applied sciences

MDPI

Review

Applications of Natural Language Processing Tools in Orthopaedic Surgery: A Scoping Review

Francesca Sasanelli [1,*,†], Khang Duy Ricky Le [2,3,4,5,†], Samuel Boon Ping Tay [6], Phong Tran [1] and Johan W. Verjans [7,8]

1 Department of Orthopaedic Surgery, Western Health, Melbourne, VIC 3011, Australia
2 Department of General Surgical Specialties, The Royal Melbourne Hospital, Melbourne, VIC 3052, Australia
3 Department of Surgical Oncology, Peter MacCallum Cancer Centre, Melbourne, VIC 3052, Australia
4 Department of Medical Education, Melbourne Medical School, The University of Melbourne, Melbourne, VIC 3010, Australia
5 Geelong Clinical School, Deakin University, Geelong, VIC 3220, Australia
6 Eastern Health, Box Hill, VIC 3128, Australia
7 Australian Institute for Machine Learning (AIML), University of Adelaide, Adelaide, SA 5000, Australia
8 Lifelong Health Theme (Platform AI), South Australian Health and Medical Research Institute, Adelaide, SA 5000, Australia
* Correspondence: francesca.sasanelli@wh.org.au
† These authors contributed equally to this work.

Abstract: The advent of many popular commercial forms of natural language processing tools has changed the way we can utilise digital technologies to tackle problems with big data. The objective of this review is to evaluate the current research and landscape of natural language processing tools and explore their potential use and impact in the field of orthopaedic surgery. In doing so, this review aims to answer the research question of how NLP tools can be utilised to streamline processes within orthopedic surgery. To do this, a scoping review was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) and Arksey and O'Malley framework for scoping reviews, as well as a computer-assisted literature search on the Medline, Embase and Google Scholar databases. Papers that evaluated the use of natural language processing tools in the field of orthopaedic surgery were included. Our literature search identified 24 studies that were eligible for inclusion. Our scoping review captured articles that highlighted multiple uses of NLP tools in orthopaedics. In particular, one study reported on the use of NLP for intraoperative monitoring, six for detection of adverse events, five for establishing orthopaedic diagnoses, two for assessing the patient experience, two as an informative resource for patients, one for predicting readmission, one for triaging, five for auditing and one for billing and coding. All studies assessed these various uses of NLP through its tremendous computational ability in extracting structured and unstructured text from the medical record, including operative notes, pathology and imaging reports, and progress notes, for use in orthopaedic surgery. Our review demonstrates that natural language processing tools are becoming increasingly studied for use and integration within various processes of orthopaedic surgery. These AI tools offer tremendous promise in improving efficiency, auditing and streamlining tasks through their immense computational ability and versatility. Despite this, further research to optimise and adapt these tools within the clinical environment, as well as the development of evidence-based policies, guidelines and frameworks are required before their wider integration within orthopaedics can be considered.

Keywords: natural language processing; artificial intelligence; generative artificial intelligence; machine learning; deep learning; ChatGPT; GPT-3; GPT-4; chatbot; generative pre-training transformer; orthopaedic surgery; orthopaedics

Citation: Sasanelli, F.; Le, K.D.R.; Tay, S.B.P.; Tran, P.; Verjans, J.W. Applications of Natural Language Processing Tools in Orthopaedic Surgery: A Scoping Review. *Appl. Sci.* **2023**, *13*, 11586. https://doi.org/10.3390/app132011586

Academic Editor: Carlos A. Iglesias

Received: 20 September 2023
Revised: 16 October 2023
Accepted: 18 October 2023
Published: 23 October 2023

1. Introduction

Natural language processing (NLP) refers to a subfield of artificial intelligence (AI) technology that includes generative AI tools that are designed to utilise large language models to receive, rationalise and subsequently generate human language [1,2]. These technologies have gained significant popularity, attributed to widely accessible and advanced iterations including ChatGPT, a generative pre-training transformer (GPT) chatbot tool released by OpenAI (San Francisco, CA, USA), Bing Chat by Microsoft (Redmond, WA, USA) and BARD by Google (San Francisco, CA, USA). These tools have led to a paradigm shift in the ways we can approach digital problems. Specifically, the underlying algorithms of these popular tools have been extensively trained to receive, recognise and interpret input from human language. In doing so, these tools have the ability to generate output data for various applications including content creation, education, restructuring or re-organisation of data as well as provide contextually relevant answers to prompts in real time. Popular ways this has been utilised have been to use Chatbot NLP tools such as ChatGPT or Google Bard to answer frequently asked questions, provide frameworks for academic assignments, and even author academic research papers. The advanced computational ability of these tools to process big data and generate natural human language has therefore understandably seen global implementation of these technologies across multiple industries including education, finance and business.

