A literature review of NLP based on Activity Theory

Mitrasadat Mirshafiee Khoozani
University of Calgary
Calgary, AB, , Canada
mitrasadat.mirshafie@ucalgary.ca

ABSTRACT

To improve and facilitate the process of decision making, we need to gather the right type of information. Certainly, one of the most important resources at our disposal is natural language. In this report, we will focus on natural language processing - a sub-field of Artificial Intelligence (AI)- and its applications in forming decisions in various fields of study (e.g., healthcare and finance). However, while numerous articles describe NLP tools and practices, there is little research on their application in context. This is a significant gap given that these applications are a highly complex and employed in our everyday life. Therefore, in this research we apply activity theory (AT) to investigate the development of artificial intelligence in creating Natural language processing systems that can aid us in decision making and designing decision support systems.

CCS CONCEPTS

• Information systems → Decision support systems; **• Computing methodologies → Natural language processing**; **• Human-centered computing → HCI theory, concepts and models**.

KEYWORDS

activity theory, decision support systems, neural networks, natural language processing, decision support systems, healthcare

ACM Reference Format:
Mitrasadat Mirshafiee Khoozani. 2022. A literature review of NLP based on Activity Theory. In *Proceedings of University of Calgary (Calgary, AB, CA).* ACM, New York, NY, USA, 6 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Natural language processing (NLP) is a branch of linguistics, computer science, and artificial intelligence that studies the interactions between computers and human language, with a focus on how to program computers to process and analyze huge amounts of textual data. Even though there is a lot of natural language text in the linked world and it contains a lot of knowledge, it is getting harder for humans to disperse it and find the knowledge or wisdom in it, especially given time constraints. The automated NLP is designed to perform this task as accurately and efficiently with minimal human intervention, but in a much larger scale and much faster way.

This field has been studied using numerous techniques. These methods can be divided into two overall groups: structural linguistic ([9]) and statistical methods ([40], [40]). In this study we will take a look at the latest paper which use mathematical and statistical methods that lead to automated systems for users.

NLP contributes to numerous fields of research. For example, the goal of computerised clinical decision support (CDS) is to facilitate decision-making among healthcare professionals and the general public by making health-related data available when demanded. With the help of textual data we are able to to drive decision support systems, by representing clinical knowledge and generate narratives that can guide users in their everyday efforts. All of this is made possible by utilising machine learning and deep learning techniques related to natural language processing.

Although anecdotal evidence suggests that natural Language processing technologies are highly effective in our systems and rapidly growing in popularity. empirical and collective research on AI-powered decision support systems (DSS) is sparse.

Activity Theory is a philosophical and interdisciplinary framework for researching various human actions as developmental processes, with simultaneous interlinkages between individual and social levels ([30], [31], [19]). Activity Theory has its roots on the classical German philosophy. Recently, some scholars have committed their efforts to converting Activity Theory into models which can be used in various areas.

This field has previously inspired several theoretical studies on the adoption and implementation of technologies, including information systems ([2], [27], [17]), mobile technologies ([28], Kietzmann, 2008; Ryu et al., 2005), and intermodal mobility ecosystems ([2]). Activity theory has been software development ([15]), entrepreneurship ([25]), organisational learning ([20]), etc. Inspired by all the previous work done in this field, this study uses Activity Theory (AT) as a theoretical lens to examine NLP systems, and their contribution to automatic systems developed by previous researchers and practitioners.

2 RESEARCH QUESTIONS

Based on the above gaps observed and motivations mentioned, in this work, we classify the body of available knowledge and aim at answering the following research questions (RQs):

RQ1. What are the main components of NLP applications through the lens of activity theory?
RQ2. What studies present evidence quantifying the impact of NLP results and benefit in decision making? (here we present a collection of NLP applications that were used to form decision making systems - we can map the three stages of intelligence, design, and choice to each study we review)
RQ3. How these NLP applications/pipelines components of action

Calgary, AB, CA, July and August, 2022,

https://doi.org/XXXXXXX.XXXXXXX

theory fit into the categories of action theory? (Here we describe what is defined as objective-rules-community-etc.)
RQ4. How many studies consider multiple NLP algorithms, and what is the distribution of algorithms out of all the studies sampled?

