatment by developing a range of input platforms for different sensory devices, such as cameras and Internet of Things wearable sensors. The SenseCare web platform already includes a number of emotion detection analysis techniques [16]. The conceptual architecture of the SenseCare KM-EP AC platform based on audiovisual emotion recognition is detailed in [19]. The SenseCare KM-EP is a web-based application that integrates AR and Virtual activities for home rehabilitation [18]. It uses audiovisual sensors to monitor patients' emotional states and body movements during gaming, providing therapists with immediate input. The platform contains VR and

AR games that use the WebXR API. Following an online survey, 81.8% of consumers were satisfied with the SenseCare KM-EP.

[20] discusses a survey of serious games for tele-rehabilitation of upper arm stroke patients. It investigates gaming for tele-rehabilitation, including the advantages of game-based treatments for stroke rehabilitation, such as improved motivation and efficacy. According to the authors, more studies are required in areas such as game design and evaluation, customized game development, and interaction with tele-rehabilitation systems. They predict that game-based tele-rehabilitation can enhance stroke rehabilitation outcomes.

The use of the Leap Motion Device for stroke therapy was investigated by Khademi et al. [21]. They developed a game-like exercise prototype by tracking hand movements with the device. Experiments included 12 stroke patients who improved their hand function after four weeks of exercising. The authors believe the device could be helpful for stroke rehabilitation, but more study is required to corroborate this and investigate its possibilities for home-based rehabilitation. Yang et al [22] used the Leap Motion Controller and a two-layer bidirectional recurrent neural network to create a system that detects dynamic hand motions in real-time. The algorithm obtained promising accuracy rates on a dataset of hand gestures and could be used in VR and human-computer interaction. It senses motions and is accurate to 0.01 mm [23].

3. Conceptual Design

The conceptual design of the proposed system was detailed in our previous work [18].

In this section, we will look at how UML [24] is used to design our software. The system is designed to satisfy the requirements of a specified application domain. We explain the system's main components and interactions, as well as providing UML class and interaction diagrams to demonstrate its structure and behavior. Our UML diagrams provide a picture of the system's architecture and are beneficial to developers and stakeholders engaged in its development.

UML classes: *Patient*, *Record*, *Caregiver*, *Game*, *Feedback*, *Emotions monitoring*, and *Tele-Rehab-Session* are represented in Table 1. Figure 1 illustrates the UML class diagram for the SenseCare KM-EP.

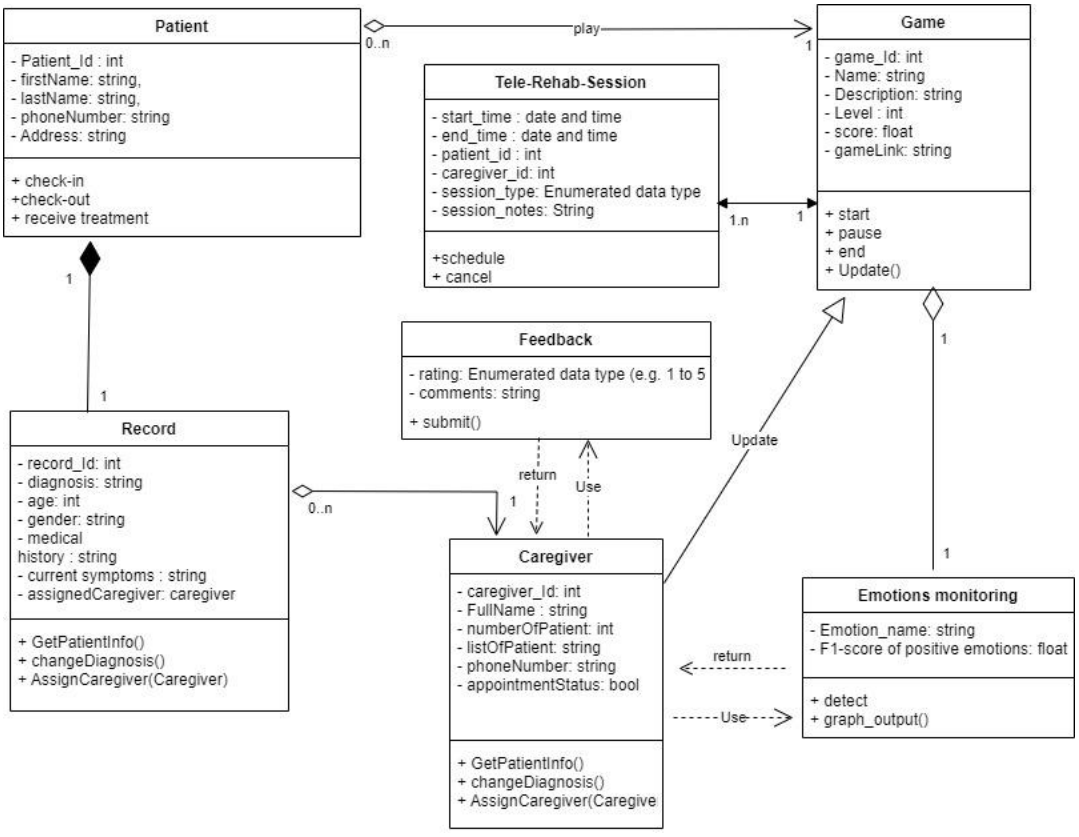


Figure 1: UML class diagram for the SenseCare KM-EP tele-rehabilitation application

Table 1
SenceCare KM-EP UML classes

Class	Attributes	Methods
Patient	Patient_Id: int , firstName : string, last-Name: string, phoneNumber: string, Address : string.	check-in, check-out, receive treatment.
Record	record_Id: int, diagnosis: string, age: int, gender: string, medical history: string, current symptoms: string, assignedCare-giver: caregiver.	GetPatientInfo(), changeDiagnosis(), AssignCare-giver(Caregiver)
Caregiver	caregiver_Id: int , FullName : string, numberOfPatient: int, listOfPatient: string, phoneNumber: string, appointmentStatus: bool.	Getinfo(), ChangeNumberOfPatient(), AddNewPatient(Patient), GetAllPatients()
Game	game_Id: int, Name: string , Description: string, level : int, score: float, gameLink: string.	start, pause, end, Update()
Feedback	rating: Enumerated data type (e.g. 1 to 5 rating scale), comments: string	Submit
Emotions monitoring	Emotion_name: string, F1-score of positive emotions: float.	detect, graph_output()
Tele-Rehab-Session	start_time : date and time , end_time : date and time , patient_id: int, caregiver_id: int, session_type: Enumerated data type (e.g. physical therapy), session_notes: String, emotions_outcome: string	schedule, cancel

In Figure 1 the *Patient* class has an association relationship with the *Tele-Rehab-Session* class, which means that a *Patient* can have multiple *Tele-Rehab-Session* and a *Tele-Rehab-Session* is related to one *Patient*. The *Feedback* class has an aggregation relationship with the *Emotions monitoring* class, which means that the *Feedback* class can have one or more *Emotions monitoring* and an *Emotions monitoring* can be part of multiple *Feedback*. The *Emotion monitoring* class has a composition relationship with the *Tele-Rehab-Session* class, which means that the *Tele-Rehab-Session* class is responsible for creating and destroying the *Emotions monitoring*, and the *Emotions monitoring* cannot exist without the *Tele-Rehab-Session* class.