Progressive digitalisation of healthcare in the field of orthopaedic surgery, which utilises large amounts of data-based infrastructure through electronic medical records, telehealth, intra-operative stereotaxis and the use of radiology therefore offers a valuable opportunity for implementing these tools to improve the delivery of healthcare for orthopaedic surgeons. To date, the use of natural language processing tools in orthopaedic surgery is limited, with minimal research into clinical application of these tools and therefore no guidelines as to how orthopaedic surgery can effectively implement these technologies. Despite this, the translatability of generative AI and NLP tools is significantly high given the vast amount of unstructured free text in progress notes, operative notes, radiology reports and pathology reports. There is a growing field of research into the ways in which NLP tools can be applied in orthopaedic surgery to improve the experience for all stakeholders including surgeons, researchers, nursing and allied health staff as well as patients. In particular, there is a potential for these tools to enhance auditing processes, research capabilities, prognostication and triaging as well as optimisation of healthcare delivery for surgeons and healthcare access for patients. To date, however, these outcomes are poorly characterised, with no current studies that have explored the potential of NLP tools within the global journey of orthopaedic surgery. Furthermore, the challenges of implementing these technologies in practice, including issues related to privacy, clinician acceptance, resources and funding, as well as appropriate stewardship of data, remain poorly considered.

The main objective of this scoping review, which has been conducted in accordance with the Arksey and O'Malley framework, is to provide an up-to-date review of the current landscape of research exploring the potential avenues of implementation for NLP in orthopaedic surgery. This review aims to answer the research question of how NLP tools can be utilised to streamline processes within orthopedic surgery by characterising the efficacious evidence-based methods in which NLP tools can be applied through the journey of orthopaedic surgery. This review also aims to further contribute key insights and challenges of integrating contemporary NLP technologies into orthopaedic practice.

2. Materials and Methods

2.1. Literature Search Strategy

A scoping review was systematically conducted with adherence to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines and the Arksey and O'Malley framework for scoping reviews. The PRISMA guidelines were utilised to provide a structured and comprehensive foundational search to ensure all

relevant research in this field was captured in a systematic way. The Arksey and O'Malley framework was utilised to ensure a systematic and robust methodological approach was followed in the generation of this review.

A computer-assisted search of electronic databases Medline, Embase and Google Scholar was performed on 18 April 2023. The search query combined medical subject headings (MeSH) terms and keywords related to NLP, ChatGPT and surgery for all papers between 18 April 1974 to 18 April 2023. Additional articles were captured using hand searching of reference lists of included articles. The full search strategy can be viewed in Appendix B.

2.2. Inclusion and Exclusion Criteria

Full-text, peer-reviewed publications in the English language were assessed. Papers that were included evaluated the applications of NLP tools, large language models and other forms of generative AI within the field of orthopaedic surgery.

Papers were excluded if they (1) were not available in full-text or English language, (2) assessed AI tools that were not related to NLP, large language models or generative AI, (3) were pre-clinical, animal-based, cell-based or lab-based research, or (4) evaluated NLP tools in fields other than orthopaedic surgery.

2.3. Literature Screening and Data Extraction

Initial title and abstract screening were completed independently by two investigators (FS, KL). Studies that met eligibility based on the aforementioned inclusion and exclusion criteria were eligible and selected for full-text analysis. The same investigators (FS, KL) subsequently reviewed these articles for inclusion in this review. Disagreement during this process was resolved by consensus.

2.4. Quality Assessment

The quality of evidence of included studies was assessed utilising the Newcastle–Ottawa Scale (NOS) by two independent investigators (FS, KL). Disagreements during this process were resolved by consensus. If consensus could not be achieved, a third investigator (ST) was consulted for resolution.

3. Results

3.1. Overview of Included Studies

A total of 810 publications were retrieved following a computer-assisted search (Figure 1). Following the removal of duplicates, 519 articles were screened by title and abstract to assess eligibility, resulting in the exclusion of 305 articles. The remaining 214 articles progressed to full-text analysis in which 190 articles were excluded: 156 due to wrong population of interest, 12 due to wrong intervention, 12 due to wrong study design, 7 due to wrong outcomes and 1 article due to retraction. A total of 24 articles were included in this scoping review (Table 1). Of the 24 articles, 23 were retrospective observational studies utilising NLP tools within various fields of orthopaedics. The remaining article was a case study. All retrospective observational studies were of level III evidence and the case study was of level IV evidence based on the Oxford Centre for Evidence-Based Medicine guidelines.

Of the included studies, one study reported on the use of NLP for intraoperative monitoring, six for detection of adverse events, five for establishing orthopaedic diagnoses, two for assessing the patient experience, two as an informative resource for patients, one for predicting readmission, one for triaging, five for auditing and one for billing and coding (Figure 2).