Each and every one of the above mentioned research questions, help us move from one research stage to the next. First we take a look at studies available on artificial intelligence and its applications based on the components in activity theory, namely (object, subject, tool, rule, division of labor, and community)

The first question helps us define and understand the effect that language has, if any, on the process of decision making. The second question aims at summarizing all the methods that researchers use to collect textual data and then prepare it for mathematical, machine learning, and deep learning algorithms for better interpretation. After understanding the methods of data preparation, we jump into algorithms that can help us better understand text. These methods are normally divided into statistical and deep learning models with various levels of complexity and depth. Having looked at the methods, we dive into applications of natural language processing studied by scholars in different fields (i.e., NLP in healthcare, customer service, smart cities, etc.) that help build decision making systems.

3 RESEARCH STAGES

Inspired by [16] and [34] and with the goal of summarizing and categorizing research in the intersection of NLP and DSS, we set out to follow the following steps in this study:

1. Question formulation (keywords, scope, etc.)
2. Locating studies and evaluation (selection of specific topics)
3. Outcomes and interventions

3.1 Question Formulation

Before beginning any research, a research question (RQ) must be developed. It seeks to examine a current area of ambiguity and suggests a need for focused inquiry. Therefore, creating a strong RQ is important.

We must investigate several earlier articles and sources in order to formulate a clear research question.To look for the relevant literature, we used the following scientific databases: ScienceDirect, DBLP, Google Scholar,Scopus, ACM Digital Library, Springer, IEEE Xplore DigitalLibrary, and arXiv.

Through this research, we are able to gather a few important keywords that will assist us and guide us in the right direction, motivate us, and help us develop our idea further.

Among one of the first keywords used in this research paper are "Natural language processing", "Decision making", "Decision support systems", "activity theory", "activity systems", and "Applications of NLP in decision support systems".

3.2 Locating studies and evaluation

Preliminary studies were chosen based on two inclusion criteria: (i) they had to be published during the past 20 years, and (ii) they had to be published in journals or conferences. This resulted in a decrease in the number of papers from more than 453,000 to 37,500. Three exclusion criteria were applied to these chosen papers: I studies not given in English, (ii) studies that were copies of other research, and (iii) studies outside the field of AI. This further decreased the number of papers to 31,600. while further exclusion constraints were experimented with (e.g., inclusion of the phrase "artificial intelligence" AND "activity theory" AND "natural language processing"), no search results would appear after these filters as there has been no study previously done on the intersection of these three fields.

3.3 Outcomes and interventions

4 RQ1.

How NLP applications/pipelines can be broken down and analyzed through the lens of Activity Theory?

First, We provide a brief overview of the aspects of AT relevant to this work in this section. Activity theory is a conceptual framework created to form a unity of activity and consciousness. The foundational concept of the framework is “activity”, which is understood as purposeful, transformative, and developing interaction between actors (“subjects”) and the world (“objects”). [3] The theoretical foundations of AT in general can be found in the works of [47] and [?]

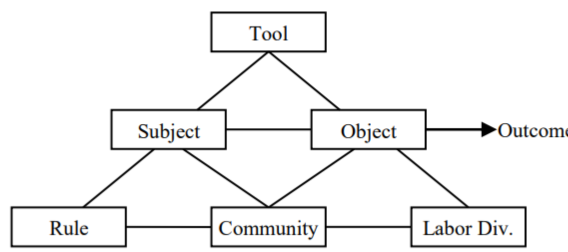

Figure 1: An activity system, from Engeström [18]

4.1 Object, Subject, and their Interaction

Interaction between "subjects" and "objects" is one of the analytical units suggested by activity theory. As a result, computers and text processing algorithms may appear to be objects if we consider that people and research teams studying NLP are subjects. This viewpoint is similar to the earlier studies on "human-computer" interaction in HCI. However, adopting an activity theoretical perspective had important implications for understanding how people use interactive technologies. The fact is that "computer and text processing algorithms" are often not an object of action but are rather a mediating artefact. Therefore, in general, individuals do not engage directly with computers; instead, they do so via computers. This idea is inspired by [26] which played a key role in introducing activity theory to HCI, reflected this perspective on interactive technologies in its very title: “Through the interface: An activity-theoretical perspective on human-computer interaction”. From this we can conclude that the object will be the information extracted from various forms of unstructured textual data.

