



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/196474/>

Version: Published Version

---

**Article:**

Zirikly, A., Desmet, B., Newman-Griffis, D. et al. (2022) Information extraction framework for disability determination using a mental functioning use-case. *JMIR Medical Informatics*, 10 (3). e32245. ISSN: 2291-9694

<https://doi.org/10.2196/32245>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

Viewpoint

# Information Extraction Framework for Disability Determination Using a Mental Functioning Use-Case

Ayah Zirikly<sup>1,2,3</sup>, PhD; Bart Desmet<sup>1</sup>, PhD; Denis Newman-Griffis<sup>1,4</sup>, PhD; Elizabeth E Marfeo<sup>1,5</sup>, MPH, PhD; Christine McDonough<sup>1,6</sup>, PhD; Howard Goldman<sup>1,7</sup>, MD, PhD; Leighton Chan<sup>1</sup>, MPH, MD

<sup>1</sup>Rehabilitation Medicine Department, Clinical Center, National Institutes of Health, Bethesda, MD, United States

<sup>2</sup>Whiting School of Engineering, Johns Hopkins University, Baltimore, MD, United States

<sup>3</sup>Malone Center for Engineering in Healthcare, Johns Hopkins University, Baltimore, MD, United States

<sup>4</sup>Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, United States

<sup>5</sup>Department of Occupational Therapy, Tufts University, Medford, MA, United States

<sup>6</sup>School of Health and Rehabilitation Science, University of Pittsburgh, Pittsburgh, PA, United States

<sup>7</sup>Department of Psychiatry, School of Medicine, University of Maryland, Baltimore, MD, United States

**Corresponding Author:**

Ayah Zirikly, PhD

Rehabilitation Medicine Department

Clinical Center

National Institutes of Health

10 Center Drive

Bethesda, MD, 20892

United States

Phone: 1 301 827 6558

Email: [ayah.zirikly@nih.gov](mailto:ayah.zirikly@nih.gov)

## Abstract

Natural language processing (NLP) in health care enables transformation of complex narrative information into high value products such as clinical decision support and adverse event monitoring in real time via the electronic health record (EHR). However, information technologies for mental health have consistently lagged because of the complexity of measuring and modeling mental health and illness. The use of NLP to support management of mental health conditions is a viable topic that has not been explored in depth. This paper provides a framework for the advanced application of NLP methods to identify, extract, and organize information on mental health and functioning to inform the decision-making process applied to assessing mental health. We present a use-case related to work disability, guided by the disability determination process of the US Social Security Administration (SSA). From this perspective, the following questions must be addressed about each problem that leads to a disability benefits claim: *When did the problem occur and how long has it existed? How severe is it? Does it affect the person's ability to work?* and *What is the source of the evidence about the problem?* Our framework includes 4 dimensions of medical information that are central to assessing disability—temporal sequence and duration, severity, context, and information source. We describe key aspects of each dimension and promising approaches for application in mental functioning. For example, to address temporality, a complete functional timeline must be created with all relevant aspects of functioning such as intermittence, persistence, and recurrence. Severity of mental health symptoms can be successfully identified and extracted on a 4-level ordinal scale from absent to severe. Some NLP work has been reported on the extraction of context for specific cases of wheelchair use in clinical settings. We discuss the links between the task of information source assessment and work on source attribution, coreference resolution, event extraction, and rule-based methods. Gaps were identified in NLP applications that directly applied to the framework and in existing relevant annotated data sets. We highlighted NLP methods with the potential for advanced application in the field of mental functioning. Findings of this work will inform the development of instruments for supporting SSA adjudicators in their disability determination process. The 4 dimensions of medical information may have relevance for a broad array of individuals and organizations responsible for assessing mental health function and ability. Further, our framework with 4 specific dimensions presents significant opportunity for the application of NLP in the realm of mental health and functioning beyond the SSA setting, and it may support the development of robust tools and methods for decision-making related to clinical care, program implementation, and other outcomes.

**KEYWORDS**

natural language processing; text mining; bioinformatics; health informatics; machine learning; disability; mental health; functioning; NLP; electronic health record; framework; disability; EHR; automation; eHealth; decision support; functional status; whole-person function

## ***Introduction***

Over the past 2 decades, the use of data-driven informatics techniques to aid in clinical decision-making has increased across the fields of computer science, bioinformatics, and medicine [1]. Natural language processing (NLP), which enables analysis of complex information recorded in narrative text format, has been a key driver of informatics successes in health care. Applications such as automated report analysis for clinical decision support and adverse event monitoring in the electronic health record (EHR) have been widely adopted [2-5]. However, informatics technologies for mental health have consistently lagged because of the complexity of measuring and modeling mental health and illness. The expansion of medical NLP technologies from clinical applications into the realm of complex administrative tasks such as claims evaluations and benefits administration [6,7] has highlighted the need for improved tools for analyzing information on mental health and function, which play a significant role in key health outcomes such as disability [8].

One of the primary goals of NLP in health data is to analyze narrative medical texts such as medical histories, physical examinations, and standardized assessments to extract the data needed to inform decision-making processes. The use of NLP to support these goals for management of mental health conditions has not yet been explored in depth. Our research group develops NLP models to support the information needs of the US Social Security Administration (SSA) disability determination process. Through its disability programs—Social Security Disability Insurance (SSDI) and Supplemental Security Income (SSI)—the SSA is the largest federal provider of financial assistance to workers with disabilities and their families. Because of the impact of functional abilities on both the individuals with disabilities and the society, it is essential that a person's functional abilities are characterized both comprehensively and efficiently in the disability determination process. Multiple sources of information are used to understand a person's ability to work. Given the complexity of the disability determination process, there is interest in developing approaches that enhance the validity of and confidence in the information across sources. While our work is motivated by the SSA's focused use-case, the SSA setting reflects fundamental challenges in the development and broad application of medical informatics technologies. The SSA leverages data from all types of health care providers and EHR systems across the United States. Therefore, informatics tools must be robust to significant heterogeneity in documentation—a known challenge for medical NLP research [9]. The volume of applications for disability benefits that the SSA must process is also extraordinarily high—over 2 million applications every year since 2004 [10]—and informatics tools must therefore support rapid

processing of high-volume data. Finally, and a key motivating factor for our work, the SSA's decision-making processes must incorporate diverse health and function information from all domains of human experience. The SSA setting thus provides an invaluable environment for learning how to translate the potential of NLP tools into practical, reliable tools for real-world applications.

## **Contributions of This Paper**

In this paper, we propose a framework for the advanced application of NLP methods to identify, extract, and organize functioning information to inform the decision-making process applied to assessing functioning and disability. While the framework is applicable to mental and physical functioning use cases alike, this paper focuses on mental functioning. We found no literature that directly addresses our decision-support use-case for mental health and function; therefore, we developed a conceptual framework for synthesizing prior NLP research to create decision support tools for use in assessing SSA's definition of disability. Our framework includes 4 dimensions of medical information that are central to assessing disability—temporal sequence and duration, context, severity, and information source. Findings of this work are intended to inform the development of instruments that will support the decisions of disability adjudicators in the SSA's stepwise process of disability determination and have implications for a broad array of individuals and organizations responsible for assessing mental health function and ability. Further, our framework presents significant opportunity for the application of NLP in the realm of mental and physical health and functioning beyond the SSA setting, and it can support the development of robust tools and methods for decision-making related to clinical care, program implementation, and other outcomes.

## **Background**

The US SSA administers the largest federal assistance programs in the United States, including 2 disability programs: SSDI and SSI. Both programs are based on a statutory definition of disability as the inability to engage in any substantial gainful activity by reason of any medically determinable physical or mental impairment(s), which can be expected to result in death or which has lasted or can be expected to last for a continuous period of not less than 12 months. The SSA's disability determination process is a stepwise process for evaluating individuals according to criteria that operationalize the statutory definition of disability. The process is based on federal regulatory standards that include both financial and medical criteria. Applicants are either allowed or denied at each step or move on to further evaluation in the subsequent steps. The process is administered by state Disability Determination Service agencies. In step 1, applicants are denied if they work and earn

more than the threshold for substantial gainful employment. In step 2, applicants are screened based on whether medical evidence supports the existence of a severe impairment. In step 3, the applicant's medical evidence is compared to codified clinical criteria for various medical impairments, called the Listing of Impairments (Listings). When impairments "meet" or "equal" the Listings, the applicants are allowed. Applicants who are not allowed at step 3 move on to steps 4 and 5, which assess vocational factors such as the "residual functional capacity" of the individual as well as the applicant's age, education, and relevant work experience. In step 4, adjudicators within the Disability Determination Service assess whether the applicant can work in any of their past jobs. If the adjudicator determines that an applicant can work in a previous job, the applicant is denied. Otherwise, in step 5, the Disability Determination Service adjudicators evaluate whether the applicant can perform any work in the national economy. There has been internal effort at the SSA to improve accuracy and timeliness of the disability adjudication process, and external groups have been engaged to assist with this. Expert panels and evaluations of the processes have resulted in recommendations for more systematic integration of functional information into adjudication decisions [11,12].

