Natural Language Processing for Medical Texts – A Taxonomy to Inform Integration Decisions into Clinical Practice

Marvin Braun
University of Goettingen, marvin.braun@uni-goettingen.de

Lutz Kolbe
University of Göttingen, lkolbe@uni-goettingen.de

Caspar Neumann
University of Goettingen, c.neumann02@stud.uni-goettingen.de

Natural Language Processing for Medical Texts – A Taxonomy to Inform Integration Decisions into Clinical Practice

Completed Research Paper

Marvin Braun
University of Göttingen
Göttingen, Germany
marvin.braun@uni-goettingen.de

Caspar Clemens Neumann
University of Göttingen
Göttingen, Germany
c.neumann02@stud.uni-goettingen.de

Lutz M. Kolbe
University of Göttingen
Göttingen, Germany
l.kolbe@uni-goettingen.de

Abstract

Electronic health records (EHR) have significantly amplified the volume of information accessible in the healthcare sector. Nevertheless, this information load also translates into elevated workloads for clinicians engaged in extracting and generating patient information. Natural Language Process (NLP) aims to overcome this problem by automatically extracting and structuring relevant information from medical texts. While other methods related to artificial intelligence have been implemented successfully in healthcare (e.g., computer vision in radiology), NLP still lacks commercial success in this domain. The lack of a structured overview of NLP systems is exacerbating the problem, especially with the emergence of new technologies like generative pre-trained transformers. Against this background, this paper presents a taxonomy to inform integration decisions of NLP systems into healthcare IT landscapes. We contribute to a better understanding of how NLP systems can be integrated into daily clinical contexts. In total, we reviewed 29 papers and 36 commercial NLP products.

Keywords: Healthcare, Natural Language Processing, Taxonomy, Artificial Intelligence

Motivation

In the past decade, electronic health records (EHR) have been adopted by a wide range of healthcare institutions across the world. The US government alone spent roughly 40 billion dollars to promote the implementation of EHR systems (Fierce Healthcare, 2019). In 2021, EHRs saw an adoption rate of nearly 100% by hospitals and around 80% by office-based physicians (HealthGovIT, 2021). By increasing the information quantity and quality, EHR adoption will ultimately improve patient care. Reduced documentation times, higher information quality through implemented documentation policies, and a reduction of medication errors are some advantages that are commonly mentioned (Campanella et al., 2016; Domaney et al., 2018). Despite this, many of the promised advantages have not materialized (yet) (Colicchio et al., 2019). On the contrary, EHR adoption in healthcare introduced several unintended consequences for clinicians (Gephart et al., 2015). These unintended consequences ultimately result in increased documentation times for clinicians. For every hour of patient contact, clinicians spend about two hours documenting (Arndt et al., 2017). One example is the scanning of medical documents of a patient for relevant information and understanding of the patient's history as a whole (Colicchio et al., 2019). These documents often include many redundant and outdated information (Wrenn et al., 2010) that further increases the time needed to read and process patient information. Research suggests that the time spent on such documentation processes puts clinicians under increasing pressure and fosters burnout syndromes (Adler-Milstein et al., 2020; Domaney et al., 2018). Moreover, pressure on clinicians has been increased further by the burdens of the pandemic (e.g., lack of personnel) (Stuijfzand et al., 2020).

Natural Language Processing (NLP) aims to tackle this problem by providing several functions to extract, structure, and summarize information from unstructured texts and, thereby, save the valuable time of clinicians (Velupillai et al., 2018; Wang et al., 2018). The ever-increasing NLP capabilities with new models such as generative pre-trained transformers (GPT) show great potential for the medical domain (Korngiebel & Mooney, 2021). Its capabilities have been proven in various healthcare areas such as treating cardiovascular diseases and nutritional and mental disorders (Wang et al., 2018). However, these studies mainly utilize NLP for research purposes and not to relieve physicians in their daily clinical work.

Several research studies suggest that the lack of correct integration of NLP systems into the *socio-technical environment* (i.e., the interaction resulting from humans, systems, and the tasks at hand - Zhang & Li, 2004) of clinical contexts is one of the main limitations for successful utilization (Liu et al., 2012; Zheng et al., 2015). The clinical workflow is especially important to ensure usability for clinicians (Davenport & Kalakota, 2019; Petitgand et al., 2020). We argue that bringing together the huge, demonstrated potential of NLP in healthcare on the one hand, and the challenge of implementing the technology into clinical practice, on the other hand, through a structured overview of integration dimensions is of great use for both practitioners and researchers. For this work, we define integration dimensions as parameters that need to be considered when implementing NLP systems into daily clinical work. To provide this overview, we choose the method of taxonomy development to structure the body of knowledge and derive characteristics and underlying dimensions of integration decisions for NLP systems in healthcare. Thereby, we derive the following research question:

> *What are the integration dimensions and characteristics of NLP systems for medical information processing in healthcare?*

To answer the research question, we follow the proposed method of Nickerson et al. (2013) together with the updated extension of Kundisch et al. (2022) and design a taxonomy for NLP systems in healthcare. This conceptual study addresses the calls for unlocking the potential of AI and related technologies in healthcare (Davenport & Kalakota, 2019) by systematizing integrational dimensions that are crucial for the adoption of such technologies and applications in daily clinical contexts.

Natural Language Processing for Medical Texts

The amount of available medical texts that can be utilized to increase the quality of patient care but also to fasten up research processes in healthcare is massively expanding (Wang et al., 2018). However, most of the information is unstructured in the sense that it is not directly readable by machines (Jensen et al., 2017). NLP is seen as a core technology to overcome this issue and structure the unstructured information for further use. NLP provides clinicians with a viable alternative to the laborious and erroneous process of extracting information from patients' records (Murff et al., 2011). The power of NLP has been demonstrated for a wide range of diseases (e.g., diabetes - Kumar et al., 2014, smoking cessation - Liu et al., 2012) and other topics such as adverse drug events (Sohn et al., 2011).

However, to achieve this goal, NLP needs to be successfully implemented into the information technology (IT) infrastructure of the respective healthcare provider. For a successful implementation, both the individual requirements of the organization and the characteristics of the NLP system need to be considered (Braun et al., 2022). As demonstrated by Wang et al. (2018), NLP systems strongly vary in their used technology and scope. NLP encompasses a series of functions including syntactic processing (splitting texts into sentences or words), information extraction (extracting information of interest), and relationship detection (e.g., correlation of medications, time, and symptoms) (Koleck et al., 2019) that are used in dependence of the respective task. Moreover, there are different technology stacks an NLP system can be based upon. In this work we differentiate between AI-based systems and rule-based systems, sometimes also referred to as expert systems. Rule-based systems utilize strict patterns (i.e., rules) that are compared to the text to find matches. These rules are usually created manually. Rule-based systems are still commonly used, especially in clinical settings (Wang et al., 2018). However, in recent years, AI-based NLP systems have gained traction due to the advances made by the underlying technologies (e.g., AI algorithms such as

Support Vector Machines) and they usually beat rule-based systems in terms of accuracy (Wang et al., 2018). However, clinical practice still favors rule-based systems since they provide greater interpretability in an environment where accurate information processing and comprehension are vital. Moreover, most commercial software vendors' NLP systems are rule-based (Braun et al., 2022; Wang et al., 2018).

To sum up, NLP is already applied across many different isolated research endeavors in healthcare. However, to be fully implementable into daily clinical practice, there is still a lack of knowledge on how NLP-based systems vary in their scope, function, and ability to be integrated into the IT infrastructure of healthcare providers. While there are existing taxonomies for adjacent system types such as decision support systems in healthcare (Berlin et al., 2006), we specifically focus on the individual context of NLP systems and its unique requirements for integration into daily clinical practice. To successfully integrate NLP systems into practice, knowledge of the system itself and the implications of the system's characteristics for integration is necessary.

Method

This paper sheds light on relevant integration decisions when implementing NLP systems into the IT infrastructure of healthcare providers. Hence, we identify and systematize the dimensions and characteristics of NLP systems that impact integration decisions. To achieve this, we use the extended method of taxonomy development proposed by Kundisch et al. (2021), which originates from Nickerson et al. (2013). The development process comprises several steps that we outline in the following. First, we describe the object of analysis (1), the target user groups (2), and the intended purpose (3). We analyze NLP systems that are available to the public. In general, the systems we analyze can be grouped as systems that process or generate medical texts with NLP. We specifically exclude systems of research that are not publicly available or proof-of-concept systems. The purpose of the taxonomy is to inform practitioners and researchers about different types of NLP systems and which dimensions of integration need to be considered. Second, we define the meta-characteristic (4) of our taxonomy as *'integration parameters for NLP systems in healthcare providers*.' Moreover, we adopt all subjective ending conditions (5) and three objective ending conditions (6) (unique dimensions, unique characteristics, and at least one object for every characteristic) from Nickerson et al. (2013).