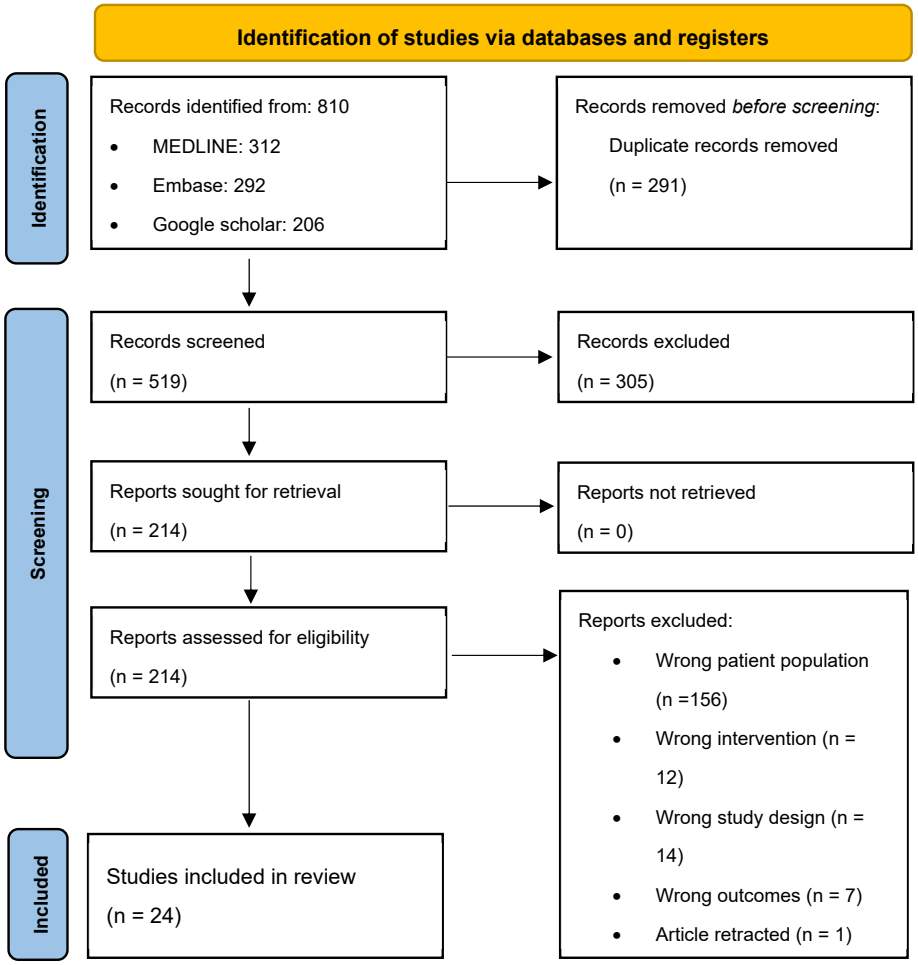

Figure 1. Search strategy and workflow in accordance with PRISMA guidelines.

Table 1. Overview of included studies.

Author	Year	Study Type	Location	Intervention	Cohort	Level of Evidence (CEBM)
Agaronnik et al. [3]	2022	Retrospective comparative study	United Kingdom	NLP (Bio_ClinicalBERT) for clinical data extraction, identifying intraoperative neuromonitoring in spine surgery, compared against traditional codes.	13,718 patients who had spinal surgeries, with 23,243 operative reports in total	III
Borjali et al. [4]	2021	Retrospective comparative study	United States	NLP (Generalised Linear Model, K-NN, Random Forest, SVM, Shallow Neural Network, Multilayer Bidirectional Long Short-term Memory (BiLSTM) and Convolutional Neural Network (CNN)) for clinical data extraction, identifying adverse events in free-text clinical notes.	6617 patients presenting for primary THR, with 7156 surgeries total	III
Fu et al. [5]	2021	Retrospective Observational Study	United States	NLP (MedTaggerIE) for clinical data extraction, identifying periprosthetic joint infections compared against manual chart review.	1179 surgeries.	III