As part of the adjudication process, adjudicators amass a body of evidence referred to as the Medical Evidence of Record (MER), composed of information collected from multiple sources to characterize a person's potential ability to work. The MER forms the primary resource from which it is determined if an individual meets the SSA's statutory definition of disability. Therefore, the adjudicator must extract a variety of information from the MER, including medical evidence, medical opinion, and lay evidence, to support a decision on disability under this statutory definition.

A primary challenge for accuracy in the disability determination process is that adjudicators must access *all* relevant information from the MER for their decision, including information about both health conditions and functional abilities that relate to work. MER for a single individual may include dozens to hundreds of clinical reports, which imposes a significant burden on the adjudicator to rapidly process extensive medical evidence. Automated analysis of these documents with NLP thus has significant potential to assist adjudicators in the evidence review process and to support efficiency of the process. Our research group has developed novel NLP technologies for automated identification of functional status information in medical evidence [7], thus providing high-coverage retrieval of information related to mobility limitations [13-15] and categorization of this evidence according to the World Health Organization's International Classification of Functioning, Disability and Health [16]. Expansion of these technologies to mental health and function requires adaptation to the conceptual frameworks that characterize mental function, as outlined in the sections that follow.

For the purposes of this paper, we focus on mental health functioning, that is ways in which a person's underlying

cognition, emotions, and behaviors affect their ability to perform daily activities including work tasks and participation, for example, a person's ability to regulate their emotions in stressful situations, multitask, or solve problems. This operationalization is based on the biopsychosocial model of health and function by the World Health Organization's International Classification of Functioning, Disability and Health. In this model, disability results from a gap between a person's underlying ability and the context in which they are performing various activities (eg, work participation [17,18]). This model highlights the fact that diagnostic information is necessary but not sufficient to understand a person's ability to participate in meaningful activities such as employment. In clinical contexts, information on functioning is critical to understanding the impact of conditions on people in their personal and environmental contexts and to develop an effective management plan. Recent work has demonstrated initial feasibility of applying NLP methods to mental health-related topics, including psychiatric readmission and symptoms of severe mental illness (SMI), as well as to mental health and suicide risk within nonclinical texts [19-23]. There is little evidence of the potential for NLP methods to characterize functional and behavioral manifestations of mental health in a person's daily life.

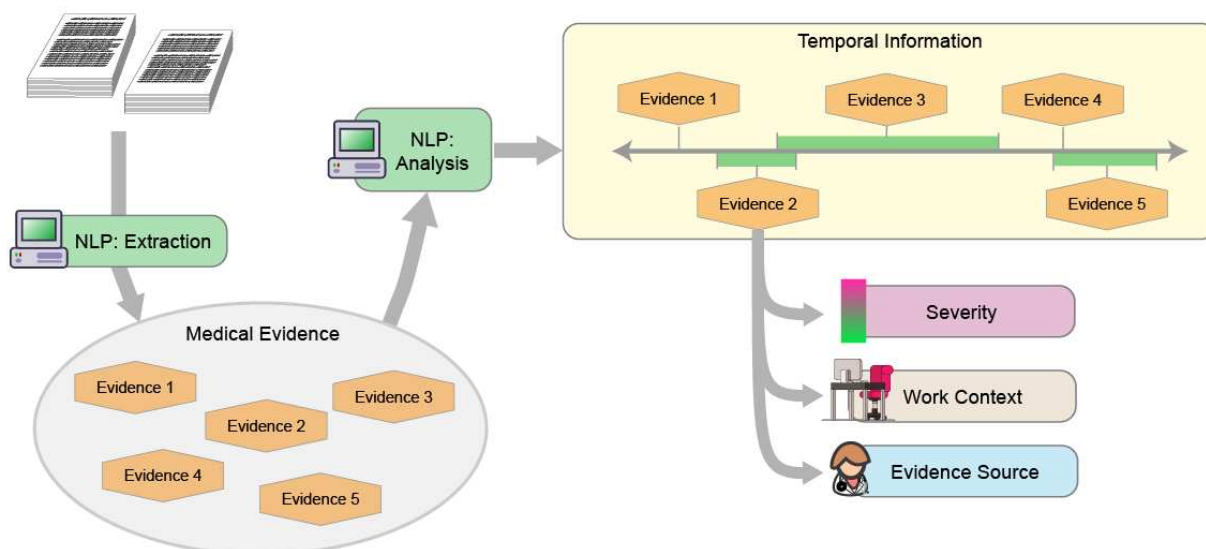
### Information Needs for Analyzing Mental Health and Functioning

For disability adjudication, a wide array of information is needed about the following specific areas of mental functioning that a person uses in a work setting: understanding, remembering, or applying information; interacting with others; concentrating, persisting, or maintaining pace; and adapting and managing oneself. Evidence in the MER may reflect a physical or mental *impairment* that may affect these areas of functioning, an observed *limitation* in one of these areas of functioning, or both. The adjudicator's task is therefore to evaluate the level of severity or degree to which the medically determinable mental impairment affects the 4 areas of mental functioning (ie, limitations) and an individual's ability to function independently, appropriately, effectively, and on a sustained basis.

Thus, adjudicators must organize and synthesize the medical evidence in the following 4 distinct dimensions to understand the trajectory of mental function in an individual and its impact on work capacity: temporal information including sequence and duration, severity of extracted mentions of functioning, the context with respect to work and work-related information, and the source of the information.

We envision the use of NLP technologies to transform raw evidence found in medical records (MER in the SSA context) into a structured presentation of evidence illustrating each of these 4 factors to SSA adjudicators. Figure 1 illustrates the conceptual structure of such an analytic pipeline. Evidence is first extracted from the MER documents and then ordered into a temporal sequence, with each piece of evidence annotated with the severity of impact on function, the relevant work context, and the source of the evidence.

**Figure 1.** Conceptual pipeline for analysis of information on function in medical records. NLP: natural language processing.



The remainder of this paper describes the existing NLP literature related to these 4 tasks and highlights how each can be addressed in the area of mental health and function.

## Methods

### Literature Search

We conducted a scoping review of NLP approaches, models, and methods to characterize functional status in free text in the biomedical, clinical, mental health, and disability domains from 1994 to 2021. We searched Google Scholar, which indexes not only PubMed but also conferences and workshops that may be relevant to our scope of interest such as the Association for Computational Linguistics (ACL) conferences, BioNLP, Clinical NLP, and CLPsych. Our search yielded a small number of publications in special workshops, such as AI4Function, that discuss the extraction of physical function information (eg, mobility) but do not address mental health function. We did not find any articles in our area of focus. To expand our search, we used keywords in Google Scholar related to the 4 dimensions of our proposed framework to find articles that describe approaches relevant to each dimension but an area different from functioning. Examples of keywords used include “temporal ordering,” “clinical temporal ordering,” “event extraction,” “NLP and mental health,” “symptom severity,” “environmental context,” “personal context,” and “author attribution.”

### Findings From Existing Work

Disability in the SSA context is defined as the inability to engage in substantial gainful activity for at least 12 months because of a physical or mental impairment and is assessed both in terms of the *trajectory* of a person’s function and the *context* of how it relates to work. As the disability adjudication process also involves collecting MER data from a variety of providers, it is critical to understand how different pieces of evidence relate to one another in terms of the perspective of the information’s *source* (eg, the disability claimant, a medical professional with an established relationship with the claimant, an outside

consultant). Thus, for each piece of evidence in the MER, an adjudicator must be able to answer the following 4 questions: *When and for how long was an impairment or associated limitation true? How severe or intense is the impairment or limitation? Does the impairment or limitation affect work? and Who reported the impairment or limitation and how convincing is it as evidence?*

In this paper, we present an overview of relevant NLP research and methodologies that can help the adjudicator extract relevant information for these 4 questions. However, it is important to note that building solutions to address these 4 questions requires the ability to identify mental impairments either manually or via automated algorithms. In this paper, we choose to focus on addressing the 4 dimensions or questions only and assume that information on mental impairment is available. We justify our choice by the following factors: the availability of extraction systems that, given annotated data, can extract mentions of mental health impairments with high confidence in EHRs [24] and clinical text [21] and can extract these mentions using available International Classification of Diseases codes; and the novelty and urgency of the proposed 4 dimensions and the lack of available studies to address them. The mentions include observations such as “The patient was not able to concentrate on the given tasks for more than 5 minutes during the exam.”

### Temporal Information

Temporal information includes duration and temporal sequencing. In our use-case, the SSA’s statutory definition of disability requires specific definition and sequential information. The disability is due to a mental impairment, so the impairment must precede the functional limitation.

Temporal sequencing or temporal reasoning has been an active research area in NLP and data mining for a long time. Its importance comes from its applicability to many tasks such as summarization [25], question answering [26], and medical informatics [27]. Given the similarity of the medical utilization task to our use-case, we will mainly focus on reviewing NLP

techniques developed in that area and how we can integrate them into our framework.