The following elements in the UML design provide a visual roadmap for the implementation and development of the rehabilitation system by providing a structured representation of the relationships and interactions within the platform.

Game and *Tele-Rehab-Session*: A *Tele-Rehab-Session* is initiated when a *Patient* plays a *Game*. The *Game* is part of the *Tele-Rehab-Session* and the *Tele-Rehab-Session* tracks the progress of the patient's game play.

Patient and *Tele-Rehab-Session*: A *patient* is associated with a *Tele-Rehab-Session*, and the *Tele-Rehab-Session* is associated with a particular *patient*. The *patient*'s plays *Game* and the *Emotions monitoring* is done during the *Tele-Rehab-Session*.

Feedback and *Tele-Rehab-Session*: *Feedback* is generated during a *Tele-Rehab-Session*, based on the *Patient*'s comments and rating. The *Caregiver* uses this *Feedback* to update the *Game*.

Emotions Monitoring and *Tele-Rehab-Session*: *Emotions monitoring* is an integral part of a *Tele-Rehab-Session*, and it provides information about the *Patient*'s emotional state throughout the *Tele-Rehab-Session*.

Record and *Caregiver*: The *Caregiver* is responsible for recording the *Patient*'s information, including their emotional state and progress in the *Game*. The *Record* of the *Patient* is the database that stores this information for future reference.

Caregiver and *Game*: The *Caregiver* is responsible for overseeing the *Tele-Rehab-Session*, and updating the *Game* after showing the result of monitoring the *Patient*'s emotional state and progress in the *Game*. The *Patient* can also submit written *Feedback* to the *Caregiver* based on their observations during *Game* play.

The employment of UML design plays a part in capturing information flow, and relationships between system elements, and assessing the effect of interactions between users on the system's behavior as a whole.

4. Implementation

Figure 2 depicts a concept architecture overview of the SenseCare KM-EP for hyper-limb rehabilitation.

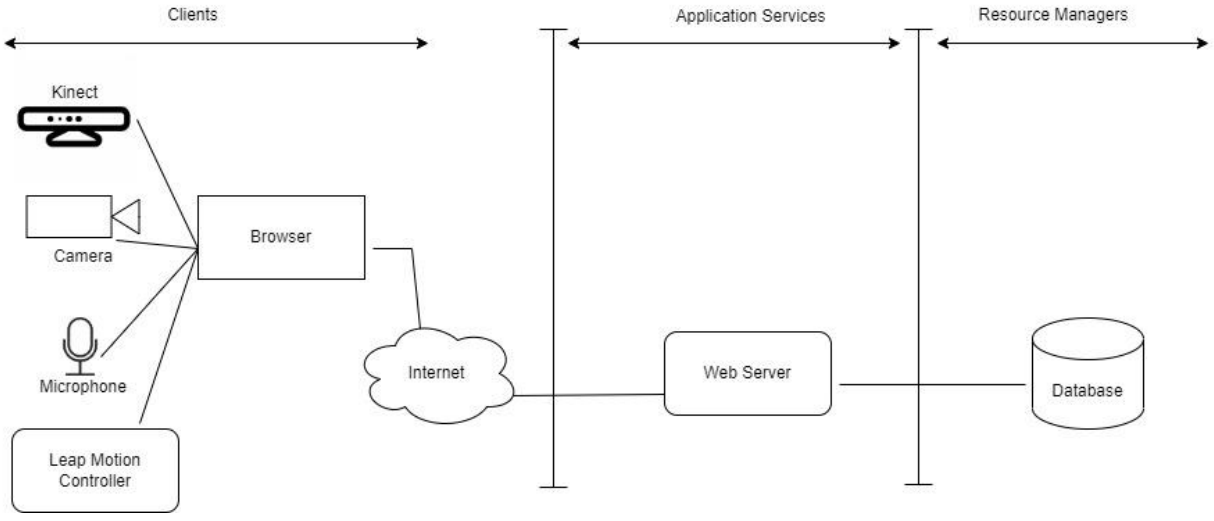


Figure 2: SenseCare KM-EP Concept Architecture

The client accesses the application's interface for the SenseCare KM-EP through his/her web browser. The hand movements and interaction with the game are enabled by a Leap Motion device. In addition, emotion monitoring of the client is achieved using a Kinect camera and a microphone. The web server is implemented using NodeJS, and data storage is managed by MongoDB.

We integrated the Leap Motion controller [25] into our design because it allows for hand tracking in 3D web settings when used in conjunction with Three.js and WebGL. Three.js is a JavaScript tool that allows developers to produce WebGL-based 3D animations and images for web devices. It provides a straightforward API for creating, changing, and displaying 3D models and situations.

Infrared cameras and LED lights on the Leap Motion device work with Three.js to monitor hand and finger motions in three dimensions. The 3D environment produced by Three.js is then managed using this information. Users can move and transform 3D items using their hands or motions to start animations or events, for instance.

Leap Motion device [1] integration into our design enables hand tracking in 3D online settings when used in conjunction with Three.js and WebGL. A JavaScript tool called Three.js allows developers to produce WebGL-based 3D animations and images for web devices. It provides a straightforward API for creating, altering, and displaying 3D items and situations. In order to link to the Leap Motion controller and retrieve info from it, the LeapJS library [26] [27] is used. This library allows for the tracking of hand and finger motions and recognize gestures.

4.1. Rehabilitation Games

We opted for a game-based web approach for two reasons. Firstly, browser games, typically developed with HTML and JavaScript, offer a widely recognized and accessible platform. Secondly, these games are popular due to their ease of access, as they do not require any downloads or installations, and they generally have low hardware requirement.

Figure 3 displays game-based web applications that use the Leap Motion controller for interaction, which have been obtained from the GitHub repository. Figure 3 (a): In this game the player manipulates the camera and objects within the THREE.js scene using the leap motion controller. It supports actions such as rotation, zooming/scaling, and panning, with customizable gestures. Figure 3 (b): is a 3D game created with CSS and HTML. It uses Leap Motion devices and LeapJS to control a jet plane. CSS3 3D transforms are used for the city and objects. JavaScript starts the game. The buildings, plane, and other 3D elements are HTML elements transformed with CSS3. CSS transitions handle the day-night rotation. CSS has limitations for complex 3D games, including performance issues, browser support problems, and graphics constraints. Figure 3 (c): this game is a straightforward browser-based space game specifically designed for use with the Leap Motion controller. It allows users to play a simple space game by simply waving their finger in the air. Figure 3 (d): this VARTIST game [28] delivers an effect that closely resembles CSS 3D transforms or the THREE.js CSS3DRenderer, in this game, hand tracking with Leap Motion offers intuitive controls. Players may sketch on the canvas by pinching their fingers, click or choose by moving their hand up, control items with fist gestures, and dismiss the menu by pushing their hand forward. These interactions improve immersion and the user experience.