First iteration: We chose a conceptual-to-empirical cycle for our first iteration to develop a profound knowledge of NLP systems in healthcare. To identify relevant dimensions and characteristics and grasp all relevant papers from different research streams, we conducted a systematic literature review, following the proposed methodology of Webster and Watson (2002) (see Figure 1).

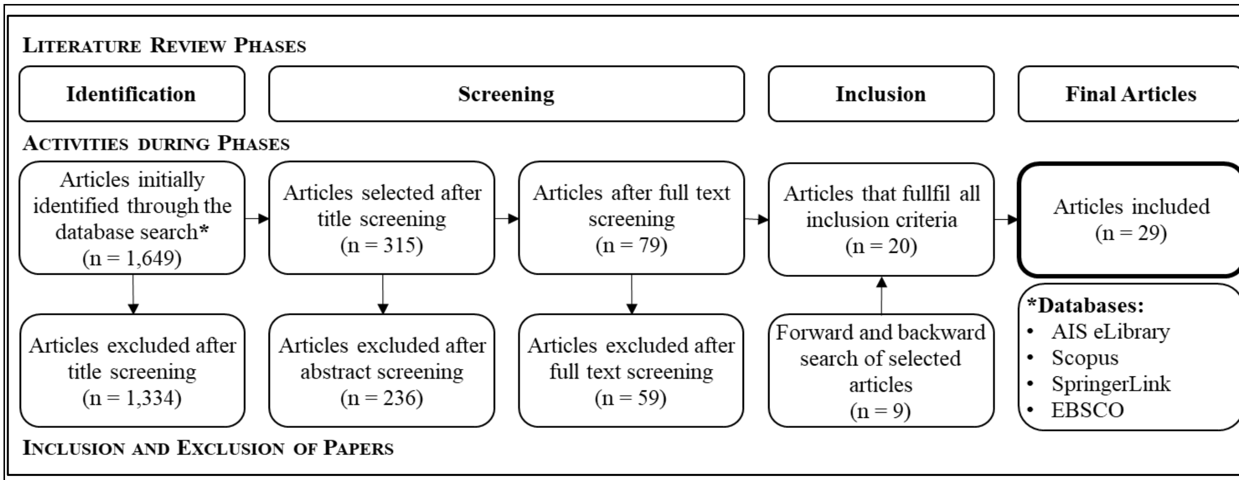


Figure 1. Literature Selection Process

The search string consisted of *'natural language processing / NLP / information extraction / text mining'* and *'health* / clinic* / EHR / electronic health record / medic* text / patient notes'*. This search was conducted between October 2022 and February 2023 and yielded 1,649 search results, which were screened for exclusion (see Fig. 1). We conducted the keyword search in the top IS journals (e.g., European Journal of Information Systems, Information Systems Journal, Information Systems Research, Journal of Association for Information Systems, Journal of Strategic Information Systems, and MIS Quarterly). Moreover, we specifically screened relevant journals of adjacent research streams such as IS in healthcare (e.g., Journal of Biomedical Informatics) and information technology (e.g., Frontiers in Computer Science).

The journals were selected to specifically focus on the integration aspect of NLP systems into daily clinical practice. The final 29 articles included dimensions and characteristics that were used as a starting point. Following the methodology of Webster & Watson, we used a concept matrix to identify and synthesize dimensions and characteristics of integration aspects of NLP systems (see Table 1 for examples).

Article	**Dimensions and Characteristics**	**Synthesized Concepts**
Chen et al., 2019	algorithm: [rule-based, artificial intelligence, hybrid], unstructured data: [operation notes, ct, mri, pathology reports], language: [english, chinese]	algorithm, input documents, language
Gao et al., 2018	deployment mode: [public, private, community], service form: [saas, paas, iaas], supported task: [clinical, administrative, strategy, research]	deployment mode, supported task
Kreimeyer et al., 2017	nlp type: [rule-based, machine-learning, hybrid], open source: [yes, no]	algorithm, pricing
Lockey et al., 2021	applications: [predicting hospital admission of patients from the emergency department, augmentation of the existing triage process, classifying radiology reports to identify the appropriate clinical response, extracting extra detail from free text notes], tasks: [text categorization, information extraction, semantic analysis, machine translation, question answering, chatbots]	supported task
Iroju and Olaleke, 2015	nlp techniques used in healthcare: [symbolic/logical approaches, statistical approaches, connectionist approach, hybrid approach], applications of nlp in healthcare: [information extraction, information retrieval, question and answering, user interface document, categorization, machine translation, text summarization]	algorithm, input documents, supported task
Tremblay et al., 2009	text mining: [rule-based matching (e.g., keywords or regular expressions), natural language processing, machine learning], note-types: [clinical notes, diagnostic reports, radiology reports, pathology reports, colonoscopy reports]	algorithm, input documents

Table 1. Concepts Used for the Taxonomy

Most literature focuses on the underlying algorithms that are used to conduct NLP in healthcare (Dalianis, 2018; Zheng et al., 2015). Moreover, these often use different characteristics to classify the type of algorithm (e.g., rule-based vs. AI-based, supervised vs. unsupervised). By analyzing the literature, we were able to identify the following dimensions to start our taxonomy with*: algorithm, supported task, input documents, supported languages, pricing, and deployment mode.*

Second Iteration: For our second iteration, we chose an empirical-to-conceptual iteration. In this cycle, we systematically searched outlets such as *crunchbase.com* and *healthskout.com* to identify real-world NLP systems for medical texts (see Appendix B for a full list of the reviewed applications). For each application found, we searched for related applications by name. For example, after we found the product "Amazon Comprehend Medical" and added it to our database, we performed a search using the term "Amazon Comprehend Medical alternatives." For all detected applications we directly performed a quick check to see if they were still available and solved a task around medical text analysis in the context of medical text analysis. In total, we initially identified 22 real-world applications in this iteration. We then reviewed each database entry in detail and noted the characteristics of the tools. For this purpose, we checked websites, documentation, and product video demonstrations. Upon closer inspection, four tools were not relevant for the classification (e.g., focus on conversational agents or focus on research). Thereby, the number of tools was reduced to 18. While we initially derived a very granular taxonomy from literature, for example by listing different types of AI-based algorithms and the individual analyzed document types, we refined several characteristics in this iteration because we found that real-world applications' descriptions and scientific publications strongly differ in their focus. For example, while scientific publications usually focus on a single medical event (e.g., a disease), applications usually do not differentiate on such a granular level. In this iteration, we synthesized the *algorithm* dimension and added the dimensions of *customizing* (i.e., training models vs. pre-defined models), *accuracy reporting* (i.e., confidence scores for extracted information or assigned terms), *interface*, *input type,* and *output type*. Due to adjustments that were made to the dimensions and characteristics in this iteration, our ending conditions were not fulfilled, and we needed to perform a third iteration.

Third Iteration: We performed a third iteration (again empirical-to-conceptual) and searched for additional systems that we probably missed. Here we used the databases of capterra.com and quicksprout.com to find additional systems that probably use different wordings. In total, we identified twelve additional relevant systems. Similar to the second iteration, we checked the vendors' documentation, images, and videos to assess the tools. The analysis of the twelve additional tools led to the addition of the following dimensions: *medical dictionary* (i.e., the inclusion of standardized medical term dictionaries such as SNOMED-CT), *data protection* (i.e., the compliance with HIPAA requirements), and *environment* (i.e., describing the ability of the system to either function as an integral part of a larger ecosystem or operate independently as a standalone solution).

Fourth Iteration: Our ending conditions demanded another iteration because we changed and added dimensions. We searched for general NLP solutions that could be adapted to our use case and identified six additional systems. While we were able to classify all four with our taxonomy, the iteration did not result in any change in the taxonomy. Since no further changes were made to the dimensions and characteristics, we ended the taxonomy development after the fourth iteration. Finally, we inductively determined meta-dimensions that contribute to our meta-characteristic and grouped the identified dimensions into groups of linked integration decisions. The full process of the taxonomy development is summarized in Appendix A.