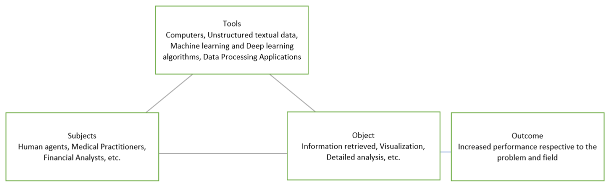

Figure 2: Figure 2 An activity system modified based on our research direction [18]

4.2 Tools, Communities, and their Interaction

As mentioned earlier, tools are considered exteriorized forms of mental processes manifested in constructs, whether physical or psychological. As a results the notion of tools in AT is broad and can involve digital devices, packages providing ML algorithms, and software.

On the other hand, communities consist of groups that each subject is placed into. Therefore, tools are mediating artifacts that serve each of the communities in a certain way, and in this relationship, rules act as a bridge between the two, regulating the boundaries and defining the responsibilities of each group. It is also important to note that there are social hierarchies and strata that help communities achieve goals (objects). These division of activities among actors/subjects are called divisions of labor.

To clarify the matter further, We exemplify some of the applications developed by researchers who utilized NLP to facilitate the action of agents and subjects.

One of the best examples of developed tools is Medline [23]. Medline was developed by the National Library of Medicine (group of subjects - communities in the Library of Medicine) in the United States of America[41]. It is a bibliographic database which contains information in biomedicine, biology, biochemistry, molecular evolution and life sciences. Medline contains journal citations and abstracts in the medical field which are collected from a set of different medical journals by the US National Institutes of Health. Hence, it is most widely used by another group of human agents (community) specified as the clinicians and research scholars (each with their own hierarchy of communities) in the field of medicine.

Metamap [6] is another tool that employs a knowledge intensive approach based on symbolic natural language processing and computational linguistic techniques . Metamap maps biomedical text to concepts in the UMLS Metathesaurus [43] - another tool for mapping biomedical text. Metamap has been extensively used for Information Retrieval and Information Extraction to give medical practitioners (subjects) effective access to hierarchical knowledge graph representations in medicine to facilitate the process of medical prescription and scientific exploration (objective).

Furthermore, in digital medicine, [50] developed a user-friendly platform called the desiderata (tool) for efficient collection of domain expert inputs, seamless integration with clinical data, and a highly scalable computing infrastructure. This platform is used by medical practitioners, subjects, making medical judgments.

As another instance, NLP systems such as cTAKES (clinical Text Analysis and Knowledge Extraction System) [42] and MedLEE (Medical Language Extraction and Encoding system) [21] have been designed to extract meaningful information from unstructured clinical text by identifying the domain entities/concepts in the text, annotating the domain entities with standard vocabularies, understanding the different linguistic and semantic relationship amongst the sentences. This is with a view to providing machine interpretable information, facilitating automatic analysis of healthcare information as well as providing users with meaningful information.

And last but not least, in [34], two NLP-empowered Information Retrieval (IR) models that utilize Part of speech tagging (POS-tagging), namely automatic POS-based term weighting schemes into bag-of-word (BoW) and Markov Random Field (MRF) IR models - all considered tools for subjects, were incorporated to improve the accuracy of patient (a target community) document creation. These efforts strengthen the conventional IR models and improve processing clinical documents or biomedical literature.

Based on all the above mentioned uses of NLP and its applications, we decide to classify the type of activities into the following categories: Monitor, Alarm/Alert, Predict, and Assign code.