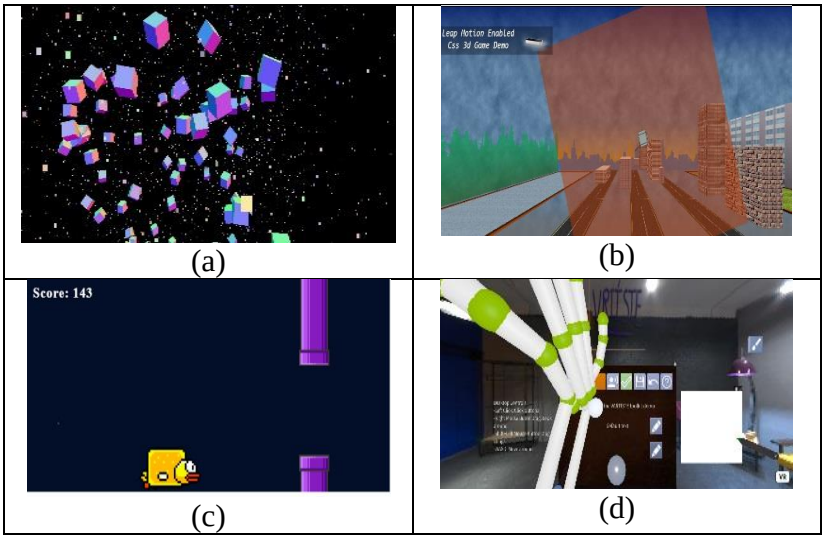


Figure 3: Examples of games-based web with a leap motion controller

Additional examples will be created, taking into consideration specific patient criteria and clinician advice, to be utilized in experiments involving real patients.

XR experiences can be built, and accessed through a web browser rather than by installing an additional app (see Figure 3). Leap Motion enables more organic and intuitive activities in VR and AR settings when used with WebXR. It can be used to move through virtual environments, operate virtual interfaces and objects, and conduct out activities while playing the game. This can increase the feeling of presence and total immersion in the virtual environment, making the experience more captivating and realistic.

The web game interface gives access to input and output audiovisual sensors available to the browser including camera for facial emotion recognition and contactless vital signs, and microphone for speech emotion recognition, for monitoring the patient during gameplay. Our previous study [18] provided a detailed account of the methods used to detect emotions from facial expressions and speech, as well as techniques for monitoring body gestures and pulse through cameras.

4.2. Emotion Score

The Weighted Average approach used to calculate global sensation from audiovisual sensors. The weighted average formula for calculating an overall emotion score based on input from the face, voice, emotional gestures, and vital signs is as follows:

Overall Emotion Score =
(Weight_Face * Emotion_Face) +
(Weight_Speech * Emotion_Speech) +
(Weight_Gestures * Emotion_Gestures) +
(Weight_Gestures * Emotion_Gestures) +
(VitalSigns of Weight and Emotion)

Individual emotions (Emotion_Face, Emotion_Speech, Emotion_Gestures, Emotion_VitalSigns) are multiplied by their weights (Weight_Face, Weight_Speech, Weight_Gestures, Weight_VitalSigns) in this formula. The weighted ratings are then added together to determine the total emotion score.

Calculating an Overall Emotion Score during game sessions (e.g. from time 0 to time n) involves using a weighted average formula to calculate the emotion score for each time interval, and then combining them to obtain a single score for the entire rehabilitation session. Here's the aggregation technique based on the significance of each time interval:

Overall Emotion Score =
(Overall Emotion Score_0 + Overall Emotion Score_1 +
... + Overall Emotion Score_n) / (n + 1)

Here, (n + 1) (n + 1) denotes the total count of time intervals considered throughout the duration of game play monitoring where: Emotion_Score_i represents the emotion score at time point i, and Weight_i represents the weight assigned to that time point. The resulting value represents the calculated global sensation score. The interpretation of this score will depend on the specific scale. For instance, if we use scale from 1 to 10, a score of 8.01 might be interpreted as a moderately positive sensation.

5. Experiments

To conduct experiments with real patients, we will choose patients who require specific hyper limb treatment, such as hand rehabilitation after a fracture. We will collaborate with medical experts to design new games or update existing ones that incorporate diagnostic criteria to address the patients' needs. Once the games are developed, they will be integrated into the platform. Finally, we will create a session plan tailored to each patient's requirements.

Initially, a small group of 6 players used the leap motion controller device to interact with games (Figure 3(a), (b), (c), & (d)) and move their limbs while playing.

During experimentation, we opted to utilize the Leap Motion device for hyper limb rehabilitation in order to monitor more effectively the player's movements during gaming (see Table 2).

Table 2
Affective monitoring / Device interaction

Audio visual monitoring Emotion recognition	Headsets with VR controllers	Leap motion controller
Facial	✗	✓
Speech	✓	✓
Body	✗	✓
Contactless vital sign	✗	✓

Correctly identifying emotions from audiovisual sensors can be challenging. Furthermore, the technology poses social issues such as privacy and accuracy.

A variety of methods can be used to assess the effectiveness of a game for tele-rehabilitation such as questionnaires. A questionnaire can provide a quantitative assessment of the user's experience and satisfaction with the game, as well as feedback on its usability and efficacy as a therapy tool.

In order to calculate the total patient satisfaction, a rating system with a line of stars in CSS for each game was added to the patient's dashboard. We created a weighted patient satisfaction score as follows by dividing the patient rating by the weight:

Weight_Rating * Patient_Rating = Patient_Satisfaction_Score
(e.g., Weight_Rating = 10 stars)

To measure total patient satisfaction, we employ a weighted average method that combines the emotion ratings with the patient satisfaction score. The calculation is as follows:

Total_Patient_Satisfaction =
(Emotion_Score + Patient_Satisfaction_Score) /
(Weight_Face + Weight_Speech + Weight_Gestures + Weight_VitalSigns + Weight_Rating)

Emotion_Score reflects the aggregate emotion ratings from multiple sensors, and Patient_Satisfaction_Score is the patient's evaluation weighted by its appropriate weight for each game session. The denominator comprises the weights allocated to each element, ensuring a fair assessment of total patient satisfaction.

The use of leap motion in a browser is still in its early phases, but there are a few experimental tools and frameworks that allow developers to use Leap Motion with WebXR, such as A-Frame [29] and three.js. It should be mentioned that support for this technology is still limited, and it may not work on all browsers.

Other challenges exist such as a difficulty in finding suitable patients for experimentation. In order to collect text input from patients and deliver it to caregivers, a player-fillable form was included into the game's Frontend in our prototype. Additionally, we collected emotional information from connected audiovisual sensors. As a result, there will be two distinct interpretations of the input provided by patients. Finally, interpreting the game and feedback outcomes necessitates detailed analysis of both the qualitative and quantitative data gathered from the game UI and audiovisual sensors. It is possible to make wise choices about how to enhance the game by developing a strong knowledge of user satisfaction and emotional states.

6. Conclusion and future work

This paper presents a novel approach to upper limb remote rehabilitation that combines AC therapy with gaming. To support this approach, SenseCare KM-EP UML classes were used. The Leap Motion controller was used in our experiment to enable user input from game-playing patients. To monitor the

patients' emotional responses as they played the game, audiovisual sensors were also used. Furthermore, formulas have been derived to assess the overall satisfaction of the patients undergoing treatment using this new methodology. However, despite the promising opportunities, additional research must be conducted and other challenges must be overcome before the full potential of the proposed strategy can be realized.

Several future plans are in place for this project, including: (1) conducting further experimentation with a larger patient population, (2) refining the data analysis techniques to gain a better understanding of the results, (3) investigating alternative sensors for detecting and monitoring emotions, (4) exploring ways to optimize the WebXR platform for enhanced usability and engagement (e.g., Avatar therapist), and (5) incorporating an option for monitoring respiration rate during gameplay.

7. References

[1] C. J. Winstein *et al.*, "Guidelines for Adult Stroke Rehabilitation and Recovery: A Guideline for Healthcare Professionals from the American Heart Association/American Stroke Association," *Stroke*, vol. 47, no. 6, pp. e98–e169, 2016, doi: 10.1161/STR.0000000000000098.