Temporal reasoning usually includes the following aspects: identifying the targeted events for the task (eg, treatments, diagnosis, symptoms, or medications); and defining time in a machine-readable way that is relevant to the domain and task and extracting temporal information related to the targeted events (eg, in a medical informatics setting, we care about the duration of symptoms or frequencies of medications) [28].

In the context of mental health functioning, the events of interest are events related to mental health conditions and impairment, which can be separated into the following 3 distinct categories: *persistent*, the impairment continues to exist over a prolonged time without interruptions of some criterion duration; *intermittent*, the impairment occurs at irregular intervals and is observed in a temporal sequence with interruptions greater than the criterion duration that defines persistent; and *recurring*, the impairment occurs periodically or repeatedly. Although we can think of recurring as a special case of intermittent, a recurrent mental functioning event is observed again after a period of some specified duration that is longer than the minimum duration defining intermittent.

While this list mainly focuses on the disability use-case, it presents a framework for researchers to structure their problem using all or a subset of our temporal formalization based on the targeted use-case, task, and domain. Our suggested framework differs from other NLP techniques for temporal sequencing because we need to consider nuances that accompany the mention of temporal information. For instance, a sentence such as “The patient reports having lack of interest mainly during the morning hours when it is the weekend,” suggests the need for a system that can highlight the time: weekend morning hours, associated with lack of interest.

Although we introduce a slightly different framework for temporal sequencing, existing NLP methods can be applied to mental health functioning, especially given that most time expressions in medical notes are in the format of date and frequency (eg, how many times per week/day). For instance, temporal recognition and reasoning have played a significant role in information extraction systems [28-30]. Denny et al [29] developed a system that identifies the temporal information and status of colonoscopy events within EHRs with high precision and recall (>.9). In the area of mental health, Viani et al [28] focused on temporal expression extraction to help estimate the duration of untreated psychosis. The temporal information extraction helps in identifying in EHRs when the psychosis symptoms started (onset) and when the treatment was first initiated. Examples of temporal expressions from the paper include mentions such as “started hearing voices at the age of 16, these hallucinations were not elicited during today’s exam.” This is highly relevant to our use cases and to identify the 3 temporal formalizations of persistent, intermittent, and recurring.

To build NLP systems that can identify temporal information, the availability of annotated data sets with temporal information is critical. Although such data sets outside the clinical and medical domains have been publicly available and more easily accessible, such as the ACE 2005 Multilingual Training Corpus

[31], the clinical domain imposes more limitations, especially mental health, given the sensitivity of this information and privacy concerns. Examples of annotated data sets for temporal reasoning in clinical text are THYME corpus [32], where 1254 deidentified oncology notes from the Mayo Clinic have been annotated using the ISO-TimeML specification [33]. Sun et al [34] introduced one of the most popular data sets for temporal reasoning in their i2b2 data set that contains 310 discharge summaries. In both these data sets, the focus is on 4 time annotation categories: date, includes both actual dates in addition to mentions such as yesterday and tomorrow and duration, frequency, and time (eg, 3 PM, in the afternoon).

In another data genre, but within the area of mental health, there have been efforts to introduce temporally annotated data sets such as RSDD-Time [35]. This data set is extracted from social media posts that focus on self-reported patients who are diagnosed as having depression. The annotation includes temporal information relevant to when the diagnosis occurred and if the condition is still present.

Given our use-case, we believe that the i2b2 annotation scheme would serve our goals for identifying when the impairment or symptoms occurred and determining for how long the symptoms or impairments lasted.

With regard to methods, researchers used a variety of machine learning techniques such as logistic regression that is especially effective for a small training sample size of less than 500 [36]. Recent advances in the contextualized embeddings [37,38] improved the performance of NLP tasks, including temporal ordering in the clinical domain [39-41]. For instance, Med-BERT [42], a language model that is trained and fine-tuned on the EHR data set, yields a performance that is comparable to that of deep learning techniques on data sets that are almost 10 times larger.

## Severity

The severity of a symptom or functional limitation is an important factor for psychological assessments and psychometric benchmarks, where it is often recorded using a 4-level ordinal scale or as a score that is discretized into that scale. A typical scale includes absent, mild, moderate, and severe labels [43]. The latter 3 labels are frequently employed for the disorders described in the Diagnostic and Statistical Manual of Mental Disorders Text Revision Fourth Edition (DSM-IV-TR), which permit severity specifiers. An advantage of this scale is that mental health clinicians and laypeople alike readily understand it, and it has been adopted in computational approaches for severity classification as well.

Filannino et al [44] describe an NLP shared task focused on symptom severity prediction in neuropsychiatric evaluation records with an exclusive focus on positive valence events, objects, or situations that are harmful but attractive to patients to the point that they are actively engaged despite the consequences. Positive valence is classified on the aforementioned 4-level scale at the patient level and assesses lifetime maximum severity. As such, this task differs from our approach in that it is not time dependent or resolved at the individual mention level. Filannino et al [44] report that in this

relaxed use-case, the task can be accomplished automatically with close to human performance.

Severity classification has also been actively researched in suicide risk assessment for patients and individuals on social media. For instance, Shing et al [45] and Zirikly et al [46] introduce an annotated Reddit data set for users with and without depression, each of whom received a suicide risk assessment score on a 4-level scale (none, low, moderate, high). Zirikly et al [46] organized a shared task for advanced automatic user suicide risk classification and provided baseline systems using deep learning models (eg, convolutional neural network) and machine learning models that require feature engineering. Examples of features that are commonly used in NLP methods for the mental health domain and emotion detection and classification are n-grams, lexicons such as the Linguistic Inquiry and Word Count [47], and emotion-word dictionaries [48], topic models, and Reddit usage metafeatures. Top-ranked systems could distinguish between low-risk and high-risk users, but fine-grained 4-level scale classification results indicate the need for further research.

Jackson et al [21] introduced an annotated data set for SMI using clinical text from the Clinical Record Interactive Search system in a cohort of 18,761 patients with SMI and 57,999 individuals without SMI. The authors used a support vector machine model to extract symptoms associated with SMI from discharge summaries. While the data and model for this task are relevant for the severity classification use-case, it does not address severity classification directly.

We can conclude that no severity classification models currently exist for mental health signs and symptoms, but there is a growing body of work on severity classification at the patient level. For clinical symptoms more broadly, Koleck et al [49] performed a systematic review of NLP approaches for processing symptoms in free-text EHR narratives. They found that out of 14 studies, the large majority used documentation occurrence or frequency of occurrence to investigate symptoms, and symptom severity was explicitly evaluated in only 1 study: Heintzelman et al [50] used NLP on oncology provider encounter notes to classify the severity of cancer patients' pain symptoms into no, some, controlled, or severe pain. Koleck et al [49] report accurate extraction of symptom severity with location and duration as important directions for future work on EHR NLP algorithms.

From our literature review, we find that for both mental and physical health, there is ample opportunity for novel work on severity classification of symptoms and functioning and for continued efforts at the patient level.

## Context

The context in which a functional impairment or limitation is experienced or observed is critical to understanding its impact on work-related activities. Functional activity is an outcome of the interaction between an individual (including physical or cognitive impairments in addition to personal identities and preferences) and their physical, social, and cultural environment [51]. Characterization of this multidimensional relationship between environment, personal factors, and functioning is thus

highly complex, as reflected by the wide variety of strategies used to capture contextual information in functional measurement [52]. Two themes have emerged in prior literature that make a useful distinction between different types of contextual factors: *social context* (ie, social determinants of health), broader characteristics of an individual's social situation such as socioeconomic status, education, zip code, and race and ethnicity identifiers, which inform available resources and opportunities for activity [53]; and *individual context*, factors that are more specific to an individual's activity performance, such as the physical environment for an activity, social roles such as work requirements, and personal preferences such as transportation access or personal values.

While research on social determinants of health has grown rapidly [54-57] due in part to their strong correlation with population-level health outcomes [58], research on individual context and environment—which more directly impacts functioning [59,60]—remains a significant challenge. Conceptual frameworks of disability have grown to recognize the role of both environmental factors and personal factors in functional outcomes, as seen in the World Health Organization's International Classification of Functioning, Disability and Health. Measures have been developed to characterize environmental factors of function, including physical [61] and psychosocial environment [62]. Such measures can be highly informative regarding functional outcomes [63]. However, they are not systematically used in clinical contexts [64,65] and some work-related aspects of environment remain underspecified even in conceptual models [66]. Functional assessment measures, on the other hand, frequently either control for environment (as in standardized performance measures) or embed environmental characteristics directly into the measurement of function [67,68] rather than capturing them as related variables. In either case, the details of a person's environment and its role in their functional outcomes are difficult to extract reliably.