Activity	Remind\Alert	Monitor	Predict	Assign Codes
Objectives	➢ Alert adverse events ➢ immunization	➢ drug-drug interaction ➢ syndromic presentation ➢ monitoring patient progress	➢ suicide intent ➢ adjustment to cancer ➢ cognitive/physical impairment	➢ assign a pre-defined code to reports for billing purposes
Tools	doctor reports, one-to-one conversation with the patient	Libraries such as Medline and applications developed based on them, clinical patient records	doctor reports, one-to-one conversation with the patient	Manually assigned and encoded datasets, billing reports
Subjects	Clinicians, Administrators,	Patients, Researchers, Clinicians	Researchers, Clinicians	human coders, Researchers, Administrators
Sample papers	1. Automated identification of adverse events related to central venous catheters 2. Using Electronic Medical Records to Enhance Detection and Reporting of Vaccine Adverse Events	1. Toward a complete dataset of drug–drug interaction information from publicly available sources 2. TIEVis: a Visual Analytics Dashboard for Temporal Information Extracted from Clinical Reports	1. Prediction of breast cancer distant recurrence using natural language processing and knowledge-guided convolutional neural network 2. Suicide Note Classification Using Natural Language Processing: A Content Analysis	1. Automatic construction of rule-based ICD-9-CM coding systems 2. "Can NLP techniques be utilized as a reliable tool for medical science?" - Building an NLP Framework to Classify Medical Reports

Figure 3: Categorization of activities in clinical NLP applications, inspired by [14]

5 RQ2.

What studies present evidence quantifying the impact of NLP results and benefit in decision making?

With the explosion of textual data, knowledge, and guidelines that specifies how to analyse text has far exceeded human cognitive capacity. It is very likely that decision support systems will make this information available and easily understandable to people (the agents in activities). Support systems are computer-based software programmes created to assist professionals, namely doctors, patients, financial analysts, etc. in making informed decisions, facilitating quick access to evidence-based advice, and identifying when additional information or different hypotheses should be taken into account. The goal of natural language processing, which enables computers to interpret meaning from natural language, is to significantly improve the functionality of these systems by automating and increasing the accuracy of our decisions.

Inspired by [38] To categories and filter out papers, we look for one or more of the following points in a paper.

- incorporate archival and unstructured notes ([35], [1], and [33])
- create models for tracking and interpreting community development ([**?**], [12])
- improve reasoning ([48])
- provide reliable and reproducible advice - addressed thoroughly by [36] and [49] ([29]), [22]
- prioritize evidence and test results
- engage professionals and the community to promote effective communication and coordination of the activity.[24]

To look at the impact of an application or study in detail, we need to explore our search results and papers further and specifically look at the evaluation stage in each article.

It is important to mention that all papers we looked at until now were in the fields of medicine and healthcare; therefore, we'll be focusing on healthcare in particular. But first there is an important issue we need to address. Reportedly, the absence of access to a standard corpus of data makes it difficult to develop metrics for assessing clinical decision systems for a specific clinical setting. Researchers must address legal, privacy, and institutional review issues in order to enable access to the enormous amount of data that is already available. That is why most of the evaluation methods were developed for specific users, document types and CDS goals. Many papers (for instance [38], [14], and [6]) reviewed in this study mentioned this fact that in order to evaluate CDS engines, an unified database containing anonymised medical data would be helpful. Some others such as [46] have extensively studied and addressed some of the challenges faced by researchers in evaluating medical systems and proposed some solutions for future work. Among the studies, there were some recent ones that dedicated their efforts to creating one or more gold standards such as [8].

For example, the earlier discussed MetaMap [6] information retrieval system features several evaluation processes that determines the accuracy of information acquired from a the main collection of documents [4]. However, this study mentions that this evaluation is specific to the information retrieval field and the accuracy of results, but a direct and thorough comparison between this system and other biomedical text retrieving system has never been performed. Never the less, the NLM's medical text indexer (MTI) [5], [7] was used in an indexing experiment called MetaMap-enhanced retrieval to compare MTI's indexing performance to official, manually compiled MEDLINE indexing. Performance of augmented documents always outperformed that of the baseline, and in all but one instance, MEDLINE indexing outperformed MTI indexing.