[2] K. R. Anderson, M. L. Woodbury, K. Phillips, and L. V. Gauthier, "Virtual reality video games to promote movement recovery in stroke rehabilitation: A guide for clinicians," *Arch. Phys. Med. Rehabil.*, vol. 96, no. 5, pp. 973–976, 2015, doi: 10.1016/j.apmr.2014.09.008.

[3] R. A. Calvo, S. D'Mello, J. M. Gratch, and A. Kappas, *The Oxford Handbook of Affective Computing*. Oxford Library of Psychology, 2015.

[4] S. Doolani *et al.*, "A Review of Extended Reality (XR) Technologies for Manufacturing Training," *Technologies*, vol. 8, no. 4, 2020, doi: 10.3390/technologies8040077.

[5] M. Taj-Eldin, C. Ryan, B. O'flynn, and P. Galvin, "A review of wearable solutions for physiological and emotional monitoring for use by people with autism spectrum disorder and their caregivers," *Sensors (Switzerland)*, vol. 18, no. 12, p. 4271, 2018, doi: 10.3390/s18124271.

[6] W. World Health Organization, "World Health Organization. Dept. of Child, and Adolescent Health. The importance of caregiver-child interactions for the survival and healthy development of young children: a review.," pp. 1–106, 2004, [Online]. Available: http://apps.who.int/iris/bitstream/10665/42878/1/924159134X.pdf%5Cnabout:newtab.

[7] J. Frommel, C. Schrader, and M. Weber, "Towards Emotion-Based Adaptive Games: Emotion Recognition Via Input and Performance FeatuFrommel, J., Schrader, C. and Weber, M. (2018) 'Towards Emotion-Based Adaptive Games: Emotion Recognition Via Input and Performance Features', in Proceedings of the 2," in *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, 2018, p. 173-185, doi: 10.1145/3242671.3242672.

[8] T. Coelho *et al.*, "Promoting reminiscences with virtual reality headsets: A pilot study with people with dementia," *Int. J. Environ. Res. Public Health*, vol. 17, no. 24, pp. 1–13, Dec. 2020, doi: 10.3390/ijerph17249301.

[9] H. Ben Abdessalem, Y. Ai, K. S. Marulasidda Swamy, and C. Frasson, "Virtual Reality Zoo Therapy for Alzheimer's Disease Using Real-Time Gesture Recognition," *Adv. Exp. Med. Biol.*, vol. 1338, pp. 97–105, 2021, doi: 10.1007/978-3-030-78775-2_12.

[10] I. Glover, Glover, "Play As You Learn : Gamification as a Technique for Motivating Learners. Proceedings of World Conference on Educational Multimedia," in *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications*, 2013, pp. 1999–2008.

[11] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: Defining 'gamification,'" in *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, MindTrek 2011*, 2011, pp. 9–15, doi: 10.1145/2181037.2181040.

[12] J. Jonsdottir, R. Bertoni, M. Lawo, A. Montesano, T. Bowman, and S. Gabrielli, "Serious games for arm rehabilitation of persons with multiple sclerosis. A randomized controlled pilot study," *Mult. Scler. Relat. Disord.*, vol. 19, pp. 25–29, 2018, doi: 10.1016/j.msard.2017.10.010.

[13] F. Engel *et al.*, "Sensecare: Towards an experimental platform for home-based, visualisation of

emotional states of people with dementia," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, vol. 10084 LNCS, pp. 63–74, doi: 10.1007/978-3-319-50070-6_5.
[14] "Sensor Enabled Affective Computing for Enhancing Medical Care | SenseCare Project | H2020 | CORDIS | European Commission." https://cordis.europa.eu/project/id/690862/fr (accessed Jul. 21, 2021).
[15] "SenseCare: Sensor Enabled Affective Computing for Enhancing Medical Care | FTK – Research Institute for Telecommunication and Cooperation." https://www.ftk.de/en/projects/senscare (accessed Aug. 25, 2021).
[16] M. Healy, R. Donovan, P. Walsh, and H. Zheng, "A Machine Learning Emotion Detection Platform to Support Affective Well Being," in *Proceedings - 2018 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018*, 2019, pp. 2694–2700, doi: 10.1109/BIBM.2018.8621562.
[17] D. Maier, "Analysis of technical drawings by using deep learning," Hochschule Mannheim, Germany, 2019.
[18] H. Hadjar, P. Mckevitt, and M. Hemmje, "Home-based Immersive Web Rehabilitation Gaming with Audiovisual Sensors," in *ACM International Conference Proceeding Series*, 2022, pp. 1–7, doi: 10.1145/3552327.3552347.
[19] H. Hayette, V. Binh, D. Maier, G. Mayer, P. Mc Kevitt, and M. Hemmje, "Video-based emotion detection analyzing facial expressions and contactless vital signs for psychosomatic monitoring," *Cerc2021*, vol. 6473, 2021.
[20] P. Amorim, B. S. Santos, P. Dias, S. Silva, and H. Martins, "Serious games for stroke telerehabilitation of upper limb-a review for future research," *Int. J. Telerehabilitation*, vol. 12, no. 2, pp. 65–76, 2020, doi: 10.5195/ijt.2020.6326.
[21] M. Khademi, L. Dodakian, H. M. Hondori, C. V. Lopes, A. McKenzie, and S. C. Cramer, "Free-hand interaction with leap motion controller for stroke rehabilitation," in *Conference on Human Factors in Computing Systems - Proceedings*, 2014, pp. 1663–1668.
[22] L. Yang, J. Chen, and W. Zhu, "Dynamic hand gesture recognition based on a leap motion controller and two-layer bidirectional recurrent neural network," *Sensors (Switzerland)*, vol. 20, no. 7, p. 2106, 2020, doi: 10.3390/s20072106.
[23] F. Weichert, D. Bachmann, B. Rudak, and D. Fisseler, "Analysis of the accuracy and robustness of the Leap Motion Controller," *Sensors (Switzerland)*, vol. 13, no. 5, pp. 6380–6393, May 2013, doi: 10.3390/s130506380.
[24] P. A. Muller, *Modélisation objet avec UML*, vol. 514. Eyrolles Paris, 1999.
[25] "Tracking | Leap Motion Controller | Ultraleap." https://www.ultraleap.com/product/leap-motion-controller/ (accessed Jan. 26, 2023).
[26] "GitHub - leapmotion/leapjs: JavaScript client for the Leap Motion Controller." https://github.com/leapmotion/leapjs (accessed Mar. 16, 2023).
[27] "Hello World — Leap Motion JavaScript SDK v3.2 Beta documentation." https://developer-archive.leapmotion.com/documentation/javascript/devguide/Sample_Tutorial.html (accessed Mar. 16, 2023).
[28] "VARTISTE." https://vartiste.xyz/landing.html (accessed May 17, 2023).
[29] D. Marcos, D. McCurdy, and K. Ngo, "A-Frame – Make WebVR," 2015. https://aframe.io/ (accessed Dec. 04, 2021).