Taxonomy to Inform Integration Decisions of NLP-Systems in Healthcare

The final taxonomy consists of 14 dimensions that are grouped into 5 meta-dimensions (see Table 2). While the original development method proposed by Nickerson et al. (2013) prescribes to use only mutually exclusive characteristics, we found that many NLP systems involve different characteristics of one dimension. Hence, we followed the approach of Püschel et al. (2022) to classify dimensions as either non-exclusive (NE) or mutually exclusive (ME). Moreover, the meta-dimensions were created to group similar dimensions (i.e., high cohesion within the meta-dimension) and cluster the most important topics to inform design decisions of NLP systems in healthcare.

<table>
<tr><th colspan="3">Dimensions (D_n)</th><th colspan="5">Characteristics (C_m)</th></tr>
<tr><td rowspan="4">Application</td><td>Supported Task</td><td>NE</td><td>Medical Entity Extraction</td><td>PHI Extraction</td><td>Encoding</td><td>Querying</td><td>Medical Text Generation</td></tr>
<tr><td>Input Documents</td><td>NE</td><td colspan="2">Physician-Generated Documents</td><td colspan="3">Patient-Generated Documents</td></tr>
<tr><td>Medical Dictionary</td><td>ME</td><td colspan="2">Included</td><td colspan="3">Not included</td></tr>
<tr><td>Supported Languages</td><td>NE</td><td colspan="2">English Only</td><td colspan="3">Other Language</td></tr>
<tr><td>Technical</td><td>Algorithm</td><td>NE</td><td colspan="2">Rule-based</td><td colspan="3">Artificial Intelligence</td></tr>
</table>

	Customizing	NE	Available		Not Available	
	Deployment Mode	NE	On-Premises		Cloud	
Interaction	Input Type	NE	Plain Text	Marked-up Text		Question
	Output Type	NE	Highlighted Text	Marked-up Text	Table	Template
	Interface	NE	GUI	API	SDK	CL
Regulatory	Accuracy Reporting	ME	Confidence Score Available		No Confidence Score Available	
	Data Protection	ME	HIPAA Compliance Explicitly Stated		No HIPAA Compliance Explicitly Stated	
Service Model	Environment	ME	Part of a Medical Ecosystem		Standalone Solution	
	Pricing	ME	Open-Source	Pay-per-Request	Pay-Per-Input Volume	Pay-per-License

Table 2. Taxonomy to Inform Integration Decisions of NLP-Systems in Healthcare

The meta-dimension Application determines the purpose of the system and describes its boundaries. This dimension groups characteristics that determine the medical use case. The Technical dimension groups characteristics that inform decision-makers about the technical boundaries of the NLP system and provides a clear understanding of the necessary technical infrastructure that is needed to integrate the system into the respective healthcare organization. Next, Interaction describes the communication interfaces of the NLP system that determine how the NLP system can communicate with either users or adjacent systems. The Regulatory dimension describes laws and ethical components that need to be considered when integrating the NLP system into clinical contexts. Last, Service Model describes important aspects of pricing and other products of the system vendor. This dimension informs about general conditions that need to be considered when implementing an NLP system. The respective sub-dimensions and characteristics will be presented in the following.

Application: This meta-dimension defines the purpose and scope of the NLP system and, hence, is the most important integration dimension. The **Supported Tasks** dimension (D_1) describes the tasks that can be performed by the text analysis tool. Regarding the purpose of the systems, we identify five common functional patterns across the ten papers that provide classifications for the functions of NLP systems in healthcare. The current main purpose of NLP systems in healthcare is information extraction; however, we can split this function into several sub-categories. First, we identify systems that aim to extract medical information (D_1, C_1) (e.g., symptoms, medications, therapy decisions). While research usually segregates between different entities and focuses on the extraction of single diseases and their symptoms (e.g. Wang et al., 2018), we do not observe this for commercial NLP systems (e.g., Google, 2022). This especially makes sense because, in a clinical context, the system needs to deal with a wide range of medical texts that include all kinds of medical information, whereas research environments are more isolated.

Second, we find systems that focus on the so-called protected health information (PHI) (D_1, C_2), i.e., identifying and personal health information (e.g., name, age, special medical conditions). PHI extraction is concerned with recognizing, processing, and anonymizing information such as name, age, and medical record numbers (Meystre et al., 2010).

Besides these two extraction tasks, we identify encoding (Casey et al., 2021; Miranda-Escalada et al., 2020) and querying as further tasks. Under encoding, systems assign codes or standardized terms such as the International Statistical Classification of Diseases and Related Health Problems (ICD) or Systematized Nomenclature of Medicine – Clinical Terms (SNOMED-CT) to the diseases described in a patient care document (D_1, C_3). In clinical contexts, mapping the found medical entities onto standardized medical vocabularies is crucial for further processes (Randorff Højen & Rosenbeck Gøeg, 2012). In this way, simple and efficient billing can be done. Next, we identified querying as a supported task, which enables clinicians to formulate questions and receive an answer (e.g., the health status of a patient) (D_1, C_4). Many tools offered multiple functionalities such as medical information and PHI extraction. Last, we identify the purpose of medical text generation (D_1, C_5) which is currently emerging with the advances in NLP achieved through GPT.

The type of **Input Documents** (D_2) dimension allows tools to distinguish between physician-generated documents (D_2, C_1) such as doctor notes, clinical trial reports, patient health records or discharge summaries, and patient-generated documents (D_2, C_2) such as questionnaires or self-formulated disease descriptions. Patient-generated documents are normally less precise and professional with the use of medical terminologies compared to physician-generated documents (Reyes-Ortiz et al., 2015; Spasić et al., 2020; Tremblay et al., 2009)

The **Medical Dictionary** dimension (D_3) is about the input vocabulary required by the algorithm to perform text analysis. This vocabulary is needed to map extracted information to standardized entities (e.g., to the SNOMED-CT vocabulary). For organizations, it is important to assess whether the medical entities are pre-defined by the software vendor (D_3, C_1) or if they can (and need) to choose a vocabulary by themselves (D_3, C_2). The **Supported Language** (D_4) dimension deals with which languages the tool can process. This is especially important because we observe a huge gap between NLP systems that can process the English language and systems that can process other languages. Due to the huge focus on the English language, we defined the dimensions of 'English' (D_4, C_1) and 'other language' (D_4, C_2).

Technical: This meta-dimension groups technical aspects that are important for the integration decision. The **Algorithm** dimension (D_5) deals with the engine that performs the text processing. Here we make the major distinction between rule-based (D_5, C_1), and AI-based (D_5, C_2) systems. While there exist further subgroups and other classifications of machine-learning systems in research (e.g., deep learning, transformer-based (Casey et al., 2021)), real-world applications often did not further disclose which AI algorithm they used. Many tools used a hybrid combination of rule-based and machine-learning algorithms to increase the flexibility of the systems. The **Customizing** dimension (D_6) makes it possible to distinguish between tools that allow the user to customize the engine and those that do not. The characteristic that states that customizing is available (D_6, C_1) means that the users can either train their AI models or define their own rules. The pendant to this is that customizing is not available (D_6, C_2), and the users rely on pre-defined AI models or sets of rules. The **Deployment Mode** (D_7) dimension contains the two classic variants on-premises (D_7, C_1) and cloud (D_7, C_2) (Gao et al., 2018).

Interaction: Dimensions of this meta-dimension shape the interaction of users with the system and hence, also make implications for the integration. The **Input Type** dimension (D_8) is about the structural type of the input document. Input can be supplied to the tool as plain text (D_8, C_1) or as marked-up text (D_8, C_2). In contrast to plain text, the marked-up text includes annotations and further information that can be processed by the system. The characteristic question (D_8, C_3) refers to tools that need a question like "Which drug was prescribed to Ms. Miller during her last treatment?" as input to then answer it with the help of text analysis. Several tools offered multiple input types.