As another instance, we can once again mention Medline. The massive Medline database, created by the National Library of Medicine of the United States, has indexed numerous medical papers. To get the best possible retrieval outcomes, a variety of applications and search engines have been developed using this library. Here, we'll discuss a couple of them. For instance, [37] has created a survey of optimal search strategies for retrieving systematic reviews from Medline. To evaluate the efficiency of this system, the sensitivity, specificity, and precision of retrieval of systematic reviews of many unique terms were determined by comparison with a hand search of all articles. After the evaluation the author proclaims that systematic reviews can be retrieved from Medline with close to perfect sensitivity or specificity, or with high precision, by using empirical search strategies.

We can also name [45] which has constructed a state-of-the-art medical concept classification with an F-score >80, which is close to human agreement on the same task.

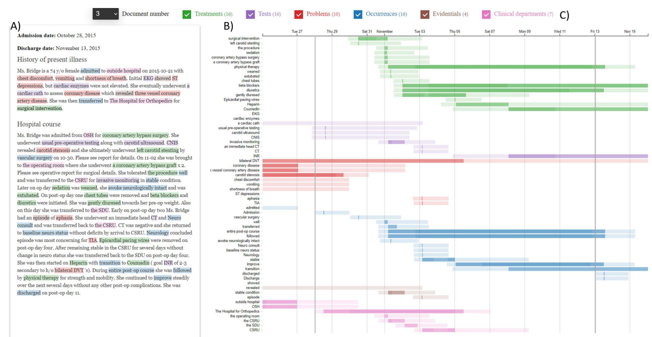

Figure 4: TIEVis [13] , A sample of assigning codes to text to facilitate decision making in identifying the disease, related departments, and potential prescription.

6 RQ3.

What type of evaluation techniques were mostly incorporated and to what level did human intervention play a role?

To enable for a study's scientific evaluation, all empirical research investigations must be examined. Studies are typically designed as clinical trials, cohort studies, or case-control studies with the goal of determining whether an intervention has a meaningful association with a desired outcome.

In recent articles when analysing their application based on a massive collection of medical records, researchers primarily attempt to use computational and automatic methods. With that in mind, we also have to consider that nearly all of these datasets were initially built, encoded, or categorized by humans.

The most fundamental and primary form of quantitative validation utilised in most papers is a 2x2 contingency table (or confusion matrix), where the quantity of correctly and wrongly assigned values for a particular binary outcome or classification label is compared with a gold (reference) standard. The performance measures precision (Positive Predictive Value), recall (sensitivity), accuracy, F-score can be computed using this table. [46] When the final product is more complex, other evaluation methods can be applied, such as continuous or ranked methodologies ([44], [10])

In the field of healthcare, clinical results can either be validated internally (using data from the original study sample) or externally (measured on a different sample). Since the aim of developing NLP methods is to create automatic solutions to particular problems, assessment methods can be intrinsic (evaluating an NLP system in terms of directly measuring its performance on achieving its immediate objective), such as in [4], and [23], or extrinsic (evaluating an NLP system in terms of its performance in accomplishing an overall goal in the system where the NLP algorithms are a part of a broader programme or pipeline), such as in [39], [44], and [11].

7 CONCLUSIONS AND THOUGHTS ON FUTURE WORK

The intersection of natural language processing, decision support systems, and activity theory is still unexplored and ripe for development. Healthcare and medicine are two areas where NLP can be useful because these are domains where doctors, patients, and researchers act as agents and work to improve communication in order to provide the optimal prescriptions and medical advice. As a result, disciplines like natural language processing are now increasingly useful in enhancing this collective effort.

Based on the papers we reviewed we can claim that there is a huge gap in the evaluation section (RQ. 2) in various papers. This is a consequence of not having an anonymized clinical database that can act as a gold standard in each task in the medical decision support systems.

8 ACKNOWLEDGMENTS

I would very much like to thank Dr. Ruhe for his patience, support, and guidance.