The management of migration at the border : accessing the territory and possible consequences of irregular migration

Linda Cottone[1]

[1]*Universidad Autónoma de Barcelona, Spain*

1. Invited Talk

This analysis examines the challenges and dangers faced by migrants, refugees, and asylum seekers attempting irregular migration, including disappearances, violence, and exploitation often linked to transnational organized crime like trafficking in persons and smuggling of migrants. Highlighting these risks emphasizes the importance of safe and regulated migration pathways, which bolster both state security and individual safety while mitigating the impact of such crimes. Visa systems are identified as tools for managing entry to ensure both national safety and the protection of individual rights. In this background, this review also discusses the implications of entry bans, pushbacks, and the risks of rights violations at borders. It considers alternatives to immigration detention to meet international standards and reflects on voluntary return as an option respecting human dignity. The analysis concludes with insights on balancing public health, security, and individual rights, as particularly highlighted by the mobility restrictions during the COVID-19 pandemic special circumstances.

CERC 2023: Collaborative European Research Conference, June 09–10, 2023, Barcelona, Spain
linda.cottone@gmail.com (L. Cottone)

ISBN 978-3-96187-021-9
https://doi.org/10.48444/h_docs-pub-518

Towards Emerging Argument Integration into Medical Informational Chatbot Dialogs

Stephanie Heidepriem[1], Alexander Duttenhöfer[1], Christian Nawroth[1] and Matthias Hemmje[1]

[1]*FernUniversität Hagen*

Abstract

In this paper, we present an interdisciplinary approach on how a medical informational chatbot can use emerging Arguments (eAs) represented as emerging Argument Tree Structure (eATS) to improve its expressiveness, decision-making ability, and explainability. For this, we use an experimental and unvalidated gold standard eATS about SARS-CoV-2 represented as a Resource Description Framework (RDF) graph as reproducible, documentable and explainable medical informational dialog control. Our approach focuses on eATS as the basis for the dialog, allowing the chatbot to explain its actions and thus comply with the artificial intelligence regulations and proposed legislation of the European Union (EU).

Keywords

Natural Language Processing, Natural Language Understanding, Medical Informational Chatbot, emerging Arguments, emerging Named Entity, Resource Description Framework

1. Motivation and Introduction

Since the launch of ChatGPT in 2022, the use of chatbots has been on the rise, including in the medical field [12, 20, 21]. A chatbot is a user interface for having a conversation with a machine. Chatbots take input in natural language, process the language, and provide an appropriate response [11]. According to Singh et al. [22], a chatbot is technically a combination of technology, Artificial Intelligence (AI), and business process design. A Medical Informational Chatbot is a chatbot that has medical knowledge and can interact with the user about it. It can convey knowledge and assist the user with various medical questions.

Factual knowledge that influences diagnosis in medical applications is highly relevant and often new knowledge is represented as emerging Named Entities (eNE) that occur in eAs. One project addressing this issue is *RecomRatio* [5, 3, 2], which aims to support medical argumentation and reasoning based on textual evidence. Until now, factual knowledge of diagnostic reasoning has been mostly based on textual information named entities and arguments whereas new knowledge and current research is formalized as emerging Arguments [17, 14, 18, 1, 15, 16]. However, medical informational chatbots can guide client dialogs to support medical decisions [12, 21]. Thus, we plan to integrate eATSs as textual information represented as Extensible

CERC 2023: Collaborative European Research Conference, June 09–10, 2023, Barcelona, Spain

stephanie.heidepriem@fernuni-hagen.de (S. Heidepriem); alexander.duttenhoefer@fernuni-hagen.de (A. Duttenhöfer); christian.nawroth@fernuni-hagen.de (C. Nawroth); matthias.hemmje@fernuni-hagen.de (M. Hemmje)

0000–0002–0163–8781 (A. Duttenhöfer); 0000–0001–8293–2802 (M. Hemmje)

ISBN 978-3-96187-021-9
https://doi.org/10.48444/h_docs-pub-518

Markup Language (XML) into reproducible, documentable and explainable medical informational chatbot dialogs.

The research method for this work is based on the framework of Nunamaker et al. (1990). The research approach consists of the four phases: Observation, theory building, system development and experimentation [19].

2. State of the Art in Technology and Related Work

Medical factual knowledge can be represented as eAs which contain newly emerging research information that is not yet available as eNEs in medical dictionaries and thus not known to an extended medical expert group [17, 14, 15, 16]. Christian Nawroths emerging Named Entity Recognition Information Retrieval System (eNER-IRS) is one system that can detect eAs in medical publications and returns them structured as mindmaps. The eNER-IRS searches eAs based on medical user given input parameters as publication and filter criteria. Additionally, further arguments in related publications are identified and returned as mindmaps grouped by publication and filter criteria. The mindmaps contain eAs, publication identifier and eNEs, but no further relationships between arguments or medical information.

A medical informational chatbot should not only work with medical knowledge, but also interact with a user. It should provide information and recommendations and can thus be described as a recommendation system for medicine. In a literature review by Stark et al. [23], recommendation systems for the selection of an appropriate medication were considered. This shows that such recommendation systems also exist in the medical field. What is not apparent here, however, is how much these systems are also used in practice. One sticking point of any recommendation system is to justify its recommendation. Since many of these systems are based on statistical methods and use AI, this recommendation will not be comprehensible to the user without an explanation and thus the trust in the system will decrease. To ensure that, among other things, such AI systems used in the EU are transparent, safe, ethical and under human control, the EU 2021 has put forward a proposal by a first ever AI legal framework, the “harmonised rules on artificial intelligence (artificial intelligence act)” [7]. According to this, AI systems are classified into four categories: minimal risk, limited risk, high risk, and unauthorized. Chatbots like here belong to the limited risk category, since the user does not have to follow the recommendations. Nevertheless, for acceptance as a recommendation system, it is essential that the chatbot can explain to the user why it is passing on which information or recommendations.

The concept presented here is part of Stephanie Heidepriem’s dissertation project: “Enabling Knowledge-based Chatbot Dialogs to Support Web-based User Interfaces to Medical Information, Recommendation, and Advice Systems”. In this project, the above mentioned problems, which are to be partially solved by this concept, have to be considered. However, the target system is exclusively related to mental illnesses and is not primarily intended as a diagnostic system, but as a support and information system for patients and their relatives.

The transformation and modeling of eAs into eATSs is addressed in the dissertation of Alexander Duttenhöfer under the current working title “AI-Based Extraction and Use of Argumentation Trees for Information Retrieval Support”. In his approach, eAs are represented as RDF graphs

and formalised as XML, resulting in eATSs. They provide structured and explainable information by metadata, relations and additional properties that are processable by Medical Informational Chatbots or other Information Retrieval Systems.

In order to achieve this explainability, the mindmap structures of Christian Nawroths eNER-IRS are transformed into understandable eATSs. Hence, the eAs must be converted into RDF and enriched by additional data.

3. SARS-CoV-2 RDF/XML Gold Standard

We use an unvalidated and experimental dataset on SARS-CoV-2 represented as RDF/XML and consisting of eAs in a tree structure called eATS as a gold standard for this work. By the gold standard, we mean a standardized approach that has been shown to produce good results when it is used [9, 10, 13]. The chatbot considers all eAs in the gold standard as truth and uses them accordingly. The dataset consists of the root node and several chained child nodes. In the end, there are 3 layers, with the outermost child nodes representing the eAs:

- Layer 1 (L1): Root node indicating the topic SARS-CoV-2
- Layer 2 (L2): Medical categories *Diagnostics*, *Symptoms* and *Therapy* that group eAs
- Layer 3 (L3): eAs regarding SARS-CoV-2 gold standard

L1 contains only the root node and represents the entry point of the SARS-CoV-2 gold standard tree structure. L2 includes three nodes *Diagnostics*, *Symptoms* and *Therapy*, where the node *Therapy* has no relevant child nodes and is used as a placeholder to demonstrate the extensibility of our tree structure. L3 consists of all 19 eAs that are part of our SARS-CoV-2 gold standard tree structure.