The **Output Type** (D_9) dimension pursues the same task as the input type dimension, only for the output provided by the system. The output can be provided as highlighted text (D_9, C_1). This includes both tools which highlight the medical entities found in the output visually and tools that tag them textually. Marked-up text (D_9, C_2) includes tools that provide the output in a structured form (e.g., JSON or XAI). This is especially important if the text needs to be transferred to adjacent systems in the context of clinical workflows. The output type table (D_9, C_3) describes tools that list the recognized results in tabular form. This allows both easy readability for humans and further processing for machines. In contrast to the first two characteristics, the tabular form only presents the recognized entities and not the whole text. Last, the template form (D_9, C_4) allows users to define a structure for the information. The **Interface** (D_{10}) dimension describes the systems' interaction possibilities. Services can be used via a Graphical User Interface (GUI) (D_{10}, C_1), Application Programming Interface (API) (D_{10}, C_2), Software Development Kit (SDK) (D_{10}, C_3), or Command Line (CL) (D_{10}, C_4). The interface dimension has a high impact on the integration of the system into the existing healthcare IT infrastructure. All respective dimensions and characteristics of the meta-dimension interaction were generated from the products.

Regulatory: This meta-dimension groups regulatory aspects such as laws and other legal requirements. The **Accuracy Reporting** dimension (D_{11}) deals with whether a tool provides a confidence score for the individual results (D_{11}, C_1) or not (D_{11}, C_2). Only some tools offered the ability to provide a confidence score. The score indicates the degree of confidence in the accuracy of the respective task (e.g., extracted medical entities or assigned code). It is particularly important for defining internal governance guidelines based on this score. Here, for example, it could be specified that every result with a confidence score of less than 90% must be checked manually. The **Data Protection** dimension (D_{12}) allows a binary distinction between tools that explicitly state that they are compliant with the *Health Insurance Portability and Accountability Act (*HIPAA) (D_{12}, C_1) and those that do not (D_{12}, C_2).

Service Model: Dimensions that shape the service model of the NLP system are grouped under this meta-dimension. The **Environment** dimension (D_{13}) consists of two categories and describes whether the text analysis tool is the software provider's only medical IT tool (standalone) (D_{13}, C_1) or whether the provider offers an entire medical-IT ecosystem (D_{13}, C_2). This is especially relevant for integration because the IT infrastructures of healthcare providers often consist of various systems that should be (but not always are) compatible with each other (Davenport & Kalakota, 2019; Palvia et al., 2012). The **Price** dimension (D_{14}) is divided into open-source (D_{14}, C_1) and three commercial dimensions (Kaur & Chopra, 2016; Negro-Calduch et al., 2021). The first of the three commercial dimensions is the *pay-per-input volume* (D_{14}, C_2). Here, billing can take place per page, per word, or according to other criteria related to the volume of the actual document. *Pay-per-request* (D_{14}, C_3) also refers to a form of volume but is based on the number of requests rather than the volume of the document itself. The last form, *pay-per-license* (D_{14}, C_4), applies to tools that sell their software as a license and do not charge according to the actual usage.

Application of the Taxonomy

The main goal of this taxonomy is to group and classify NLP systems for medical texts and inform relevant stakeholders (from research and clinical contexts) about integration aspects. In the following, we will demonstrate the usefulness of the taxonomy by classifying two different artifacts. The classification process and the derived types of NLP systems of the following tools was discussed with two medical experts in the field of NLP to ensure the rigor of this step. Moreover, we calculated the distribution of characteristics for the analyzed objects (see Table 3).

The first product is Google's *Healthcare Natural Language API* (Google, 2022). The NLP system can extract medical information and PHI and can also encode the extracted information into terms of medical vocabularies (e.g., ICD10). The language is restricted to English. Moreover, the system focuses on physician-generated documents. The built-in algorithm uses passive machine learning (AI-based), i.e., the system is not learning continuously. Moreover, the system cannot be trained by users and only offers a predefined model. The API is cloud-based and offers no possibility to host it on-premises. The input type of the system is plain text while the output type is marked-up text that includes the individually-identified information. Moreover, the API provides confidence scores for every extracted information. Several medical dictionaries are included (e.g., SNOMED-CT, ICD9/10, NCBI Taxonomy). The API is HIPAA compliant. Last, the user pays per request and the API is not embedded into a medical ecosystem.

The second product is Apache's *cTAKES* (Apache, 2022), which is an SDK to create an NLP system and has been developed by the Mayo Clinic Organization. cTAKES is used for medical information extraction only. The SDK is not restricted to the English language and can be for example used for documents in German (Becker & Böckmann, 2016). Due to the high flexibility of the SDK, the resulting NLP can also be adapted to patient-generated documents. Moreover, the system uses a hybrid variant of rule-based and AI algorithms. It is highly customizable and can be seen as a toolkit to adapt the different functionalities to individual requirements. Hence, the SDK is only available on-premises. The input and output types are similar to Google's API (plain text, marked-up text). cTAKES includes a Unified Medical Language System (UAIS) dictionary. The SDK offers the ability to include confidence scores in the system. The SDK itself is not HIPAA compliant, which makes sense because it is not a ready-to-use system. Moreover, cTAKES is open-source and offers other components that create a medical ecosystem.

The two examples were chosen because they depict two types of NLP systems that could be observed very often during the classification process. The first type we identify is cloud-based, commercial (i.e., not open-source) systems ***(tech-enabled NLP systems)***. They are usually provided by big tech companies whose focus is not medical products. These tech-enabled systems typically provide either APIs or a GUI (e.g., Amazon 2022, Microsoft, 2022) and hence, can be usually integrated into the existing EHR system or are standalone systems. These systems are only barely configurable, i.e., allow no training of their models and have strong restrictions on input and output types. Moreover, most of these are HIPAA-compliant. The offered systems can be integrated quickly and offer fixed functionalities out of the box. This type of system is generally interesting for healthcare organizations with limited resources and knowledge in this domain.

The second type we identify are highly customizable open-source systems ***(configurable NLP pipelines)*** that usually offer no hosting service, and hence, are deployed on-premises. These configurable NLP pipelines provide several functionalities that can be mixed and matched according to the individual requirements of healthcare organizations (e.g., Apache, 2022). These systems are often strongly connected to research (e.g., MetaMap, 2022) but can also be adapted to the needs of clinical practice. However, they require a customization process and will not offer the same performance out-of-the-box as the tech-enabled NLP systems. The configurable NLP pipelines are favorable for larger healthcare providers with strong resources in the domain of information systems because the implementation process is not as easy, even though it also allows stronger customization toward single-use cases.

The classification process revealed further interesting relationships of characteristics and dimensions of such products that need to be considered when integrating NLP systems to process medical texts. First, we observe that medical information extraction and encoding are the most prominent use-case (~86%) but only a fraction of the systems also includes PHI extraction (~14%). PHI is crucial because it allows the anonymization of the information for further processing (which is a legal requirement in many countries (HHS, 2022)), but also the identification of the patient, depending on the individual use case. Moreover, we observe that medical text generation is currently emerging as a new function of NLP systems for healthcare organizations, which is largely enabled by the advances of GPT.

Dimensions (D_n)			**Characteristics (C_{nm})**				
Application	Supported Task	NE*	Medical Entity Extraction **(86%)**	PHI Extraction **(14%)**	Encoding **(80%)**	Querying **(14%)**	Medical Text Generation **(5%)**
	Input Documents	NE*	Physician-Generated Documents **(100%)**	Patient-Generated Documents **(22%)**			
	Medical Dictionary	ME	Included **(59%)**	Not included **(41%)**			
	Supported Languages	ME	English Only **(86%)**	Multiple Languages **(14%)**			
Technical	Algorithm	NE*	Rule-based **(39%)**	Artificial Intelligence **(88%)**			
	Customizing	ME	Available **(34%)**	Not Available **(66%)**			
	Deployment Mode	NE*	On-Premises **(61%)**	Cloud **(66%)**			
Interaction	Input Type	NE*	Plain Text **(90%)**	Marked-up Text **(26%)**	Question **(3%)**		
	Output Type	NE*	Highlighted Text **(29%)**	Marked-up Text **(41%)**	Table **(12%)**	Template **(17%)**	
	Interface	NE*	GUI **(36%)**	API **(30%)**	SDK **(31%)**	CL **(12%)**	
Regulatory	Accuracy Reporting	ME	Confidence Score Available **(22%)**	No Confidence Score Available **(78%)**			
	Data Protection	ME	HIPAA Compliance Explicitly Stated **(33%)**	No HIPAA Compliance Explicitly Stated **(67%)**			
Service Model	Environment	ME	Part of a Medical Ecosystem **(61%)**	Standalone Solution **(39%)**			
	Pricing	NE*	Open-Source **(33%)**	Pay-per-Request **(11%)**	Pay-Per-Input Volume **(36%)**	Pay-per-License **(22%)**	

*Please note, that due to non-exclusive dimensions, the total percentage can exceed 100%.

The presented distributions are rounded.