REFERENCES

[1] Naveed Afzal, Sunghwan Sohn, Sara Abram, Christopher G. Scott, Rajeev Chaudhry, Hongfang Liu, Iftikhar J. Kullo, and Adelaide M. Arruda-Olson. 2017. Mining peripheral arterial disease cases from narrative clinical notes using natural language processing. *Journal of Vascular Surgery* 65, 6 (2017), 1753–1761. https://doi.org/10.1016/j.jvs.2016.11.031
[2] David K Allen, Andrew Brown, Stan Karanasios, and Alistair Norman. 2013. How should technology-mediated organizational change be explained? A comparison of the contributions of critical realism and activity theory. *Mis Quarterly* (2013), 835–854.
[3] Obinna Anya, Hissam Tawfik, Atulya Nagar, and Saad Amin. 2010. Context-aware decision support in knowledge-intensive collaborative e-Work. *Procedia Computer Science* 1, 1 (2010), 2281–2290.
[4] Alan R Aronson. 2001. MetaMap evaluation. *Unpublished manuscript* (2001).
[5] Alan R Aronson, Olivier Bodenreider, H Florence Chang, Thomas C Rindflesch, and W John Wilbur. 2000. The NLM Indexing Initiative. In *In Proceedings of the American Medical Informatics Association Annual Symposium (AMIA 2000)(pp. 17–21). Philadelphia: Hanley & Belfus*. Citeseer.
[6] Alan R Aronson and François-Michel Lang. 2010. An overview of MetaMap: historical perspective and recent advances. *Journal of the American Medical Informatics Association* 17, 3 (2010), 229–236.
[7] Alan R Aronson, James G Mork, Clifford W Gay, Susanne M Humphrey, and Willie J Rogers. 2004. The NLM indexing initiative's medical text indexer. In *MEDINFO 2004*. IOS Press, 268–272.
[8] Deniz Caliskan, Jakob Zierk, Detlef Kraska, Stefan Schulz, Philipp Daumke, Hans-Ulrich Prokosch, and Lorenz A Kapsner. 2021. First Steps to Evaluate an NLP Tool's Medication Extraction Accuracy from Discharge Letters. In *German Medical Data Sciences: Bringing Data to Life: Proceedings of the Joint Annual Meeting of the German Association of Medical Informatics, Biometry and Epidemiology (gmds EV) and the Central European Network-International Biometric Society (CEN-IBS) 2020 in Berlin, Germany*, Vol. 278. IOS Press, 224.
[9] Noam Chomsky. 2014. *Aspects of the Theory of Syntax*. Vol. 11. MIT press.
[10] Evangelia Christodoulou, Jie Ma, Gary S Collins, Ewout W Steyerberg, Jan Y Verbakel, and Ben Van Calster. 2019. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of clinical epidemiology* 110 (2019), 12–22.
[11] Paul R Cohen and Adele E Howe. 1989. Toward AI research methodology: Three case studies in evaluation. *IEEE Transactions on Systems, Man, and Cybernetics* 19, 3 (1989), 634–646.
[12] Hong-Jie Dai, Chu-Hsien Su, You-Qian Lee, You-Chen Zhang, Chen-Kai Wang, Chian-Jue Kuo, and Chi-Shin Wu. 2021. Deep learning-based natural language processing for screening psychiatric patients. *Frontiers in psychiatry* 11 (2021), 533949.
[13] Robin De De Croon, Artuur Leeuwenberg, Jan Aerts, Marie-Francine Moens, Vero Vanden Vanden Abeele, and Katrien Verbert. 2021. TIEVis: A Visual Analytics Dashboard for Temporal Information Extracted from Clinical Reports. In *26th International Conference on Intelligent User Interfaces - Companion* (College Station, TX, USA) *(IUI '21 Companion)*. Association for Computing Machinery, New York, NY, USA, 34–36. https://doi.org/10.1145/3397482.3450731
[14] Dina Demner-Fushman, Wendy W. Chapman, and Clement J. McDonald. 2009. What can natural language processing do for clinical decision support? *Journal of Biomedical Informatics* 42, 5 (2009), 760–772. https://doi.org/10.1016/j.jbi.2009.08.007 Biomedical Natural Language Processing.
[15] Denis Dennehy, Kieran Conboy, Jennifer Ferreira, and Jaganath Babu. 2020. Sustaining open source communities by understanding the influence of discursive manifestations on sentiment. *Information Systems Frontiers* (2020), 1–17.
[16] David Denyer and David Tranfield. 2009. Producing a systematic review. (2009).