The following properties are (partially) applied to the SARS-CoV-2 gold standard eATS nodes:

- **argStrength**: A fictitious value representing the strength of an eA
- **argText**: Text specifying the node
- **argType**: Specifies whether the eA supports or contradicts the root node
- **parent**: Explicit relation to the parent node
- **publicationID**: Reference to the publications on which the eA is based on

While each eA of L3 has each of the mentioned properties, the other layer nodes only use subsets of the property list. Nodes of L2 have a text description via the property *argText* and a parent-relation *parent* to L1. The root node does not have a parent-relation, so it only uses the *argText* property. With this approach, we add metadata information to L1 and L2, while the L3 eA nodes represent emergent and factual knowledge.

Figure 1 shows a representative extract of the gold standard eATS containing all three layers and their relationships. This extract is extendable by new nodes or layers without any restriction except that there can be only one node in L1, which is the root node.
The eAs of L3 are related to the medical categories of L2, which schematically group all eAs. The categories *Symptoms* and *Diagnostics* contain 9 and 8 eA nodes respectively. The list of symptoms includes arguments for and against a SARS-CoV-2 infection while the diagnostic

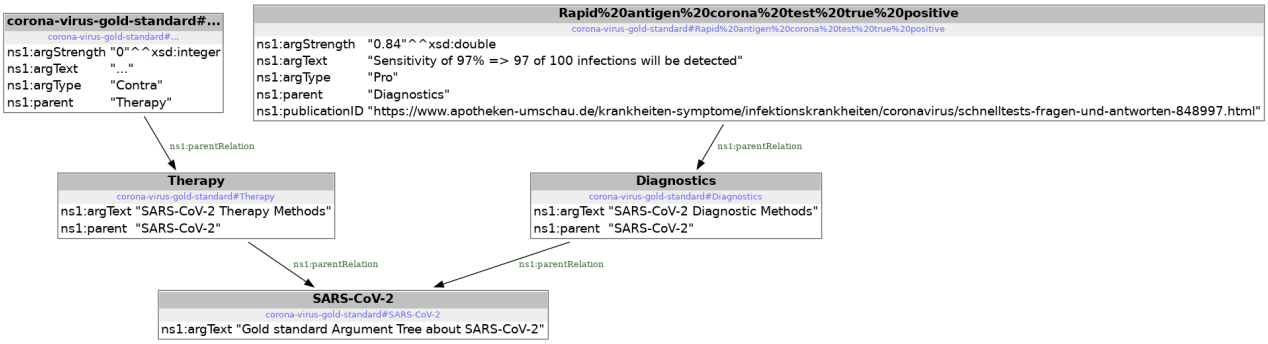


Figure 1: SARS-CoV-2 RDF Gold Standard extract of diagnostic eA and therapy placeholder node example

arguments include the accuracy of rapid antigen and Polymerase Chain Reaction (PCR) tests. Additional information regarding all eAs related to a SARS-CoV-2 infection is attached directly to each node. This includes an abstract and experimental argument strength, a type indicating whether the argument is for or against infection, and a publication id pointing to a reference where the argument appears to be understandable and comprehensible.

4. Conceptual Modeling Approach

The use of eATSs is intended to address several problems of a medical informational chatbot, such as documentation, dialog control, explanation, and decision making. The idea is that two eATSs are always used simultaneously during a dialog: an existing one, which we have already described as the gold standard, and one that is built in parallel with the chatbot user answers corresponding to the dialog, which we call the dialog tree structure.

The gold standard is initially used primarily to guide the dialog. If the user tells the chatbot that they are concerned about SARS-CoV-2 infection, the chatbot should use its Natural Language Understanding (NLU) component to recognize that the user is asking about the presence of SARS-CoV-2 infection. Accordingly, the chatbot searches for the SARS-CoV-2 root node. It then begins to query the tree structure for eAs for and against an infection, and asks the user for these arguments. In this way, it would be possible for the chatbot to draw the conclusion that an infection with SARS-CoV-2 is suspected or not, and to justify this conclusion using the collected arguments. Drawing such a conclusion by weighting arguments, as in [6], will be considered in more detail in later work.

During the dialog with the user, the dialog tree structure is built at the same time. This tree structure will be schematically similar to the gold standard, but it will only contain the parts that were queried by the conversation and their answers. This means that the conversation is also documented. This documentation serves both the documentation requirement that the chatbot has to follow as a medical system and the reproducibility test. If the medical informational dialog is repeated later with the same level of knowledge, reproducibility must be guaranteed.

Using both reasoning tree structures, it should also be possible to develop a clarification component. The EU Data Protection Regulation stipulates that an AI system must be able to explain its actions. Accordingly, two types of dialog are distinguished in a medical informational

dialog: the knowledge recommendation dialog and the meta dialog or clarification dialog. The knowledge recommendation dialog, is the standard dialog with the user, as described in the example. It gives recommendations, conveys knowledge, and asks for arguments. The clarification dialog should be enabled if desired, and should provide additional justification for the chatbot's actions. For example, if the user is being asked about the presence of a SARS-CoV-2 infection, the chatbot can justify the question about symptoms by saying "The question about the presence of a SARS-CoV-2 infection has been recognized. Existing symptoms and positive tests indicate for the presence of an infection. Question about symptoms".

5. Emerging Argument Tree Structure Chatbot Integration

To model the integration, we use the Rasa chatbot framework, according to the framework selection in [8]. Running a Rasa chatbot requires the Rasa Server, which includes the Rasa Core and Rasa NLU components. Rasa NLU parses user input and determines intentions and entities. It then passes these on to Rasa Core, which determines the next actions. Core and NLU are trained using deep learning to recognize user intent and how to respond to it.

To make the chatbot more flexible, custom actions can also be defined, which are executed on demand. Such custom actions can, for example, perform database accesses. Rasa follows the concept that these actions run in a separate component: the Rasa Action Server. This can be used independently of the Rasa server. It provides a port that is made known to the Rasa server via a configuration. When Rasa Core detects that a custom action should be executed now, it sends the corresponding request to the Rasa Action Server [22]. From there, we can then access to the reasoning trees and thus control the conversation.

The access to the existing eATS and the construction of the dialog tree structure is controlled by an Application Programming Interface (API), which we call emerging Argument Tree Structure System (eATSS). However, this still needs to be developed.

For the user interface, the Rasa server provides a port (port 5005 by default) that can be used to access the chatbot. In our modeling, the chatbot is integrated into a web page that runs on a web server. The plugin for the website is used from open source libraries and is only configured using the Rasa server's socketUrl.

Figure 2 shows the integration as a component diagram.

6. Evaluation

The present concept has not yet been implemented. A medical informational chatbot as a whole system is difficult to evaluate and there is still no standard for the evaluation of such systems, despite the growing interest in them. The best practice, is to use human judgments, but this is very time consuming.

Cases at. al [4] published an article in 2020 in which they looked at trends and methods in chatbot evaluation. The main finding was that the methods considered were generally consistent with the concept of usability ISO 9214. This primarily looks at user effectiveness, efficiency, and satisfaction in order to achieve certain goals.