Environmental factors are only one part of the contexts in which people function and must be combined with information on personal factors affecting functional outcomes. Two recent studies have developed steps toward systematically capturing personal values and capabilities to inform rehabilitative care for older adults, though automated analysis of this information remains a future direction [69,70]. Individual context is a largely unexplored area for NLP research due in part to the novelty of human functioning as an area of NLP application [71]. In an initial foray, Agaronnik et al [72] used NLP to capture wheelchair usage—which, as an assistive device, may be considered a contextual factor affecting functional outcomes—from clinical data and demonstrated clear utility of this information over structured billing and diagnosis codes alone. More broadly, the flexibility of free text and the availability of NLP tools to analyze it offers greater freedom for recording and analyzing information on salient contextual factors when the full power of more robust but burdensome environmental measures is not needed. We therefore highlight individual context, where social context meets individual activity, as a key direction for future NLP research to enable mental health and function analysis.

## Source Attribution

In the context of the SSA, disability claims can include the following sources of information (Code of Federal Regulations, SSA): objective medical evidence, medical opinions, and lay evidence.

Objective medical evidence includes signs and laboratory tests reported by recognized medical sources. It is characterized by being quantifiable and discernable. This is highly important and indicative of the intensity and persistence of the symptoms and their impact (eg, how pain severity can affect work ability). Medical opinions include relevant information received from both medical and nonmedical sources. Examples of such information are daily activity and other factors relevant to functional limitations caused by pain or symptoms; location, duration, frequency, and intensity of pain or symptoms; and treatment or any other medication used to alleviate the pain or symptoms. Lay evidence consists of information outside of objective medical evidence or medical opinions—assessments of disability or functioning limitation provided by knowledgeable nonmedical sources such as family, teachers, social services personnel, and employers. This will complement the information provided in objective medical evidence and

medical opinions to better understand the impairments from multiple perspectives. Moreover, lay evidence is very insightful and important when medical evidence does not provide enough evidence for symptoms [73]. As we note, these types of evidence carry different levels of authority and support for the symptoms of the patient. Therefore, the adjudicators need to evaluate and address each evidence separately given its source to make a more comprehensive decision for disability eligibility. In this section, we will start with an overview of related work in NLP, followed by our recommendations to customize these efforts or build on them to address the needs to source attribution within our proposed framework.

Source attribution, as proposed, correlates with multiple similar tasks in NLP. We will start with an example to showcase options for NLP techniques we can adopt from: *The patient lacks interest in doing anything, his mom mentioned. When the doctor asked the patient, he claimed that he goes to work most of the time. At the end of the visit, the doctor diagnosed the patient with depression based on multiple assessments.* First, as we mentioned previously, we are assuming the availability of an extraction system that can identify and recognize mental health functioning and diagnoses statements. Table 1 depicts the mention, its source, and type of evidence.

**Table 1.** Examples of different mental health functioning.

Mention	Source	Type
The patient lacks interest in doing anything	Mother	Lay evidence (symptom)
He claimed that he goes to work most of the time	Patient	Medical opinion (daily activity)
The doctor diagnosed the patient with depression based on multiple assessments	Doctor	Objective medical evidence (diagnosis)

Identifying who made the statements is similar to the task of identifying author attribution in a dialogue or quoted speech [74]. Although regular expressions can capture simple cases of source attribution of impairments, such as “The patient said,” Pareti et al [75] and O’Keefe et al [76] discussed more advanced techniques for quotes—direct and indirect—attribution in opinion mining. Although they show promising results, these methods have been geared and tested on newswire data. All these techniques require clean and well-structured data, an assumption that is hard to meet, especially given the noise presented in clinical notes [77].

There are some cases in which the doctor or medical expert omits mentioning the source in the note, especially when the observation is generated from them or another medical expert. In such cases, inferring source is difficult as it is not explicitly known or inferred (using coreference resolution). For these scenarios, we suggest that techniques be adopted from author attribution task. This task focuses on identifying the author of a text. This task has been well studied in multiple applications; the most traditional one is assigning anonymous literary to authors [78,79]. Additionally, it has been used in forensics to identify authors that are involved in internet-based activity in different text genres such as online messaging (eg, emails) [80], news text data set [81], and social media [82,83]. However, in our work, we focus on attribution of short text or sentences in notes.

Furthermore, we believe that it would be beneficial to adopt techniques from the intersection of author attribution and coreference resolution [84,85]. We see similarities with event extraction, where we focus on event attributes, mainly participants, when the event describes a mental health functioning mention. Techniques to extract multiple accounts from a narrative, such as the ones described by Zhang et al [86], can be adopted in our work to identify who made the observation or statement.

It is important to note that the attribution problem, as we propose it, requires systems that can identify mental health functioning mentions (eg, depression, lack of interest). For that goal, availability of annotated data to train and test machine learning systems is essential [87]. Although we earlier addressed the need for annotated data sets that have labels for mental health functioning, we need additional labels for the source attribution problem. The labels need to address who the source is and the type of evidence. However, it is worth pointing out that data sets that have labels solely on the level of evidence are sufficient for our targeted use-case. Labels can be assigned from the 3 types of evidence we mentioned above.

## Discussion

### Applications Beyond Disability Adjudication

We have presented a framework to support extraction of functional information in mental health, which includes 4 main

dimensions. There were no existing NLP applications that are directly applied to the characterization of temporality, severity, source, and context. However, we identified relevant work in mental health and other areas that could be used for the advanced application of NLP in the field. Temporal expression extraction and a relevant annotation scheme for identifying onset and duration were presented. A model for extracting severity of functional limitations was presented based on existing ordinal symptom severity ratings. An example of extraction of context was provided based on specific cases of wheelchair use in clinical settings. Finally, alternatives for source attribution were identified among existing approaches.

While our framework is tailored to the SSA disability benefits adjudication process, it has implications for a wide variety of applications outside the SSA context. For example, this approach could be used when extracting information for use in the medical case review process. This process requires expert review of patients' care history based on medical records to ensure that the treatment provided meets Medical Necessity Criteria. Additionally, this framework may be useful for other review activities, including informing the process for assessing eligibility determinations, individualized education programs, and educational placements for children under the Individuals with Disabilities Education Act. For this paper, we briefly highlight applications for mental health informatics research, functional assessment and program management in the health care setting, and consultations for case-based recommendations in treatment and managed care.

There is significant untapped potential for informatics technologies focusing on mental health and functioning, as evidenced by the interest of the mental health informatics research community in recent years. The use of informatics technologies has grown for the detection and diagnosis of mental health conditions [88], and the use as tools in mental health care delivery is beginning to be explored [89]. Our framework can inform the expansion of these technologies into a longer-term view of the trajectory of mental health and functioning in a person, thus improving the power of predictive analytics and presentation of health information to providers. Murnane et al [90] describe several technological needs for long-term mental health management, including incorporation of social contexts—a key component of our framework for NLP. Rigby et al [91] identified several aspects of mental health care that are still challenging for mental health informatics 2 decades later, including the importance of a longitudinal view. By identifying clear links to existing NLP research, our framework can serve to guide translational NLP research [92] in the mental health domain. This work can help identify both processes for translating existing NLP technologies into robust solutions for application in mental health research and care as well as new research questions for progress on the needs of mental health informatics.

There are numerous potential specific applications of our approach to using NLP for extraction of information from various sources to assist with the assessment of mental functioning. These are applications that require a review of medical and health-related information to assess functioning for the support of various clinical and other human service

processes. The most common application would be reviews of clinical records to decide on a diagnosis or a course of treatment. A similar approach might be used by a managed care organization to determine the medical necessity of an episode of care or receipt of service. A consultant who is involved in the second opinion on a diagnosis or treatment plan could benefit from a decision-support tool that extracts all the information in a medical record that is relevant to mental functioning. Outside of health care settings, educational organizations and child welfare organizations might use such a clinical review to assess a student's need for special assistance or accommodation based on impairment in mental functioning. The development of an Individual Education Plan or a 504 plan [93] could use an NLP support tool to extract information from school and medical records to assess the need for special supports.

It is worth noting that this framework and its 4 key elements can be used for and generalized to any area of functioning within the SSA disability program and its statutory definition of work disability.

### **Support Tools for Disability Adjudication Need High Sensitivity**

Current tools available to extract data related to mental health and function lack the level of sensitivity with respect to the elements in the MER on many types of mental functioning due to a mental impairment. While adjudicators ultimately need information on more fine-grained aspects of temporal sequencing using constructs such as intermittence, persistence, and recurrence, the main challenge is to create a complete timeline with all relevant aspects of functioning. Human decision makers assess the more fine-grained aspects of these characteristics of functioning. NLP systems thus need to extract the information without necessarily making fine-grained distinctions. One needs to know all the fine-grained elements to extract all relevant information even if the NLP tool does not need to make the distinctions. What is true for the granularity of temporality is also true for context, severity, and source, but most importantly, an NLP tool needs to be sensitive so that no information in the MER is overlooked.

### ***Limitations and Challenges With Respect to NLP for Mental Health Function***

Although the domain of mental health in general is attracting more NLP research, these studies focus on classification tasks in terms of diagnosis or identification of high-risk individuals and do not address how these impairments affect the patient's functioning in both personal and work environments. Thus, there will be challenges and obstacles as research evolves in this domain.