Table 3. Distribution of Characteristics

Moreover, there are only a few commercial systems that offer a combination of on-premises hosting with AI-based algorithms (~10%). This is not surprising as rule-based systems are more user-friendly and also less resource-intensive, making them more suitable for on-premises hosting. Nevertheless, cloud solutions are often not considered by healthcare providers, either due to the need to conform to formal regulations or due to privacy concerns (Al-Marsy et al., 2021). Therefore, we argue that the potential of different configurations is not fully exploited. The calculation emphasizes the crucial difference between the application of NLP applications in isolated research settings and daily clinical contexts. Large parts of current applications focus solely on the extraction of medical information from text without considering other relevant functionalities that are crucial for its adoption (e.g., PHI extraction or HIPAA conformity).

Discussion

Summary of Key Findings

Implementing automation systems into healthcare and addressing the individual requirements of individual stakeholders and users remains a critical challenge (Braun et al., 2022; Kelly et al., 2019). Recent advances in the field of NLP achieved through GPT open further possibilities for relieving clinicians in their daily clinical work.

With this taxonomy, we aim to structure the evolving body of knowledge around NLP systems for processing medical texts and provide an overview of integration decisions that need to be made in the context of healthcare IT infrastructures while tackling the underutilization of NLP in the medical domain (Wang et al., 2018). The taxonomy identified 14 integration dimensions across five meta-dimensions and systematized scattered knowledge about the function and scope of NLP systems for medical text processing. The five meta-dimensions summarize key topics that need to be considered when integrating NLP systems into daily clinical practice. We find distinct characteristics within these meta-dimensions that are unique to text-processing applications in the domain of healthcare (e.g., the type of medical dictionary and the language used) and reveal the unique challenges of dealing with textual information in this context, such as the constraints imposed by the languages, medical terms or the input and output types. However, there are also common dimensions such as the employed service model or the regulatory requirements which are also adaptable to (AI-based) visual processing applications. Moreover, we identify two archetypes of integrating NLP applications in healthcare (tech-enabled NLP systems and configurable NLP pipelines) that fundamentally differ in their characteristics and integration into daily clinical practice. While these archetypes support the same goal (enabling the analysis of textual information by NLP), they do so through a contrary integration.

Contribution to Literature

From a theoretical perspective, our work makes two main contributions. To the best of our knowledge, we are the first to investigate integration dimensions and characteristics of NLP systems in healthcare and provide an overview that applies to both real-world applications and the literature. Works investigating the highly complex socio-technical environment of healthcare can use the taxonomy to get a deeper understanding of the implications of different dimensions on integration decisions.

Moreover, the taxonomy development process revealed a large gap between NLP systems used in research projects and clinical practice. We can show that research-related NLP systems only consider a subset of important dimensions that are important for clinical practice. Research provides very narrow NLP systems that focus on a specific phenomenon. Usually, these systems also only process one type of clinical document (e.g., radiology reports (Pons et al., 2016)). In contrast, none of the found commercial applications differentiated between different types of clinical documents or diseases. We conclude that NLP systems in clinical practice need higher flexibility in the context of document type and medical events (e.g., diseases, medication) whereas those in research focus on single aspects of medical information extraction. Hence, we also decided not to further differentiate between clinical document types because it does not reflect clinical practice. Second, the taxonomy delivers a detailed overview of integration topics that need to be investigated further. As pointed out by research (Liberati et al., 2017; Nehme & Feldman, 2021) usability and integration aspects are currently limiting the successful use of AI-related technologies (such as NLP) in clinical practice.

Implications for Practice

Our work also provides contributions to practice. As pointed out in the introduction, the amount of data in healthcare has steadily increased, necessitating effective processing of this data to realize the benefits of higher information quantity. To this end, implementing such systems in the healthcare domain is very challenging as healthcare infrastructures are highly complex and often consist of many isolated, so-called *silo systems*, i.e., systems that do not exchange data with adjacent systems (Palvia et al., 2012). Our taxonomy tackles this problem by providing 14 dimensions that structure the topic of integrating a system into this domain.

Thereby, our taxonomy is of relevance to several groups of stakeholders within the healthcare system. First and foremost, the taxonomy can be beneficial for managers and provide guidance to assess which type of NLP system fits best for the respective organization. Previous studies have underlined the heterogeneity of healthcare organizations, their (maturity of) IS infrastructures, user bases, and the connected challenge of integrating NLP systems by adapting them to the custom needs of the organization (e.g., fine-tuning the system to medical terms and abbreviations) (Liu et al., 2012; Zheng et al., 2015). By following the five meta-dimensions of our taxonomy, managers consider the most important (and regulatorily necessary) aspects of implementing an NLP system in a highly complex environment. Moreover, the landscape is about to change with the integration of new technologies such as GPT-based NLP tools that will enhance capabilities (Korngiebel & Mooney, 2021). In this context, our taxonomy guides the implementation of such systems.

Additionally, the taxonomy also informs software developers of medical NLP systems and healthcare organizations about the unique integration aspects when implementing such systems in daily clinical practice. We identify the gap between AI-based NLP systems that can be hosted on-premises. The healthcare domain has very unique requirements for information (Ford et al., 2016) which partly hinders the implementation of cloud-based solutions. However, most out-of-the-box solutions are cloud (and API)-based which potentially reduces their adoption due to information privacy and security concerns. Developers could work on on-premises solutions that utilize the often more efficient and accurate AI-based algorithms.

Apart from the taxonomy itself, the identified objects reveal common system types that decision-makers need to be aware of. Our results imply that there are at least two common system types, tech-enabled NLP systems and configurable NLP pipelines, that vary strongly in their characteristics. Our taxonomy shows that the integration process of both system types is fundamentally different and demands different technological readiness of the healthcare organizations. The taxonomy can provide support to break ties between two equally-capable NLP systems and support organizations in their decision-making process. The taxonomy ultimately supports the successful implementation of NLP into the healthcare domain.

Limitations and Future Research

Of course, our work is also subject to limitations. The development process of a taxonomy and especially the building of dimensions and characteristics is partly dependent on the individual persons. We tried to overcome this limitation by independently classifying a subset of the NLP systems and comparing them.

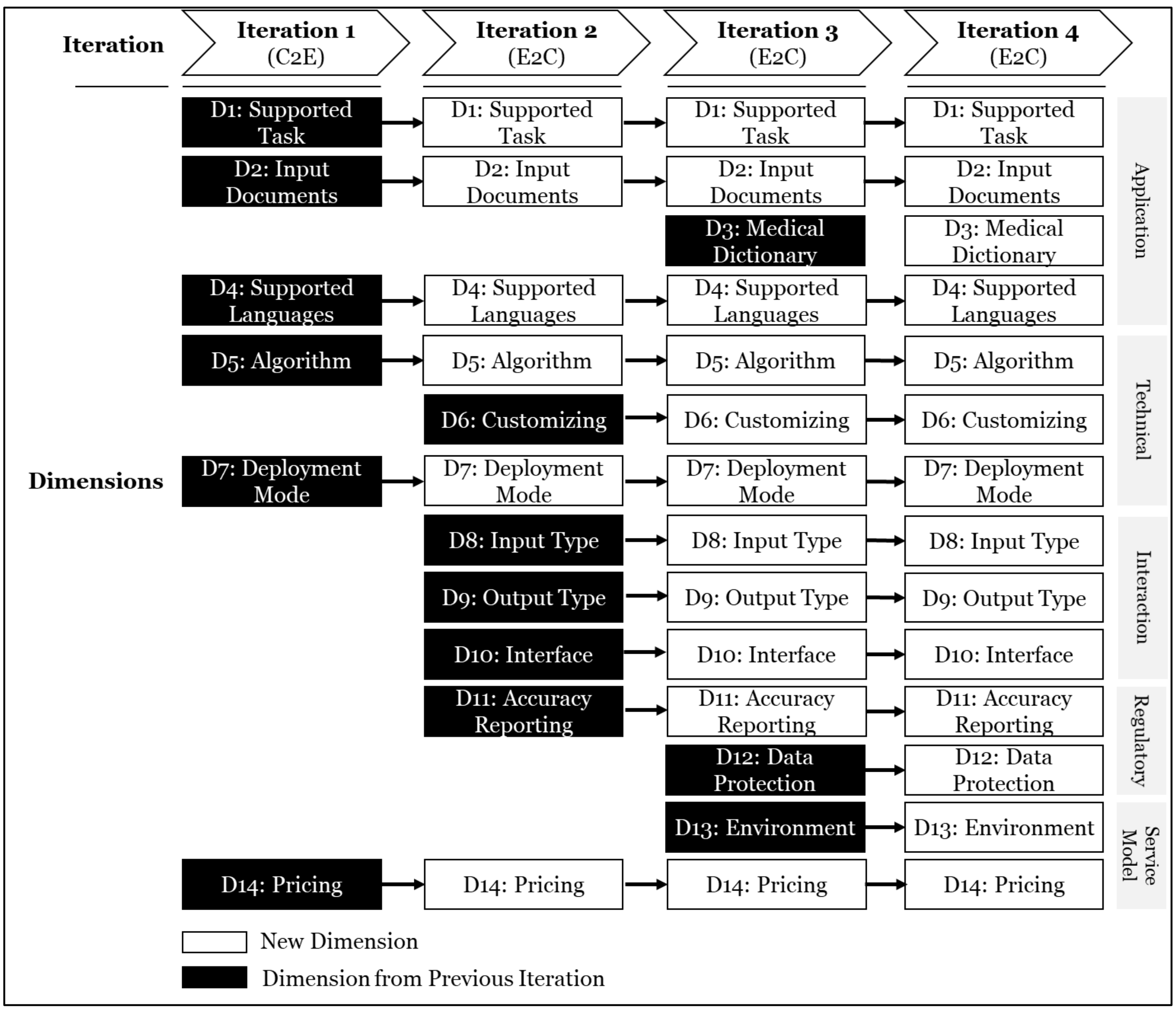
Moreover, the taxonomy could be expanded in several ways. First, the taxonomy was developed during the emergence of GPT. Recent works in the field of GPT in healthcare indicate that such applications will be able to cover a larger range of tasks, while the tasks in the context of information extraction itself remain unchanged (e.g., Yang et al. 2022). Hence, as the technology opens up new levels of performance, the integration into healthcare organizations should be similar to services such as Amazon Comprehend Medical and Microsoft's Text Analytics for Health (both are cloud-based services). Nevertheless, the taxonomy stands open to evaluation through future research that categorizes GPT-based NLP systems and further refines the taxonomy based on the gleaned insights.