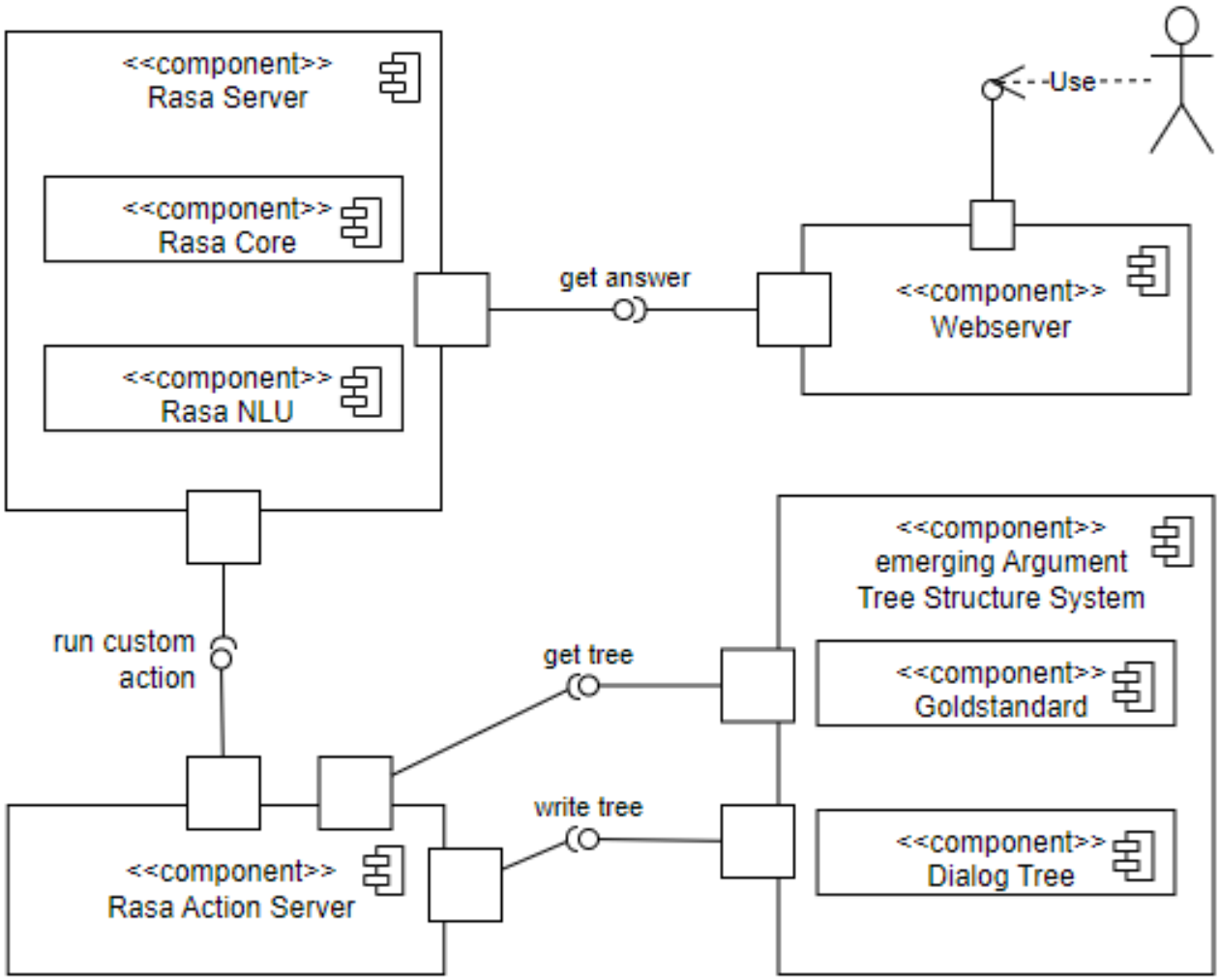


Figure 2: Chatbot Integration Component Diagram

Evaluation of the overall system will also have to be done in part by human subjects. At the same time, we plan to evaluate the individual components. For example, the Gold Standard still needs to be evaluated first. For this purpose, it is necessary to discuss our tree with experts on the subject. Then the Dialog Tree has to be tested. This can be done using sample dialogs. The gold standard can be compared with the Dialog Tree, as it should contain a subtree of the gold standard. The capabilities of NLU and Natural Language Generation (NLG) can also be assessed individually. Measures such as Precision, Recall and F1 Score can be used for this purpose.

However, the dialog guidance itself and whether the available explanations are comprehensible must be humanly tested especially in the overall concept.

7. Discussion and Future Work

The current design displayed in component diagram 2 integrates the eATSS into the Medical Informational Chatbot to provide explainable dialog structures represented as eATS.

Especially with regard to the later planned system, a few more points need to be clarified.

The dialog control should be graph-based using a gold standard. This leaves open the question of how such a gold standard can be generated as automatically as possible, and whether it is always possible to make decisions with such a standard. Automatic generation can only work if there are suitable sources from which this standard can be extracted. However, this extraction must also be thoroughly checked. In addition, it should be considered how the individual pieces of knowledge in the tree can be linked to the original source, so that it is traceable where this information comes from. In this context, the handling of knowledge that is in a state of flux is particularly interesting. It must be possible to update the gold standard if new knowledge changes the procedure.

The current use case from diagnosis is still relatively simple. But not only diagnostic dialogs are to be implemented in the later system, but also dialogs that serve exclusively for information, or dialogs that suggest e.g. techniques that the user can perform together with the chatbot: e.g. relaxation techniques, as prototypically implemented in the work of [8]. Here it will be almost impossible to create a gold standard for every available possibility and every possible conversation flow. Even in cases where a gold standard can be used, further consideration must be given to the weighting of arguments for prioritized argument retrieval.

The target system should also primarily contain and inform knowledge about mental illnesses and therapies. Here, the emotional state of the user may have to be taken into account. How well such a state can be determined in a textual chatbot remains to be tested. It also needs to be determined to what extent it should influence the dialog. A gold standard for mental illness is also difficult to implement, because the symptoms are varied and there is usually no very clear test for whether an illness is present or not.

The structure of a dialog tree in eATSS has yet to be developed. The challenge is what exactly this tree should look like. On the one hand, it should contain all responses of the user and the bot. On the other hand, it should be structurally similar to the gold standard, so that it is still recognizable why which question was chosen, and this can be easily explained.

In addition, there is the challenge of enabling the chatbot to operate in an explanation mode while converting the arguments from the trees into natural language. This requires the use of NLG techniques. Furthermore, it is necessary to address the general question of how a good natural language explanation should be structured so that it is ultimately understandable to the user.

Two other challenges, especially for a medical informational chatbot, are data protection and reproducibility. According to the presented proposal, reproducibility should always be given if the underlying gold standard has not changed. Data security has not been considered yet, but needs to be considered for future work.

The eATS, presented in this paper, is still experimental and has to be defined properly to provide a set of properties, attributes, formalisations and visualizations. Through the shown interdisciplinary approach, the eATSS and its tree structures need to be adapt- and adjustable for special use cases, but should have a common basis. The tree structures also could be improved via the Argument Interchange Format (AIF) which implements a common scheme for interdisciplinary and exchangeable arguments between researchers. Modern applications often provide APIs to integrate systems or use other data sources. The InterPlanetary File System (IPFS) protocol implements peer-to-peer networking where each user could participate in sharing or providing data. This approach could lead to growing publicly available data sets or

eATS that could be shared between multiple applications without implementing a system specific API. All applications could use and integrate the available IPFS data into other applications.