As we described in the paper, the domain of mental health in general and especially mental health functioning is ambiguous and highly semantic. This yields to different interpretations and inconsistencies in annotating documents with mental health functioning mentions and attributes, as consensus by humans is harder to attain. Furthermore, the lack of gold standard and manually annotated corpora for mental health functioning that are essential to build robust extraction solutions highlights the

need for the interested community to invest resources in building such corpora to further improve the performance of these solutions.

Although precision, specificity, and sensitivity (ie, recall) metrics are important, we believe that the interested entities, such as the SSA, can directly benefit from tools that focus on sensitivity (ie, higher recall) rather than higher precision and specificity. Such premise extends the invitation for research in categorization and relevance ranking to compensate for the low specificity and precision of such systems. Although we are aware of the importance of that line of research, this paper leaves this work for the future.

## Conclusions and Future Vision

There is tremendous opportunity for the development and application of NLP tools and methods for the characterization of mental functioning. Although we found no literature that directly applied to the 4 main dimensions in the proposed framework, relevant tools and methods were identified. Research and development leveraging this existing work to tailor approaches for the extraction of temporality, severity, source, and context will yield substantial value to the use-case of disability determination and beyond. Future work should focus on developing relevant annotated data sets and tools trained on the key aspects of the 4 mental functioning domains.

## Acknowledgments

This research is supported by the Intramural Research Program of the National Institutes of Health and the US Social Security Administration. The views expressed in this paper are those of the authors and do not necessarily reflect the official policy or position of the National Institutes of Health or the US government.

## Conflicts of Interest

None declared.

## References

- Ohno-Machado L. Realizing the full potential of electronic health records: the role of natural language processing. *J Am Med Inform Assoc* 2011 Sep 01;18(5):539-539 [FREE Full text] [doi: [10.1136/amiajnl-2011-000501](https://doi.org/10.1136/amiajnl-2011-000501)] [Medline: [21846784](https://pubmed.ncbi.nlm.nih.gov/21846784/)]
- Spyns P. Natural language processing in medicine: an overview. *Methods Inf Med* 1996 Dec;35(4-5):285-301. [Medline: [9019092](https://pubmed.ncbi.nlm.nih.gov/9019092/)]
- Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? *J Biomed Inform* 2009 Oct;42(5):760-772 [FREE Full text] [doi: [10.1016/j.jbi.2009.08.007](https://doi.org/10.1016/j.jbi.2009.08.007)] [Medline: [19683066](https://pubmed.ncbi.nlm.nih.gov/19683066/)]
- Young IJB, Luz S, Lone N. A systematic review of natural language processing for classification tasks in the field of incident reporting and adverse event analysis. *Int J Med Inform* 2019 Dec;132:103971. [doi: [10.1016/j.ijmedinf.2019.103971](https://doi.org/10.1016/j.ijmedinf.2019.103971)] [Medline: [31630063](https://pubmed.ncbi.nlm.nih.gov/31630063/)]
- Mendonça EA, Haas J, Shagina L, Larson E, Friedman C. Extracting information on pneumonia in infants using natural language processing of radiology reports. *J Biomed Inform* 2005 Aug;38(4):314-321 [FREE Full text] [doi: [10.1016/j.jbi.2005.02.003](https://doi.org/10.1016/j.jbi.2005.02.003)] [Medline: [16084473](https://pubmed.ncbi.nlm.nih.gov/16084473/)]
- Popowich F. Using text mining and natural language processing for health care claims processing. *SIGKDD Explor News* 2005 Jun;7(1):59-66. [doi: [10.1145/1089815.1089824](https://doi.org/10.1145/1089815.1089824)]
- Desmet B, Porcino J, Zirikly A, Newman-Griffis D, Divita G, Rasch E. Development of Natural Language Processing Tools to Support Determination of Federal Disability Benefits in the U.S. In: *Proceedings of the 1st Workshop on Language Technologies for Government and Public Administration (LT4Gov)*.: European Language Resources Association; 2020 Presented at: Language Resources and Evaluation Conference; 05/16/20; Marseille p. 1-6 URL: <https://aclanthology.org/2020.lt4gov-1.1/>
- Courtney-Long EA, Carroll DD, Zhang QC, Stevens AC, Griffin-Blake S, Armour BS, et al. Prevalence of disability and disability type among adults--United States, 2013. *MMWR Morb Mortal Wkly Rep* 2015 Jul 31;64(29):777-783 [FREE Full text] [doi: [10.15585/mmwr.mm6429a2](https://doi.org/10.15585/mmwr.mm6429a2)] [Medline: [26225475](https://pubmed.ncbi.nlm.nih.gov/26225475/)]
- Carrell DS, Schoen RE, Leffler DA, Morris M, Rose S, Baer A, et al. Challenges in adapting existing clinical natural language processing systems to multiple, diverse health care settings. *J Am Med Inform Assoc* 2017 Sep 01;24(5):986-991 [FREE Full text] [doi: [10.1093/jamia/ocx039](https://doi.org/10.1093/jamia/ocx039)] [Medline: [28419261](https://pubmed.ncbi.nlm.nih.gov/28419261/)]
- Social Security Administration. URL: <https://www.ssa.gov/oact/STATS/table6c7.html> [accessed 2022-02-18]
- Stobo JD, McGeary M, Barnes DK, editors. *Improving the social security disability decision process*. Washington, DC: The National Academies Press; 2007.
- Brandt DE, Houtenville AJ, Huynh MT, Chan L, Rasch EK. Connecting contemporary paradigms to the Social Security Administration's disability evaluation process. *J Disabil Policy Stud* 2011 Feb 07;22(2):116-128. [doi: [10.1177/1044207310396509](https://doi.org/10.1177/1044207310396509)]
- Newman-Griffis D, Zirikly A. Embedding Transfer for Low-Resource Medical Named Entity Recognition: A Case Study on Patient Mobility. In: *Proceedings of the BioNLP 2018 workshop*.: Association for Computational Linguistics; 2018