Second, we focused on the research field of information systems in healthcare and enriched our database selection with information technology journals. However, we did not include machine learning or computer science conferences as these often focus on the improvement of the underlying algorithms rather than on the integrational aspect. Nonetheless, the taxonomy could be evaluated and enhanced through the inclusion of these outlets. Even though we showed the usefulness of our application with the classification of two sample NLP systems, we have not validated these in concert with decision-makers and stakeholders of healthcare organizations. Future research could qualitatively investigate the importance of the dimensions to either validate or reject them. Moreover, we specifically focused on the processing of medical texts, leaving out other application areas of NLP in this domain such as conversational agents and in general, the domain of natural language understanding. Future research could set other foci in the context of NLP systems in healthcare to derive more insights into their application and possibly provide different important integration dimensions in addition to those presented in this taxonomy. Additionally, decisions made during the iterative scoping process about adding, deleting, and changing dimensions and characteristics were subjective, introducing the caveat that other researchers might create a different taxonomy.

Conclusion

Our taxonomy provides a classification scheme for the integration of NLP systems that process medical texts. Our research revealed the existing gap between NLP which is used for clinical research and clinical practice. Apart from the taxonomy itself, the distribution of characteristics revealed interesting archetypes of systems that help to inform practitioners in decision-making processes. Moreover, we highlight the need to further investigate relevant integration aspects of AI-based NLP applications in daily clinical contexts. With the ubiquity of new revolutionizing technologies such as GPTs and Large Language Models, it is crucial to understand which dimensions are decisive for realizing the benefits of AI in healthcare.

Appendix A – Iterative Development Process of the Taxonomy



Appendix B – Reviewed NLP Applications

Company	Product	Website
Agshealth (AGS1)	Clinical NLP API (NER)	https://docs.ezdi.com/#9849d545-ef9a-453c-b0d8-b1f14ce75683
Agshealth (AGS2)	ICD10-CM service/API / ICD-10-PCS API / CPT API	https://docs.ezdi.com/#32a3ea58-f9a1-4c44-b0c9-31adeefe1cda
Agshealth (AGS3)	PHI API	https://docs.ezdi.com/#32a3ea58-f9a1-4c44-b0c9-31adeefe1cda
Amazon AWS (AMA)	Amazon Comprehend Medical	https://aws.amazon.com/comprehend/medical/?nc=sn&loc=0
Amazon AWS (AMH)	Amazon Healthscribe	https://aws.amazon.com/de/healthscribe/
Anshul et al. (ANS)	HEDEA	https://pubmed.ncbi.nlm.nih.gov/29770248/
Apache (APA)	cTAKES	https://ctakes.apache.org/
Averbis (AVE)	Health discovery	https://averbis.com/health-discovery/
Brigham and Women's Hospital and Harvard Medical School (BRI1)	HITEx	https://www.i2b2.org/software/projects/hitex/hitex_manual.htAI
Brigham and Women's Hospital and Harvard Medical School (BRI2)	MTERMS	https://mterms.bwh.harvard.edu/mterms/
Circlebase (CIR)	Circlebase NLP platform	https://circlebase.com/natural-language-processing/
Clinithink (CLI)	Clix	https://www.clinithink.com/technology
CSIRO (CSI)	Medtex	https://ontoserver.csiro.au/site/our-solutions/medtex/
Dolbey (DOL)	Fusion DocCheck	https://www.dolbey.com/solutions/coding/fusion-doccheck/
Emtelligent (EMT)	emtelliPro NLP	https://www.emtelligent.com/
Eyre et al. (EYR)	medspacy	https://github.com/medspacy/medspacy
foresee medical (FOR)	HCC risk adjustment coding	https://www.foreseemed.com/
Gate (GAT)	BioYODIE Named Entity Disambiguation	https://cloud.gate.ac.uk/shopfront/displayItem/bio-yodie
Georgtown IR Lab (GEO)	QuickUAIS	https://github.com/Georgetown-IR-Lab/QuickUAIS
Google (GOO)	Healthcare Natural Language API	https://cloud.google.com/healthcare-api/docs/concepts/nlp
Google (GOM)	Med-PaLM	https://sites.research.google/med-palm/
Harvard (HAR)	Canary	https://canary.bwh.harvard.edu/
Health Fidelity (HEA)	Lumanent Insights	https://healthfidelity.edifecs.com/technology/
Inovalon (INO)	Nlpaas	https://www.inovalon.com/resource/nlpaas/
Iqvia (IQV)	NLP Platform	https://www.iqvia.com/solutions/real-world-evidence/platforms/iqvia-nlp-platform/
John Snow Labs (JOH)	Spark NLP for Healthcare	https://www.johnsnowlabs.com/spark-nlp-health/
Karandeep Singh (KAR)	clinspacy	https://github.com/AI4LHS/clinspacy
Microsoft (MIC)	Text Analytics for Health	https://learn.microsoft.com/de-de/azure/cognitive-services/language-

		service/text-analytics-for-health/overview?tabs=ner
NLM (NLM)	MetaMap	https://lhncbc.nlm.nih.gov/ii/tools/MetaMap.htAI
OHNLP (OHN1)	MedKAT/P	https://github.com/vamsithotakura/MedKAT
OHNLP (OHN2)	medXN	https://github.com/OHNLP/MedXN
Systrue (SYS1)	SySearch	https://sytrue.com/sysearch/
Systrue (SYS2)	SyReview	https://sytrue.com/syreview-medical-record-reviews/
University of Texas Health Science Center + Melax (MEL)	CLAMP	https://clamp.uth.edu/
WaveHealthTech (WAV)	CodeLogix Plus	https://wavehealthtech.com/doslogix/
Wolterskluwer (WOL)	cNLP	https://www.wolterskluwer.com/en/solutions/health-language/clinical-natural-language-processing

References

Adler-Milstein, J., Zhao, W., Willard-Grace, R., Knox, M., & Grumbach, K. (2020). Electronic health records and burnout: Time spent on the electronic health record after hours and message volume associated with exhaustion but not with cynicism among primary care clinicians. *Journal of the American Medical Informatics Association*, *27*(4), 531–538. **https://doi.org/10.1093/jamia/ocz220**

Al-Marsy, A., Chaudhary, P., & Rodger, J. A. (2021). A Model for Examining Challenges and Opportunities in Use of Cloud Computing for Health Information Systems. *Applied System Innovation*, *4*(1), 15. **https://doi.org/10.3390/asi4010015**

Amazon. (2022). *Amazon Comprehend Medical*. **https://aws.amazon.com/comprehend/medical/?nc1=h_ls**

Apache. (2022). *Apache cTAKES™ - clinical Text Analysis Knowledge Extraction System*. **https://ctakes.apache.org/**

Arndt, B. G., Beasley, J. W., Watkinson, M. D., Temte, J. L., Tuan, W.-J., Sinsky, C. A., & Gilchrist, V. J. (2017). Tethered to the EHR: Primary Care Physician Workload Assessment Using EHR Event Log Data and Time-Motion Observations. *The Annals of Family Medicine*, *15*(5), 419–426. **https://doi.org/10.1370/afm.2121**

Becker, M., & Böckmann, B. (2016). Extraction of UMLS® Concepts Using Apache cTAKES™ for German Language. *Studies in Health Technology and Informatics*, *223*, 71–76.