Overall, we can say that we have presented an idea for implementing a diagnostic conversation in a medical information chatbot. Next, the concept will be prototyped to see if it is possible to create a conversation flow with a clarification component for such dialogs. Further concepts have to be developed to evaluate which other dialog types the chatbot has to implement and how they can be implemented. The extent to which emotions can be recognized and change the conversation also needs to be investigated. Data security and reproducibility have to be considered.

Future research, which also goes in the direction of the target system, is the project Supporting Mental Health in Young People: Integrated Methodology for cLinical dEcisions and evidence-based interventions (SMILE)[1]. It aims to improve resilience in young people and explore the pathways that help people develop coping strategies. To this end, a gamification platform is being developed that focuses on interpersonal interaction and collaborative learning and problem solving. In addition, an open knowledge platform is in the background that will provide access to acquired knowledge, provide intelligent functions, and improve decision-making processes. Furthermore, an expressive multilingual, speech-enabled chatbot will be developed, which will also be coupled to digital mentors.

References

[1] Bauer, W., Warschat, J.: Smart Innovation durch Natural Language Processing: Mit Künstlicher Intelligenz die Wettbewerbsfähigkeit verbessern. Hanser, München (2020)
[2] of Bielefeld, U.: Rationalizing recommendations (recomratio). Online (2017), http://ratio.sc.cit-ec.uni-bielefeld.de/de/projekte/ratiorec/
[3] of Bielefeld, U.: Rationalizing recommendations (recomratio): Project-homepage. Online (2017), http://www.spp-ratio.de/de/projekte/ratiorec/
[4] Casas, J., Tricot, M.O., Khaled, O.A., Mugellini, E., Cudré-Mauroux, P.: Trends & methods in chatbot evaluation. In: Companion Publication of the 2020 International Conference on Multimodal Interaction. ACM (oct 2020). https://doi.org/10.1145/3395035.3425319
[5] Consortium, R.: Dfg projektantrag recommratio. Tech. rep. (2017)
[6] Duttenhöfer, A., Wagenpfeil, S., Nawroth, C., Cheddad, A., Kevitt, P.M., Hemmje, M.: Supporting argument strength by integrating semantic multimedia feature detection with emerging argument extraction (2021)
[7] EUROPEAN COMMISSION: Proposal for a regulation of the european parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts (Apr 2021), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206
[8] Heidepriem, S.: Generation of a knowledge base for chatbots to calm persons with mental disorders. mathesis, FernUniversität Hagen (Feb 2021)
[9] Idzko, M., Merkle, M., Sandow, P., Teschler, S., Töpfer, V., Voshaar, T., Vogelmeier, C.F.:

[1] https://www.horizonsmile.eu/

COPD: Neuer GOLD-standard betont die individualisierte therapie. Deutsches Ärzteblatt Online (feb 2019). https://doi.org/10.3238/perspneumo.2019.02.15.006
[10] Jones, A.M., Burnley, M., Black, M.I., Poole, D.C., Vanhatalo, A.: The maximal metabolic steady state: redefining the 'gold standard'. Physiological Reports **7**(10), e14098 (may 2019). https://doi.org/10.14814/phy2.14098
[11] Khan, R., Das, A.: Build Better Chatbots. Apress (2018). https://doi.org/10.1007/978-1-4842-3111-1
[12] King, M.R.: The future of AI in medicine: A perspective from a chatbot. Annals of Biomedical Engineering **51**(2), 291–295 (dec 2022). https://doi.org/10.1007/s10439-022-03121-w
[13] Lopez-Vazquez, C., Gonzalez-Campos, M.E., Bernabe-Poveda, M.A., Moctezuma, D., Hochsztain, E., Barrera, M.A., Granell-Canut, C., Leon-Pazmino, M.F., Lopez-Ramirez, P., Morocho-Zurita, V., Moya-Honduvilla, J., Manrique-Sancho, M.T., Montiveros, M.E., Narvaez-Benalcazar, R., Perez-Alcazar, J.D., Resnichenko, Y., Seco, D.: Building a gold standard dataset to identify articles about geographic information science. IEEE Access **10**, 19926–19936 (2022). https://doi.org/10.1109/access.2022.3150869
[14] Nawroth, C., Engel, F., Eljasik-Swoboda, T., Hemmje, M.: Towards enabling emerging named entity recognition as a clinical information and argumentation support. In: DATA 2018: Proceedings of the 7th International Conference on Data Science, Technology and Applications. pp. 47–55 (07 2018). https://doi.org/10.5220/0006853200470055
[15] Nawroth, C., Engel, F., Hemmje, M.: Emerging named entity recognition in a medical knowledge management ecosystem. pp. 29–41 (01 2020). https://doi.org/10.5220/0010061200290041
[16] Nawroth, C., Engel, F., Hemmje, M.: Utilizing emerging knowledge to support medical argument retrieval. it - Information Technology **63** (03 2021). https://doi.org/10.1515/itit-2020-0049
[17] Nawroth, C., Engel, F., Mc Kevitt, P., Hemmje, M.: Emerging Named Entity Recognition on Retrieval Features in an Affective Computing Corpus. In Proc. of 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM-2019), San Diego, CA, USA, 18-21 Nov. (2019)
[18] Nawroth, C., Schmedding, M., Fuchs, M., Brocks, H., Kaufmann, M., Hemmje, M.: Towards cloud-based knowledge capturing based on natural language processing. ScienceDirect (2015). https://doi.org/10.1016/j.procs.2015.09.236, https://doi.org/10.1016/j.procs.2015.09.236
[19] Nunamaker, J., Chen, M., Purdin, T.D.M.: Systems development in information systems research. Journal of Management Information Systems **7**(3), 89–106 (1991)
[20] Rao, A., Kim, J., Kamineni, M., Pang, M., Lie, W., Succi, M.D.: Evaluating ChatGPT as an adjunct for radiologic decision-making (feb 2023). https://doi.org/10.1101/2023.02.02.23285399
[21] Rao, A., Pang, M., Kim, J., Kamineni, M., Lie, W., Prasad, A.K., Landman, A., Dreyer, K.J., Succi, M.D.: Assessing the utility of ChatGPT throughout the entire clinical workflow (feb 2023). https://doi.org/10.1101/2023.02.21.23285886
[22] Singh, A., Ramasubramanian, K., Shivam, S.: Building an Enterprise Chatbot. Apress (2019). https://doi.org/10.1007/978-1-4842-5034-1
[23] Stark, B., Knahl, C., Aydin, M., Elish, K.: A literature review on medicine recommender systems. International journal of advanced computer science and applications **10**(8) (2019)

Award Winners

Best Presentation Award

The management of migration at the border : accessing the territory and possible consequences of irregular migration

Linda Cottone

Best Presentation Award

Improving Multiclass Classification of Cardiac Arrhythmias with Photoplethysmography using an Ensemble Approach of Binary Classifiers

Katharina Post, Marisa Mohr, Max Koeppel, Astrid Laubenheimer

CERC 2023
Collaborative European
Research Conference
Barcelona, Spain
June 9-10, 2023
www.cerc-conf.eu

ISBN 978-3-96187-021-9

https://doi.org/10.48444/h_docs-pub-518