- Presented at: Annual Meeting of the Association for Computational Linguistics; 07/19/18; Melbourne, Australia p. 1-11 URL: <https://aclanthology.org/W18-2301/> [doi: [10.18653/v1/w18-2301](https://doi.org/10.18653/v1/w18-2301)]
14. Newman-Griffis D, Fosler-Lussier E. HARE: a Flexible Highlighting Annotator for Ranking and Exploration. In: Proc Conf Empir Methods Nat Lang Process.: Association for Computational Linguistics; 2019 Presented at: Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP); 11/2019; Hong Kong, China p. 85-90 URL: <https://aclanthology.org/D19-3015> [doi: [10.18653/v1/d19-3015](https://doi.org/10.18653/v1/d19-3015)]
  15. Newman-Griffis D, Zirikly A, Divita G, Desmet B. Classifying the reported ability in clinical mobility descriptions. In: Proceedings of the 18th BioNLP Workshop and Shared Task.: Association for Computational Linguistics; 2019 Aug Presented at: Annual Meeting of the Association for Computational Linguistics; 07/2019; Florence, Italy p. 1-10 URL: <https://aclanthology.org/W19-5001> [doi: [10.18653/v1/w19-5001](https://doi.org/10.18653/v1/w19-5001)]
  16. Newman-Griffis D, Fosler-Lussier E. Automated coding of under-studied medical concept domains: linking physical activity reports to the International Classification of Functioning, Disability, and Health. *Front Digit Health* 2021 Mar 10;3:620828 [FREE Full text] [doi: [10.3389/fdgth.2021.620828](https://doi.org/10.3389/fdgth.2021.620828)] [Medline: [33791684](https://pubmed.ncbi.nlm.nih.gov/33791684/)]
  17. Marfeo EE, Haley SM, Jette AM, Eisen SV, Ni P, Bogusz K, et al. Conceptual foundation for measures of physical function and behavioral health function for Social Security work disability evaluation. *Arch Phys Med Rehabil* 2013 Sep;94(9):1645-1652.e2 [FREE Full text] [doi: [10.1016/j.apmr.2013.03.015](https://doi.org/10.1016/j.apmr.2013.03.015)] [Medline: [23548543](https://pubmed.ncbi.nlm.nih.gov/23548543/)]
  18. Chan F, Gelman J, Ditchman N, Kim JH, Chiu CY. The World Health Organization ICF model as a conceptual framework of disability. In: *Understanding Psychosocial Adjustment to Chronic Illness and Disability: A Handbook for Evidence-Based Practitioners in Rehabilitation*. NY: Springer; Jan 2009:23-49.
  19. Rumshisky A, Ghassemi M, Naumann T, Szolovits P, Castro VM, McCoy TH, et al. Predicting early psychiatric readmission with natural language processing of narrative discharge summaries. *Transl Psychiatry* 2016 Oct 18;6(10):e921-e921 [FREE Full text] [doi: [10.1038/tp.2015.182](https://doi.org/10.1038/tp.2015.182)] [Medline: [27754482](https://pubmed.ncbi.nlm.nih.gov/27754482/)]
  20. Kjell ONE, Kjell K, Garcia D, Sikström S. Semantic measures: Using natural language processing to measure, differentiate, and describe psychological constructs. *Psychol Methods* 2019 Feb;24(1):92-115. [doi: [10.1037/met0000191](https://doi.org/10.1037/met0000191)] [Medline: [29963879](https://pubmed.ncbi.nlm.nih.gov/29963879/)]
  21. Jackson RG, Patel R, Jayatilleke N, Kolliakou A, Ball M, Gorrell G, et al. Natural language processing to extract symptoms of severe mental illness from clinical text: the Clinical Record Interactive Search Comprehensive Data Extraction (CRIS-CODE) project. *BMJ Open* 2017 Jan 17;7(1):e012012 [FREE Full text] [doi: [10.1136/bmjopen-2016-012012](https://doi.org/10.1136/bmjopen-2016-012012)] [Medline: [28096249](https://pubmed.ncbi.nlm.nih.gov/28096249/)]
  22. Coppersmith G, Leary R, Crutchley P, Fine A. Natural language processing of social media as screening for suicide risk. *Biomed Inform Insights* 2018 Aug 27;10:1178222618792860 [FREE Full text] [doi: [10.1177/1178222618792860](https://doi.org/10.1177/1178222618792860)] [Medline: [30158822](https://pubmed.ncbi.nlm.nih.gov/30158822/)]
  23. Calvo R, Milne D, Hussain M, Christensen H. Natural language processing in mental health applications using non-clinical texts. *Nat Lang Eng* 2017 Jan 30;23(5):649-685. [doi: [10.1017/s1351324916000383](https://doi.org/10.1017/s1351324916000383)]
  24. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, et al. Clinical information extraction applications: a literature review. *J Biomed Inform* 2018 Jan;77:34-49 [FREE Full text] [doi: [10.1016/j.jbi.2017.11.011](https://doi.org/10.1016/j.jbi.2017.11.011)] [Medline: [29162496](https://pubmed.ncbi.nlm.nih.gov/29162496/)]
  25. Hirsch JS, Tanenbaum JS, Lipsky Gorman S, Liu C, Schmitz E, Hashorva D, et al. HARVEST, a longitudinal patient record summarizer. *J Am Med Inform Assoc* 2015 Mar;22(2):263-274 [FREE Full text] [doi: [10.1136/amiainj-2014-002945](https://doi.org/10.1136/amiainj-2014-002945)] [Medline: [25352564](https://pubmed.ncbi.nlm.nih.gov/25352564/)]
  26. Meng Y, Rumshisky A, Romanov A. Temporal Information Extraction for Question Answering Using Syntactic Dependencies in an LSTM-based Architecture. In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*.: Association for Computational Linguistics; 2017 Presented at: Conference on Empirical Methods in Natural Language Processing (EMNLP); 2017; Copenhagen, Denmark p. 887-896. [doi: [10.18653/v1/d17-1092](https://doi.org/10.18653/v1/d17-1092)]
  27. Tourille J, Ferret O, Tannier X, Névéol A. Temporal information extraction from clinical text. In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*.: Association for Computational Linguistics; 2017 Presented at: Conference of the European Chapter of the Association for Computational Linguistics (EACL); 04/2017; Valencia, Spain p. 739-745. [doi: [10.18653/v1/e17-2117](https://doi.org/10.18653/v1/e17-2117)]
  28. Viani N, Kam J, Yin L, Bittar A, Dutta R, Patel R, et al. Temporal information extraction from mental health records to identify duration of untreated psychosis. *J Biomed Semantics* 2020 Mar 10;11(1):2 [FREE Full text] [doi: [10.1186/s13326-020-00220-2](https://doi.org/10.1186/s13326-020-00220-2)] [Medline: [32156302](https://pubmed.ncbi.nlm.nih.gov/32156302/)]
  29. Denny J, Peterson JF, Choma NN, Xu H, Miller RA, Bastarache L, et al. Extracting timing and status descriptors for colonoscopy testing from electronic medical records. *J Am Med Inform Assoc* 2010;17(4):383-388 [FREE Full text] [doi: [10.1136/jamia.2010.004804](https://doi.org/10.1136/jamia.2010.004804)] [Medline: [20595304](https://pubmed.ncbi.nlm.nih.gov/20595304/)]
  30. Moharasan G, Ho TB. Extraction of temporal information from clinical narratives. *J Healthc Inform Res* 2019 Feb 27;3(2):220-244. [doi: [10.1007/s41666-019-00049-0](https://doi.org/10.1007/s41666-019-00049-0)]
  31. Doddington G, Mitchell A, Przybocki M, Ramshaw L, Strassel S, Weischedel R. The Automatic Content Extraction (ACE) Program - Tasks, Data, and Evaluation. In: *Proceedings of the Fourth International Conference on Language Resources*

- and Evaluation (LREC).: European Language Resources Association (ELRA); 2004 Presented at: Language Resources and Evaluation Conference; 2004; Lisbon, Portugal URL: <http://www.lrec-conf.org/proceedings/lrec2004/pdf/5.pdf>
32. Styler WF, Bethard S, Finan S, Palmer M, Pradhan S, de Groen PC, et al. Temporal annotation in the clinical domain. *Trans Assoc Comput Linguist* 2014 Apr;2:143-154 [FREE Full text] [Medline: 29082229]
  33. Pustejovsky J, Lee K, Bunt H, Romary L. ISO-TimeML: An International Standard for Semantic Annotation. In: *Proceedings of the Seventh International Conference on Language Resources and Evaluation.*: European Language Resources Association (ELRA); 2010 Presented at: Language Resources and Evaluation Conference; 05/2010; Valletta, Malta p. 394-397 URL: <https://aclanthology.org/L10-1027/>
  34. Sun W, Rumshisky A, Uzuner O. Evaluating temporal relations in clinical text: 2012 i2b2 Challenge. *J Am Med Inform Assoc* 2013 Sep 01;20(5):806-813 [FREE Full text] [doi: 10.1136/amiajnl-2013-001628] [Medline: 23564629]
  35. MacAvaney S, Desmet B, Cohan A, Soldaini L, Yates A, Zirikly A, et al. RSDD-Time: Temporal Annotation of Self-Reported Mental Health Diagnoses. In: *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic.* 2018 Presented at: Conference of the North American Chapter of the Association for Computational Linguistics; 06/2018; New Orleans, LA p. 168-173 URL: <https://aclanthology.org/W18-0618/> [doi: 10.18653/v1/w18-0618]
  36. Rajkumar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. *NPJ Digit Med* 2018 May 8;1(18):18 [FREE Full text] [doi: 10.1038/s41746-018-0029-1] [Medline: 31304302]
  37. Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.*: Association for Computational Linguistics; 2019 Presented at: Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; 06/2019; Minneapolis, Minnesota p. 4171-4186. [doi: 10.18653/v1/N19-1423]
  38. Peters ME, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, et al. Deep contextualized word representations. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.*: Association for Computational Linguistics; 2018 Presented at: Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; 06/2018; New Orleans, LA p. 2227-2237 URL: <https://aclanthology.org/N18-1202/> [doi: 10.18653/v1/N18-1202]
  39. Lin C, Miller T, Dligach D, Sadeque F, Bethard S, Savova G. A BERT-based one-pass multi-task model for clinical temporal relation extraction. In: *Proceedings of the 19th SIGBioMed Workshop on Biomedical Language Processing.*: Association for Computational Linguistics; 2020 Presented at: Annual Meeting of the Association for Computational Linguistics; 07/2020; Online p. 70-75. [doi: 10.18653/v1/2020.bionlp-1.7]
  40. Li Y, Rao S, Solares JRA, Hassaine A, Ramakrishnan R, Canoy D, et al. BEHRT: Transformer for Electronic Health Records. *Sci Rep* 2020 Apr 28;10(1):7155 [FREE Full text] [doi: 10.1038/s41598-020-62922-y] [Medline: 32346050]
  41. Shang J, Ma T, Xiao C, Sun J. Pre-training of graph augmented transformers for medication recommendation. In: *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence.* 2019 Presented at: International Joint Conference on Artificial Intelligence; 08/2019; Macao, China p. 5953-5959. [doi: 10.24963/ijcai.2019/825]
  42. Rasmy L, Xiang Y, Xie Z, Tao C, Zhi D. Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. *NPJ Digit Med* 2021 May 20;4(1):86 [FREE Full text] [doi: 10.1038/s41746-021-00455-y] [Medline: 34017034]
  43. Spores JM. *Clinician's Guide to Psychological Assessment and Testing: With Forms and Templates for Effective Practice.* New York City: Springer Publishing Company; Sep 18, 2012:448.
  44. Filannino M, Stubbs A, Uzuner Ö. Symptom severity prediction from neuropsychiatric clinical records: Overview of 2016 CEGS N-GRID shared tasks Track 2. *J Biomed Inform* 2017 Nov;75S:S62-S70 [FREE Full text] [doi: 10.1016/j.jbi.2017.04.017] [Medline: 28455151]
  45. Shing H, Nair S, Zirikly A, Friedenber M, Daumé III H, Resnik P. Expert, crowdsourced, and machine assessment of suicide risk via online postings. In: *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic.*: Association for Computational Linguistics; 2018 Jun Presented at: Conference of the North American Chapter of the Association for Computational Linguistics; 2018; New Orleans, LA p. 25-36 URL: <https://aclanthology.org/W18-0603/>
  46. Zirikly A, Resnik P, Uzuner Ö, Hollingshead K. CLPsych 2019 shared task: Predicting the degree of suicide risk in Reddit posts. In: *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology.*: Association for Computational Linguistics; 2019 Jun Presented at: Conference of the North American Chapter of the Association for Computational Linguistics; 2019; Minneapolis, MN p. 24-33 URL: <https://aclanthology.org/W19-3003/>
  47. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. *J Lang Soc Psychol* 2009 Dec 08;29(1):24-54. [doi: 10.1177/0261927x09351676]
  48. Mohammad SM, Turney PD. Crowdsourcing a word-emotion association lexicon. *Comput Intell* 2013;29(3):436-465. [doi: 10.1111/j.1467-8640.2012.00460.x]
  49. Koleck TA, Dreisbach C, Bourne PE, Bakken S. Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. *J Am Med Inform Assoc* 2019 Apr 01;26(4):364-379 [FREE Full text] [doi: 10.1093/jamia/ocy173] [Medline: 30726935]