[17] John Effah and Ibrahim Osman Adam. 2021. Examining Client-Vendor Relationship in the Outsourcing of a Work Environment Virtualisation: An Activity Theory Perspective. *Information Systems Frontiers* (2021), 1–15.
[18] Yrjö Engeström. 1996. Developmental work research as educational research: Looking ten years back and into the zone of proximal development. *Nordisk pedagogik* 16, 3 (1996), 131–143.
[19] Y Engestrom, R ENGESTROM, and T VAHAAHO. 1999. When the Center Does Not Hold: The Importance of Knotworking. IN CHAIKLIN, S., HEDEGAARD, M. & JUUL, UJ (Eds.) Activity Theory and Social Practice: Cultural-Historical Approaches.
[20] Yrjö Engeström and Hannele Kerosuo. 2007. From workplace learning to inter-organizational learning and back: the contribution of activity theory. *Journal of workplace learning* 19, 6 (2007), 336–342.
[21] Carol Friedman, Lyudmila Shagina, Socrates A Socratous, and Xiao Zeng. 1996. A WEB-based version of MedLEE: A medical language extraction and encoding system. In *Proceedings of the AMIA Annual Fall Symposium*. American Medical Informatics Association, 938.
[22] Natalia Grabar, Clément Dalloux, and Vincent Claveau. 2020. CAS: corpus of clinical cases in French. *Journal of Biomedical Semantics* 11, 1 (2020), 1–10.
[23] Trisha Greenhalgh. 1997. How to read a paper: the Medline database. *Bmj* 315, 7101 (1997), 180–183.
[24] Qiwei He, Bernard P Veldkamp, Cees AW Glas, and Theo de Vries. 2017. Automated assessment of patients' self-narratives for posttraumatic stress disorder screening using natural language processing and text mining. *Assessment* 24, 2 (2017), 157–172.
[25] Oswald Jones and Robin Holt. 2008. The creation and evolution of new business ventures: an activity theory perspective. *Journal of small business and enterprise development* (2008).
[26] Victor Kaptelinin and Bonnie Nardi. 2012. Introduction: Activity theory and the changing face of HCI. In *Activity Theory in HCI*. Springer, 1–13.
[27] Stan Karanasios. 2018. Toward a unified view of technology and activity: The contribution of activity theory to information systems research. *Information Technology & People* 31, 1 (2018), 134–155.
[28] Stan Karanasios and David Allen. 2014. Mobile technology in mobile work: contradictions and congruencies in activity systems. *European Journal of Information Systems* 23, 5 (2014), 529–542.
[29] Veysel Kocaman and David Talby. 2021. Spark NLP: natural language understanding at scale. *Software Impacts* 8 (2021), 100058.
[30] Kari Kuutti. 1996. Debates in IS and CSCW research: Anticipating system design for post-Fordist work. In *Information technology and changes in organizational work*. Springer, 177–196.
[31] Kari Kuutti et al. 1996. Activity theory as a potential framework for human-computer interaction research. *Context and consciousness: Activity theory and human-computer interaction* 1744 (1996).
[32]]leont1978activity Alekseĭ Leont'ev. [n. d.]. Activity, consciousness, and personality. ([n. d.]).
[33] Maxwell Levis, Christine Leonard Westgate, Jiang Gui, Bradley V Watts, and Brian Shiner. 2021. Natural language processing of clinical mental health notes may add predictive value to existing suicide risk models. *Psychological medicine* 51, 8 (2021), 1382–1391.
[34] Robert M. Losee. 2001. Natural language processing in support of decision-making: phrases and part-of-speech tagging. *Information Processing Management* 37, 6 (2001), 769–787. https://doi.org/10.1016/S0306-4573(00)00061-3
[35] Ben J Marafino, Miran Park, Jason M Davies, Robert Thombley, Harold S Luft, David C Sing, Dhruv S Kazi, Colette DeJong, W John Boscardin, Mitzi L Dean, et al. 2018. Validation of prediction models for critical care outcomes using natural language processing of electronic health record data. *JAMA network open* 1, 8 (2018), e185097–e185097.
[36] Matthew BA McDermott, Shirly Wang, Nikki Marinsek, Rajesh Ranganath, Luca Foschini, and Marzyeh Ghassemi. 2021. Reproducibility in machine learning for