Berlin, A., Sorani, M., & Sim, I. (2006). A taxonomic description of computer-based clinical decision support systems. *Journal of Biomedical Informatics*, *39*(6), 656–667. **https://doi.org/10.1016/j.jbi.2005.12.003**

Braun, M., Aslan, A., Ole Diesterhöft, T., Greve, M., Benedikt Brendel, A., & Kolbe, L. M. (2022). Just What the Doctor Ordered – Towards Design Principles for NLP-Based Systems in Healthcare. In *The Transdisciplinary Reach of Design Science Research* (pp. 183–194). DESRIST 2022. **https://doi.org/10.1007/978-3-031-06516-3_14**

Campanella, P., Lovato, E., Marone, C., Fallacara, L., Mancuso, A., Ricciardi, W., & Specchia, M. L. (2016). The impact of electronic health records on healthcare quality: a systematic review and meta-analysis. *The European Journal of Public Health*, *26*(1), 60–64. **https://doi.org/10.1093/eurpub/ckv122**

Casey, A., Davidson, E., Poon, M., Dong, H., Duma, D., Grivas, A., Grover, C., Suárez-Paniagua, V., Tobin, R., Whiteley, W., Wu, H., & Alex, B. (2021). A systematic review of natural language processing applied to radiology reports. *BMC Medical Informatics and Decision Making*, *21*(1), 179. **https://doi.org/10.1186/s12911-021-01533-7**

Chen, L., Song, L., Shao, Y., Li, D., & Ding, K. (2019). Using natural language processing to extract clinically useful information from Chinese electronic medical records. *International Journal of Medical Informatics*, *124*, 6–12. **https://doi.org/10.1016/j.ijmedinf.2019.01.004**

Colicchio, T. K., Cimino, J. J., & Del Fiol, G. (2019). Unintended Consequences of Nationwide Electronic

Health Record Adoption: Challenges and Opportunities in the Post-Meaningful Use Era. *Journal of Medical Internet Research*, *21*(6), e13313. **https://doi.org/10.2196/13313**

Dalianis, H. (2018). *Clinical Text Mining*. Springer International Publishing. **https://doi.org/10.1007/978-3-319-78503-5**

Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, *6*(2), 94–98. **https://doi.org/10.7861/futurehosp.6-2-94**

Domaney, N. M., Torous, J., & Greenberg, W. E. (2018). Exploring the Association Between Electronic Health Record Use and Burnout Among Psychiatry Residents and Faculty: a Pilot Survey Study. *Academic Psychiatry*, *42*(5), 648–652. **https://doi.org/10.1007/s40596-018-0939-x**

Fierce Healthcare. (2019). *Global EHR market hits $31B but faces usability, interoperability challenges | Fierce Healthcare*. **https://www.fiercehealthcare.com/tech/global-ehr-market-hits-31-billion-but-faces-usability-interoperability-challenges**

Ford, E., Carroll, J. A., Smith, H. E., Scott, D., & Cassell, J. A. (2016). Extracting information from the text of electronic medical records to improve case detection: a systematic review. *Journal of the American Medical Informatics Association*, *23*(5), 1007–1015. **https://doi.org/10.1093/jamia/ocv180**

Gao, F., Thiebes, S., & Sunyaev, A. (2018). Rethinking the Meaning of Cloud Computing for Health Care: A Taxonomic Perspective and Future Research Directions. *Journal of Medical Internet Research*, *20*(7), e10041. **https://doi.org/10.2196/10041**

Gephart, S., Carrington, J. M., & Finley, B. (2015). A Systematic Review of Nurses' Experiences With Unintended Consequences When Using the Electronic Health Record. *Nursing Administration Quarterly*, *39*(4), 345–356. https://doi.org/10.1097/NAQ.0000000000000119

Google. (2022). *Healthcare Natural Language API*. **https://cloud.google.com/healthcare-api/docs/concepts/nlp**

HealthGovIT. (2021). *National Trends in Hospital and Physician Adoption of Electronic Health Records*. **https://www.healthit.gov/data/quickstats/national-trends-hospital-and-physician-adoption-electronic-health-records**

HHS. (2022). *Methods for De-identification of PHI | HHS.gov*. **https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html**

Iroju, O. G., & Olaleke, J. O. (2015). A Systematic Review of Natural Language Processing in Healthcare. *International Journal of Information Technology and Computer Science*, *7*(8), 44–50. **https://doi.org/10.5815/ijitcs.2015.08.07**

Jensen, K., Soguero-Ruiz, C., Oyvind Mikalsen, K., Lindsetmo, R.-O., Kouskoumvekaki, I., Girolami, M., Olav Skrovseth, S., & Augestad, K. M. (2017). Analysis of free text in electronic health records for identification of cancer patient trajectories. *Scientific Reports*, *7*(1), 46226. **https://doi.org/10.1038/srep46226**

Kaur, A., & Chopra, D. (2016). Comparison of text mining tools. *2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, 186–192. **https://doi.org/10.1109/ICRITO.2016.7784950**

Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, *17*(1), 195. **https://doi.org/10.1186/s12916-019-1426-2**

Koleck, T. A., Dreisbach, C., Bourne, P. E., & Bakken, S. (2019). Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. *Journal of the American Medical Informatics Association*, *26*(4), 364–379. **https://doi.org/10.1093/jamia/ocy173**

Korngiebel, D. M., & Mooney, S. D. (2021). Considering the possibilities and pitfalls of Generative Pre-trained Transformer 3 (GPT-3) in healthcare delivery. *Npj Digital Medicine*, *4*(1), 93. **https://doi.org/10.1038/s41746-021-00464-x**

Kreimeyer, K., Foster, M., Pandey, A., Arya, N., Halford, G., Jones, S. F., Forshee, R., Walderhaug, M., & Botsis, T. (2017). Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review. *Journal of Biomedical Informatics*, *73*, 14–29. **https://doi.org/10.1016/j.jbi.2017.07.012**

Kumar, V., Liao, K., Cheng, S.-C., Yu, S., Kartoun, U., Brettman, A., Gainer, V., Cagan, A., Murphy, S., Savova, G., Chen, P., Szolovits, P., Xia, Z., Karlson, E., Plenge, R., Ananthakrishnan, A., Churchill, S., Cai, T., Kohane, I., & Shaw, S. (2014). NATURAL LANGUAGE PROCESSING IMPROVES PHENOTYPIC ACCURACY IN AN ELECTRONIC MEDICAL RECORD COHORT OF TYPE 2 DIABETES AND CARDIOVASCULAR DISEASE. *Journal of the American College of Cardiology*,