50. Heintzelman NH, Taylor RJ, Simonsen L, Lustig R, Anderko D, Haythornthwaite JA, et al. Longitudinal analysis of pain in patients with metastatic prostate cancer using natural language processing of medical record text. *J Am Med Inform Assoc* 2013 Sep 01;20(5):898-905 [FREE Full text] [doi: [10.1136/amiajnl-2012-001076](https://doi.org/10.1136/amiajnl-2012-001076)] [Medline: [23144336](https://pubmed.ncbi.nlm.nih.gov/23144336/)]
51. Playford D. The International Classification of Functioning, Disability, and Health. In: Dietz V, Ward N, editors. *Oxford Textbook of Neurorehabilitation*. Oxford, UK: Oxford University Press; Feb 2015:3-7.
52. Iezzoni LI, Marsella SA, Lopinsky T, Heaphy D, Warsett KS. Do prominent quality measurement surveys capture the concerns of persons with disability? *Disabil Health J* 2017 Apr;10(2):222-230. [doi: [10.1016/j.dhjo.2017.01.007](https://doi.org/10.1016/j.dhjo.2017.01.007)] [Medline: [28185857](https://pubmed.ncbi.nlm.nih.gov/28185857/)]
53. World Health Organization. In: Wilkinson R, Marmot M, editors. *Social Determinants of Health: The Solid Facts 2nd ed*. Copenhagen: WHO Regional Office for Europe; 2003.
54. Conway M, Keyhani S, Christensen L, South BR, Vali M, Walter LC, et al. Moonstone: a novel natural language processing system for inferring social risk from clinical narratives. *J Biomed Semantics* 2019 Apr 11;10(1):6 [FREE Full text] [doi: [10.1186/s13326-019-0198-0](https://doi.org/10.1186/s13326-019-0198-0)] [Medline: [30975223](https://pubmed.ncbi.nlm.nih.gov/30975223/)]
55. Dorr D, Bejan CA, Pizzimenti C, Singh S, Storer M, Quinones A. Identifying patients with significant problems related to social determinants of health with natural language processing. *Stud Health Technol Inform* 2019 Aug 21;264:1456-1457. [doi: [10.3233/SHTI190482](https://doi.org/10.3233/SHTI190482)] [Medline: [31438179](https://pubmed.ncbi.nlm.nih.gov/31438179/)]
56. Deferio JJ, Breitinger S, Khullar D, Sheth A, Pathak J. Social determinants of health in mental health care and research: a case for greater inclusion. *J Am Med Inform Assoc* 2019 Aug 01;26(8-9):895-899 [FREE Full text] [doi: [10.1093/jamia/ocz049](https://doi.org/10.1093/jamia/ocz049)] [Medline: [31329877](https://pubmed.ncbi.nlm.nih.gov/31329877/)]
57. Feller DJ, Bear Don't Walk Iv OJ, Zucker J, Yin MT, Gordon P, Elhadad N. Detecting social and behavioral determinants of health with structured and free-text clinical data. *Appl Clin Inform* 2020 Jan 04;11(1):172-181 [FREE Full text] [doi: [10.1055/s-0040-1702214](https://doi.org/10.1055/s-0040-1702214)] [Medline: [32131117](https://pubmed.ncbi.nlm.nih.gov/32131117/)]
58. Closing the gap in a generation: health equity through action on the social determinants of health - Final report of the commission on social determinants of health. World Health Organization. 2008. URL: <https://www.who.int/publications/item/WHO-IER-CSDH-08.1> [accessed 2022-02-22]
59. Stolwijk C, Castillo-Ortiz JD, Gignac M, Luime J, Boonen A, OMERACT Worker Productivity Group. Importance of contextual factors when measuring work outcome in ankylosing spondylitis: a systematic review by the OMERACT Worker Productivity Group. *Arthritis Care Res (Hoboken)* 2015 Sep 26;67(9):1316-1327 [FREE Full text] [doi: [10.1002/acr.22573](https://doi.org/10.1002/acr.22573)] [Medline: [25732705](https://pubmed.ncbi.nlm.nih.gov/25732705/)]
60. Sinclair CM, Meredith P, Strong J, Feeney R. Personal and contextual factors affecting the functional ability of children and adolescents with chronic pain: a systematic review. *J Dev Behav Pediatr* 2016 May;37(4):327-342. [doi: [10.1097/DBP.0000000000000300](https://doi.org/10.1097/DBP.0000000000000300)] [Medline: [27096569](https://pubmed.ncbi.nlm.nih.gov/27096569/)]
61. Brownson RC, Hoehner CM, Day K, Forsyth A, Sallis JF. Measuring the built environment for physical activity: state of the science. *Am J Prev Med* 2009 Apr;36(4 Suppl):S99-123.e12 [FREE Full text] [doi: [10.1016/j.amepre.2009.01.005](https://doi.org/10.1016/j.amepre.2009.01.005)] [Medline: [19285216](https://pubmed.ncbi.nlm.nih.gov/19285216/)]
62. Stansfeld S, Candy B. Psychosocial work environment and mental health--a meta-analytic review. *Scand J Work Environ Health* 2006 Dec;32(6):443-462 [FREE Full text] [doi: [10.5271/sjweh.1050](https://doi.org/10.5271/sjweh.1050)] [Medline: [17173201](https://pubmed.ncbi.nlm.nih.gov/17173201/)]
63. Orstad SL, McDonough MH, Stapleton S, Altincekic C, Troped PJ. A systematic review of agreement between perceived and objective neighborhood environment measures and associations with physical activity outcomes. *Environ Behav* 2016 Sep 29;49(8):904-932. [doi: [10.1177/0013916516670982](https://doi.org/10.1177/0013916516670982)]
64. Madans J. Proposed purpose of an internationally comparable general disability measure. 2004 Presented at: Third Meeting Washington Group on Disability Statistics; February 19-20, 2004; Brussels, Belgium.
65. Altman BM. Appendix A: Population Survey Measures of Functioning: Strengths and Weaknesses. In: Wunderlich GS, editor. *Improving the Measurement of Late-Life Disability in Population Surveys: Beyond ADLs and IADLs: Summary of a Workshop National Academies of Sciences, Engineering, and Medicine*. 2009. *Improving the Measurement of Late-Life Disability in Population Surveys: Beyond ADLs and IADLs: Summary of a Workshop*. Washington, DC: The National Academies Press; 2009:99-156.
66. Heerkens YF, de Brouwer CP, Engels JA, van der Gulden JW, Kant I. Elaboration of the contextual factors of the ICF for Occupational Health Care. *Work* 2017;57(2):187-204. [doi: [10.3233/WOR-172546](https://doi.org/10.3233/WOR-172546)] [Medline: [28582939](https://pubmed.ncbi.nlm.nih.gov/28582939/)]
67. Haley SM, Coster WJ, Binda-Sundberg K. Measuring physical disablement: the contextual challenge. *Phys Ther* 1994 May;74(5):443-451. [doi: [10.1093/ptj/74.5.443](https://doi.org/10.1093/ptj/74.5.443)] [Medline: [8171106](https://pubmed.ncbi.nlm.nih.gov/8171106/)]
68. Jette DU, Halbert J, Iverson C, Miceli E, Shah P. Use of standardized outcome measures in physical therapist practice: perceptions and applications. *Phys Ther* 2009 Feb;89(2):125-135. [doi: [10.2522/ptj.20080234](https://doi.org/10.2522/ptj.20080234)] [Medline: [19074618](https://pubmed.ncbi.nlm.nih.gov/19074618/)]
69. Stephens C, Breheny M, Mansvelt J. Healthy ageing from the perspective of older people: a capability approach to resilience. *Psychol Health* 2015 Apr 29;30(6):715-731. [doi: [10.1080/08870446.2014.904862](https://doi.org/10.1080/08870446.2014.904862)] [Medline: [24678916](https://pubmed.ncbi.nlm.nih.gov/24678916/)]
70. Yeung P, Breheny M. Quality of life among older people with a disability: the role of purpose in life and capabilities. *Disabil Rehabil* 2021 Jan;43(2):181-191. [doi: [10.1080/09638288.2019.1620875](https://doi.org/10.1080/09638288.2019.1620875)] [Medline: [31335217](https://pubmed.ncbi.nlm.nih.gov/31335217/)]