health research: Still a ways to go. *Science Translational Medicine* 13, 586 (2021), eabb1655.

[37] Victor M Montori, Nancy L Wilczynski, Douglas Morgan, and R Brian Haynes. 2005. Optimal search strategies for retrieving systematic reviews from Medline: analytical survey. *BMJ* 330, 7482 (2005), 68. https://doi.org/10.1136/bmj.38336.804167.47 arXiv:https://www.bmj.com/content/330/7482/68.full.pdf

[38] Vinay M Pai, Mary Rodgers, Richard Conroy, James Luo, Ruixia Zhou, and Belinda Seto. 2013. Workshop on using natural language processing applications for enhancing clinical decision making: an executive summary. *Journal of the American Medical Informatics Association* 21, e1 (08 2013), e2–e5. https://doi.org/10.1136/amiajnl-2013-001896 arXiv:https://academic.oup.com/jamia/article-pdf/21/e1/e2/5950060/21-e1-e2.pdf

[39] Patrick Paroubek, Stéphane Chaudiron, and Lynette Hirschman. 2007. Principles of evaluation in natural language processing. *Revue TAL* 48, 1 (2007), 7–31.

[40] Barbara BH Partee, Alice G ter Meulen, and Robert Wall. 2012. *Mathematical methods in linguistics*. Vol. 30. Springer Science & Business Media.

[41] Rune Sætre. 2006. GeneTUC: Natural Language Understanding in Medical Text. (2006).

[42] Guergana K Savova, James J Masanz, Philip V Ogren, Jiaping Zheng, Sunghwan Sohn, Karin C Kipper-Schuler, and Christopher G Chute. 2010. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *Journal of the American Medical Informatics Association* 17, 5 (2010), 507–513.

[43] Peri L Schuyler, William T Hole, Mark S Tuttle, and David D Sherertz. 1993. The UMLS Metathesaurus: representing different views of biomedical concepts. *Bulletin of the Medical Library Association* 81, 2 (1993), 217.

[44] Hanna Suominen. 2009. Performance evaluation measures for text mining. In *Handbook of research on text and web mining technologies*. IGI Global, 724–747.

[45] Özlem Uzuner, Brett R South, Shuying Shen, and Scott L DuVall. 2011. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. *Journal of the American Medical Informatics Association* 18, 5 (2011), 552–556.

[46] Sumithra Velupillai, Hanna Suominen, Maria Liakata, Angus Roberts, Anoop D. Shah, Katherine Morley, David Osborn, Joseph Hayes, Robert Stewart, Johnny Downs, Wendy Chapman, and Rina Dutta. 2018. Using clinical Natural Language Processing for health outcomes research: Overview and actionable suggestions for future advances. *Journal of Biomedical Informatics* 88 (2018), 11–19. https://doi.org/10.1016/j.jbi.2018.10.005

[47] Lev S Vygotsky. 1978. Mind in society Harvard university press. *Cambridge, MA* (1978).

[48] Hao Wang, Xianhai Meng, and Xingyu Zhu. 2022. Improving knowledge capture and retrieval in the BIM environment: Combining case-based reasoning and natural language processing. *Automation in Construction* 139 (2022), 104317.

[49] Shirley V Wang, Olga V Patterson, Joshua J Gagne, Jeffrey S Brown, Robert Ball, Pall Jonsson, Adam Wright, Li Zhou, Wim Goettsch, and Andrew Bate. 2019. Transparent reporting on research using unstructured electronic health record data to generate 'real world'evidence of comparative effectiveness and safety. *Drug safety* 42, 11 (2019), 1297–1309.

[50] Andrew Wen, Sunyang Fu, Sungrim Moon, Mohamed El Wazir, Andrew Rosenbaum, Vinod C Kaggal, Sijia Liu, Sunghwan Sohn, Hongfang Liu, and Jungwei Fan. 2019. Desiderata for delivering NLP to accelerate healthcare AI advancement and a Mayo Clinic NLP-as-a-service implementation. *NPJ digital medicine* 2, 1 (2019), 1–7.