63(12), A1359. **https://doi.org/10.1016/S0735-1097(14)61359-0**
Kundisch, D., Muntermann, J., Oberländer, A. M., Rau, D., Röglinger, M., Schoormann, T., & Szopinski, D. (2022). An Update for Taxonomy Designers. *Business & Information Systems Engineering*, *64*(4), 421–439. **https://doi.org/10.1007/s12599-021-00723-x**
Liberati, E. G., Ruggiero, F., Galuppo, L., Gorli, M., González-Lorenzo, M., Maraldi, M., Ruggieri, P., Polo Friz, H., Scaratti, G., Kwag, K. H., Vespignani, R., & Moja, L. (2017). What hinders the uptake of computerized decision support systems in hospitals? A qualitative study and framework for implementation. *Implementation Science*, *12*(1), 113. **https://doi.org/10.1186/s13012-017-0644-2**
Liu, M., Shah, A., Jiang, M., Peterson, N. B., Dai, Q., Aldrich, M. C., Chen, Q., Bowton, E. A., Liu, H., Denny, J. C., & Xu, H. (2012). A study of transportability of an existing smoking status detection module across institutions. *AMIA ... Annual Symposium Proceedings. AMIA Symposium*, *2012*, 577–586. **http://www.ncbi.nlm.nih.gov/pubmed/23304330**
Lockey, S., Gillespie, N., Holm, D., & Someh, I. A. (2021). A review of trust in artificial intelligence: Challenges, vulnerabilities and future directions. *Proceedings of the Annual Hawaii International Conference on System Sciences*, *2020-Janua*, 5463–5472. **https://doi.org/10.24251/hicss.2021.664**
MetaMap. (2022). *MetaMap*. **https://lhncbc.nlm.nih.gov/ii/tools/MetaMap.html**
Meystre, S. M., Friedlin, F. J., South, B. R., Shen, S., & Samore, M. H. (2010). Automatic de-identification of textual documents in the electronic health record: a review of recent research. *BMC Medical Research Methodology*, *10*(1), 70. **https://doi.org/10.1186/1471-2288-10-70**
Microsoft. (2022). *What is the Text Analytics for health in Azure Cognitive Service for Language? - Azure Cognitive Services | Microsoft Learn*. **https://learn.microsoft.com/en-us/azure/cognitive-services/language-service/text-analytics-for-health/overview?tabs=ner**
Miranda-Escalada, A., Farré-Maduell, E., & Krallinger, M. (2020). *Named Entity Recognition, Concept Normalization and Clinical Coding: Overview of the Cantemist Track for Cancer Text Mining in Spanish, Corpus, Guidelines, Methods and Results*. **https://doi.org/10.5281/zenodo.3773228**
Murff, H. J., FitzHenry, F., Matheny, M. E., Gentry, N., Kotter, K. L., Crimin, K., Dittus, R. S., Rosen, A. K., Elkin, P. L., Brown, S. H., & Speroff, T. (2011). Automated Identification of Postoperative Complications Within an Electronic Medical Record Using Natural Language Processing. *JAMA*, *306*(8). **https://doi.org/10.1001/jama.2011.1204**
Negro-Calduch, E., Azzopardi-Muscat, N., Krishnamurthy, R. S., & Novillo-Ortiz, D. (2021). Technological progress in electronic health record system optimization: Systematic review of systematic literature reviews. *International Journal of Medical Informatics*, *152*, 104507. **https://doi.org/10.1016/j.ijmedinf.2021.104507**
Nehme, F., & Feldman, K. (2021). Evolving Role and Future Directions of Natural Language Processing in Gastroenterology. *Digestive Diseases and Sciences*, *66*(1), 29–40. **https://doi.org/10.1007/s10620-020-06156-y**
Nickerson, R. C., Varshney, U., & Muntermann, J. (2013). A method for taxonomy development and its application in information systems. *European Journal of Information Systems*, *22*(3), 336–359. **https://doi.org/10.1057/ejis.2012.26**
Palvia, P., Lowe, K., Nemati, H., & Jacks, T. (2012). Information Technology Issues in Healthcare: Hospital CEO and CIO Perspectives. *Communications of the Association for Information Systems*, *30*. **https://doi.org/10.17705/1CAIS.03019**
Petitgand, C., Motulsky, A., Denis, J. L., & Régis, C. (2020). Investigating the barriers to physician adoption of an artificial intelligence-based decision support system in emergency care: An interpretative qualitative study. *Studies in Health Technology and Informatics*, *270*, 1001–1005. **https://doi.org/10.3233/SHTI200312**
Pons, E., Braun, L. M. M., Hunink, M. G. M., & Kors, J. A. (2016). Natural Language Processing in Radiology: A Systematic Review. *Radiology*, *279*(2), 329–343. **https://doi.org/10.1148/radiol.16142770**
Puschel, L. C., Roglinger, M., & Brandt, R. (2022). Unblackboxing Smart Things—A Multilayer Taxonomy and Clusters of Nontechnical Smart Thing Characteristics. *IEEE Transactions on Engineering Management*, *69*(5), 2129–2143. **https://doi.org/10.1109/TEM.2020.2988981**
Randorff Højen, A., & Rosenbeck Gøeg, K. (2012). SNOMED CT Implementation. *Methods of Information in Medicine*, *51*(06), 529–538. **https://doi.org/10.3414/ME11-02-0023**
Reyes-Ortiz, J. A., González-Beltrán, B. A., & Gallardo-López, L. (2015). Clinical Decision Support Systems:

A Survey of NLP-Based Approaches from Unstructured Data. *2015 26th International Workshop on Database and Expert Systems Applications (DEXA)*, 163–167. **https://doi.org/10.1109/DEXA.2015.47**

Sohn, S., Kocher, J.-P. A., Chute, C. G., & Savova, G. K. (2011). Drug side effect extraction from clinical narratives of psychiatry and psychology patients. *Journal of the American Medical Informatics Association*, *18*(Supplement_1), i144–i149. **https://doi.org/10.1136/amiajnl-2011-000351**

Spasić, I., Uzuner, Ö., & Zhou, L. (2020). Emerging clinical applications of text analytics. *International Journal of Medical Informatics*, *134*, 103974. **https://doi.org/10.1016/j.ijmedinf.2019.103974**

Stuijfzand, S., Deforges, C., Sandoz, V., Sajin, C.-T., Jaques, C., Elmers, J., & Horsch, A. (2020). Psychological impact of an epidemic/pandemic on the mental health of healthcare professionals: a rapid review. *BMC Public Health*, *20*(1), 1230. **https://doi.org/10.1186/s12889-020-09322-z**

Tremblay, M. C., Berndt, D. J., Luther, S. L., Foulis, P. R., & French, D. D. (2009). Identifying fall-related injuries: Text mining the electronic medical record. *Information Technology and Management*, *10*(4), 253–265. **https://doi.org/10.1007/s10799-009-0061-6**

Velupillai, S., Suominen, H., Liakata, M., Roberts, A., Shah, A. D., Morley, K., Osborn, D., Hayes, J., Stewart, R., Downs, J., Chapman, W., & Dutta, R. (2018). Using clinical Natural Language Processing for health outcomes research: Overview and actionable suggestions for future advances. *Journal of Biomedical Informatics*, *88*, 11–19. **https://doi.org/10.1016/j.jbi.2018.10.005**

Wang, Y., Wang, L., Rastegar-Mojarad, M., Moon, S., Shen, F., Afzal, N., Liu, S., Zeng, Y., Mehrabi, S., Sohn, S., & Liu, H. (2018). Clinical information extraction applications: A literature review. *Journal of Biomedical Informatics*, *77*, 34–49. **https://doi.org/10.1016/j.jbi.2017.11.011**

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, *26*(2), xiii–xxiii. **https://doi.org/10.1.1.104.6570**

Wrenn, J. O., Stein, D. M., Bakken, S., & Stetson, P. D. (2010). Quantifying clinical narrative redundancy in an electronic health record. *Journal of the American Medical Informatics Association*, *17*(1), 49–53. **https://doi.org/10.1197/jamia.M3390**

Yang, X., Chen, A., PourNejatian, N., Shin, H. C., Smith, K. E., Parisien, C., Compas, C., Martin, C., Costa, A. B., Flores, M. G., Zhang, Y., Magoc, T., Harle, C. A., Lipori, G., Mitchell, D. A., Hogan, W. R., Shenkman, E. A., Bian, J., & Wu, Y. (2022). A large language model for electronic health records. *Npj Digital Medicine*, *5*(1), 194. **https://doi.org/10.1038/s41746-022-00742-2**

Zhang, P., & Li, N. (2004). An assessment of human-computer interaction research in management information systems: Topics and methods. *Computers in Human Behavior*, *20*(2), 125–147. **https://doi.org/10.1016/j.chb.2003.10.011**

Zheng, K., Vydiswaran, V. G. V., Liu, Y., Wang, Y., Stubbs, A., Uzuner, Ö., Gururaj, A. E., Bayer, S., Aberdeen, J., Rumshisky, A., Pakhomov, S., Liu, H., & Xu, H. (2015). Ease of adoption of clinical natural language processing software: An evaluation of five systems. *Journal of Biomedical Informatics*, *58*, S189–S196. **https://doi.org/10.1016/j.jbi.2015.07.008**