71. Newman-Griffis D, Porcino J, Zirikly A, Thieu T, Camacho Maldonado J, Ho P, et al. Broadening horizons: the case for capturing function and the role of health informatics in its use. *BMC Public Health* 2019 Oct 15;19(1):1288 [FREE Full text] [doi: [10.1186/s12889-019-7630-3](https://doi.org/10.1186/s12889-019-7630-3)] [Medline: [31615472](https://pubmed.ncbi.nlm.nih.gov/31615472/)]
72. Agaronnik ND, Lindvall C, El-Jawahri A, He W, Iezzoni LI. Challenges of Developing a Natural Language Processing Method With Electronic Health Records to Identify Persons With Chronic Mobility Disability. *Arch Phys Med Rehabil* 2020 Oct;101(10):1739-1746 [FREE Full text] [doi: [10.1016/j.apmr.2020.04.024](https://doi.org/10.1016/j.apmr.2020.04.024)] [Medline: [32446905](https://pubmed.ncbi.nlm.nih.gov/32446905/)]
73. Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations, Board on the Health of Select Populations; Institute of Medicine. *Psychological Testing in the Service of Disability Determination*. Washington DC: National Academies Press; Jun 29, 2015.
74. Elson D, McKeown K. Automatic attribution of quoted speech in literary narrative. In: AAI. 2010 Presented at: The Twenty-Fourth AAI Conference on Artificial Intelligence; July 2010; Atlanta, Georgia.
75. Pareti S, O'Keefe T, Konstas I. Automatically Detecting and Attributing Indirect Quotations. 2013 Presented at: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing.; October 18-21, 2013; Seattle, Washington URL: <https://openreview.net>
76. O'Keefe T. A sequence labelling approach to quote attribution. 2012 Jul Presented at: EMNLP-CoNLL '12: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning; July 2012; Jeju Island, Korea.
77. Nguyen H, Patrick J. Text Mining in Clinical Domain: Dealing with Noise. 2016 Presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13, 2016; San Francisco, CA. [doi: [10.1145/2939672.2939720](https://doi.org/10.1145/2939672.2939720)]
78. Burrows J. 'Delta': a measure of stylistic difference and a guide to likely authorship. *Lit Linguistics Comput* 2002 Sep 01;17(3):267-287. [doi: [10.1093/litc/17.3.267](https://doi.org/10.1093/litc/17.3.267)]
79. Hoover DL. Testing Burrows's Delta. *Lit Linguistics Comput* 2004 Nov 01;19(4):453-475. [doi: [10.1093/litc/19.4.453](https://doi.org/10.1093/litc/19.4.453)]
80. Nirxhi S, Dharaskar R, Thakare V. Authorship verification of online messages for forensic investigation. *Procedia Comput Sci* 2016;78:640-645. [doi: [10.1016/j.procs.2016.02.111](https://doi.org/10.1016/j.procs.2016.02.111)]
81. Lambers M, Veenman CJ. Forensic Authorship Attribution Using Compression Distances to Prototypes. 2009 Presented at: International Workshop on Computational Forensics; August 13-14, 2009; The Hague, The Netherlands p. 13-24. [doi: [10.1007/978-3-642-03521-0\\_2](https://doi.org/10.1007/978-3-642-03521-0_2)]
82. Rocha A, Scheirer WJ, Forstall CW, Cavalcante T, Theophilo A, Shen B, et al. Authorship attribution for social media forensics. *IEEE Trans Inform Forensic Secur* 2017 Jan;12(1):5-33. [doi: [10.1109/tifs.2016.2603960](https://doi.org/10.1109/tifs.2016.2603960)]
83. Frye RH, Wilson DC. Defining Forensic Authorship Attribution for Limited Samples from Social Media. 2018 Presented at: FLAIRS Conference 2018; September 10-14, 2018; Melbourne Beach.
84. O'Keefe T. Examining the Impact of Coreference Resolution on Quote Attribution. 2013 Presented at: Proceedings of Australasian Language Technology Association Workshop; 2013; Brisbane, Australia p. 43-52.
85. Almeida M. A Joint Model for Quotation Attribution and Coreference Resolution. 2014 Presented at: The 14th Conference of the European Chapter of the Association of Computational Linguistics; April 2014; Gothenburg, Sweden p. 39-48. [doi: [10.3115/v1/e14-1005](https://doi.org/10.3115/v1/e14-1005)]
86. Zhang H, Boons F, Batista-Navarro R. Whose story is it anyway? Automatic extraction of accounts from news articles. *Inf Process Manag* 2019 Sep;56(5):1837-1848. [doi: [10.1016/j.ipm.2019.02.012](https://doi.org/10.1016/j.ipm.2019.02.012)]
87. Walker V. The Need for Annotated Corpora from Legal Documents, and for (Human) Protocols for Creating Them: The Attribution Problem. Maurice A. Deane School of Law at Hofstra University. Scholarly Commons at Hofstra Law. 2016. URL: [https://scholarlycommons.law.hofstra.edu/cgi/viewcontent.cgi?article=2231&context=faculty\\_scholarship](https://scholarlycommons.law.hofstra.edu/cgi/viewcontent.cgi?article=2231&context=faculty_scholarship) [accessed 2022-02-22]
88. Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. *Psychol Med* 2019 Jul;49(9):1426-1448. [doi: [10.1017/S0033291719000151](https://doi.org/10.1017/S0033291719000151)] [Medline: [30744717](https://pubmed.ncbi.nlm.nih.gov/30744717/)]
89. Kemp J, Zhang T, Inglis F, Wiljer D, Sockalingam S, Crawford A, et al. Delivery of compassionate mental health care in a digital technology-driven age: scoping review. *J Med Internet Res* 2020 Mar 06;22(3):e16263 [FREE Full text] [doi: [10.2196/16263](https://doi.org/10.2196/16263)] [Medline: [32141833](https://pubmed.ncbi.nlm.nih.gov/32141833/)]
90. Murnane EL, Walker TG, Tench B, Volda S, Snyder J. Personal informatics in interpersonal contexts. *Proc ACM Hum-Comput Interact* 2018 Nov;2(CSCW):1-27. [doi: [10.1145/3274396](https://doi.org/10.1145/3274396)]
91. Rigby M, Lindmark J, Furlan PM. The importance of developing an informatics framework for mental health. *Health Policy* 1998 Jul;45(1):57-67. [doi: [10.1016/s0168-8510\(98\)00028-1](https://doi.org/10.1016/s0168-8510(98)00028-1)] [Medline: [10183013](https://pubmed.ncbi.nlm.nih.gov/10183013/)]
92. Newman-Griffis D. Translational NLP: A New Paradigm and General Principles for Natural Language Processing Research. 2021 Jun Presented at: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics; June 2021; Online p. 4125-4138. [doi: [10.18653/v1/2021.naacl-main.325](https://doi.org/10.18653/v1/2021.naacl-main.325)]
93. Bishop T. Mental disorders and learning disabilities in children and adolescents: learning disabilities. *FP Essent* 2018 Dec;475:18-22. [Medline: [30556687](https://pubmed.ncbi.nlm.nih.gov/30556687/)]

## Abbreviations

**ACL:** Association for Computational Linguistics  
**DSM-IV-TR:** Diagnostic and Statistical Manual of Mental Disorders Text Revision Fourth Edition  
**EHR:** electronic health record  
**MER:** Medical Evidence of Record  
**NLP:** natural language processing  
**SMI:** severe mental illness  
**SSA:** Social Security Administration  
**SSDI:** Social Security Disability Insurance  
**SSI:** Supplemental Security Income

*Edited by C Lovis; submitted 19.07.21; peer-reviewed by T Ntalindwa, J Coquet, I Mircheva; comments to author 18.08.21; revised version received 08.10.21; accepted 16.01.22; published 18.03.22*

*Please cite as:*

Zirikly A, Desmet B, Newman-Griffis D, Marfeo EE, McDonough C, Goldman H, Chan L  
*Information Extraction Framework for Disability Determination Using a Mental Functioning Use-Case*  
*JMIR Med Inform* 2022;10(3):e32245  
URL: <https://medinform.jmir.org/2022/3/e32245>  
doi: [10.2196/32245](https://doi.org/10.2196/32245)  
PMID: [35302510](https://pubmed.ncbi.nlm.nih.gov/35302510/)

©Ayah Zirikly, Bart Desmet, Denis Newman-Griffis, Elizabeth E Marfeo, Christine McDonough, Howard Goldman, Leighton Chan. Originally published in JMIR Medical Informatics (<https://medinform.jmir.org>), 18.03.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org/>, as well as this copyright and license information must be included.