Table 1. *Cont.*

Author	Year	Study Type	Location	Intervention	Cohort	Level of Evidence (CEBM)
Thirukumaran et al. [6]	2019	Retrospective observational study	United States	NLP (Python Software, v3.12.0) for data extraction: Surgical site infections in orthopaedic surgeries.	372 patients undergoing orthopaedic surgery.	III
Karhade et al. [7]	2020	Retrospective observational study	United States	NLP (Extreme Gradient Boosting (XGBoost)) for identifying clinically relevant outcome: Reoperation for wound infection, compared against manual review and ICD coding.	5860 patients who underwent spinal surgery.	III
Karhade et al. [8]	2021	Retrospective observational study	United States	NLP (Extreme Gradient Boosting (XGBoost)) for identifying clinically relevant outcome: Intraoperative vascular injury.	1035 patients who underwent spinal surgery.	III
Karhade et al. [9]	2022	Retrospective observational study	Netherlands	NLP (Extreme Gradient Boosting (XGBoost)) for clinical data extraction: Incidental Durotomy, compared against manual review.	3223 patients who underwent spinal surgery.	III
Li et al. [10]	2022	Retrospective observational study	United States	NLP (Python Software) for data extraction: Meniscal tear detection.	3593 Magnetic-Resonance Imaging reports of Knees.	III
Olthof et al. [11]	2021	Retrospective comparative / observational study	Ireland	NLP (Rule Based Classification, Naive Bayes, ANN, Random Forest and Bidirectional Encoder Representations from Transformers (BERT)) for data extraction: Identification of injuries in radiology reports.	2469 radiology reports of injured extremities, and 799 chest radiographs.	III
Groot et al. [12]	2020	Retrospective observational study	United Kingdom	NLP (Extreme Gradient Boosting (XGBoost)) for clinical data extraction, identifying bone metastases from bone scintigraphic reports.	704 reports from 704 patients who had bone scintigraphy performed.	III
Tibbo et al. [13]	2019	Retrospective observational study	United States	NLP (MedTaggerIE) for data extraction: Identifying periprosthetic femur fractures and further classification of this.	2982 total hip arthroplasty reports	III
Tan et al. [14]	2018	Retrospective comparative study	United States	NLP (Java Apache Lucene, Porter Stemmer (Python Software), NegEx and the Caret Package was used to implement the machine based model) as a clinical predictive model: Identifying radiological features suggestive of low back pain	871 lumbar spinal radiological reports	III
Bovonratwet et al. [15]	2021	Retrospective observational study	United States	NLP (Press Ganey Associates) analyses of patient comments to assess for sentiment compared to traditional measures of satisfaction.	319 patients who underwent primary total knee arthroplasty, 1048 patient comments in total	III

Table 1. *Cont.*

Author	Year	Study Type	Location	Intervention	Cohort	Level of Evidence (CEBM)
Langerhuizen et al. [16]	2021	Retrospective observational study	Netherlands	NLP (Latent Dirichlet Allocation (Python Software)) for data extraction and analyses: Common themes from patient reviews, compared against manual reviewing.	11,614 free-text reviews relating to orthopaedic surgeons.	III
Dubin et al. [17]	2023	Observational study	United States	NLP (ChatGPT) analyses of 20 Frequently Asked Questions.	20 FAQs	III
Seth et al. [18]	2023	Case study	Australia	NLP (ChatGPT) analyses of clinical questions relating to surgical management of knee osteoarthritis.	N/A	IV
Karhade et al. [19]	2022	Retrospective observational study	United States	NLP (Extreme Gradient Boosting (XGBoost)) for identifying clinically relevant outcome: 90-day inpatient readmission rates post-surgery.	708 patients who underwent spinal surgery.	III
Krebs et al. [20]	2023	Retrospective observational (pilot) study	Germany	NLP (in-house model) as a clinical predictive model compared against simple clinical variables.	398 patients presenting to spinal surgery clinic.	III
Sagheb et al. [21]	2021	Retrospective observational study	United States	NLP (MedTaggerIE) for data extraction in knee arthroplasty operative reports.	1592 knee arthroplasty operative reports	III
Wyles et al. [22]	2019	Retrospective observational study	United States	NLP (MedTaggerIE) for data extraction: Relevant data from total hip arthroplasty, compared against manual review.	250 total hip arthroplasty operative notes	III
Wyles et al. [23]	2022	Retrospective observational study	United States	NLP (MedTaggerIE) for data extraction: Relevant data from total-hip arthroplasty notes, compared against manual review.	39 total hip arthroplasty operative notes.	III
Shah et al. [24]	2020	Retrospective observational study	United Kingdom	NLP (CloudMedX) for data extraction: Relevant elements in knee arthroplasty.	1000 clinical notes	III
Jungmann et al. [25]	2022	Retrospective observational study	Germany	NLP (in-house model) for data extraction: Identifying incidence of fractures by analyses of radiographic reports.	5397 limb radiograph reports	III
Zaidat et al. [26]	2023	Retrospective observational study	United Kingdom	NLP (Natural Language Toolkit (Python Software) for data extraction: Generation of billing codes, compared against manual review.	922 spinal surgical operative notes	III

Of the 24 studies, all were published within the last 5 years, specifically from the year 2018. The year with the most publications was 2021 (n = 7) and the year with the least publications was 2018 (n = 1) (Figure 3). Fifty-eight percent were published in the United States (n = 14), 16.6% were published in the United Kingdom (n = 4), 8.3% were published in Germany and the Netherlands (n = 2) and 4.16% were published in Ireland and Australia (n = 1) (Figure 4).

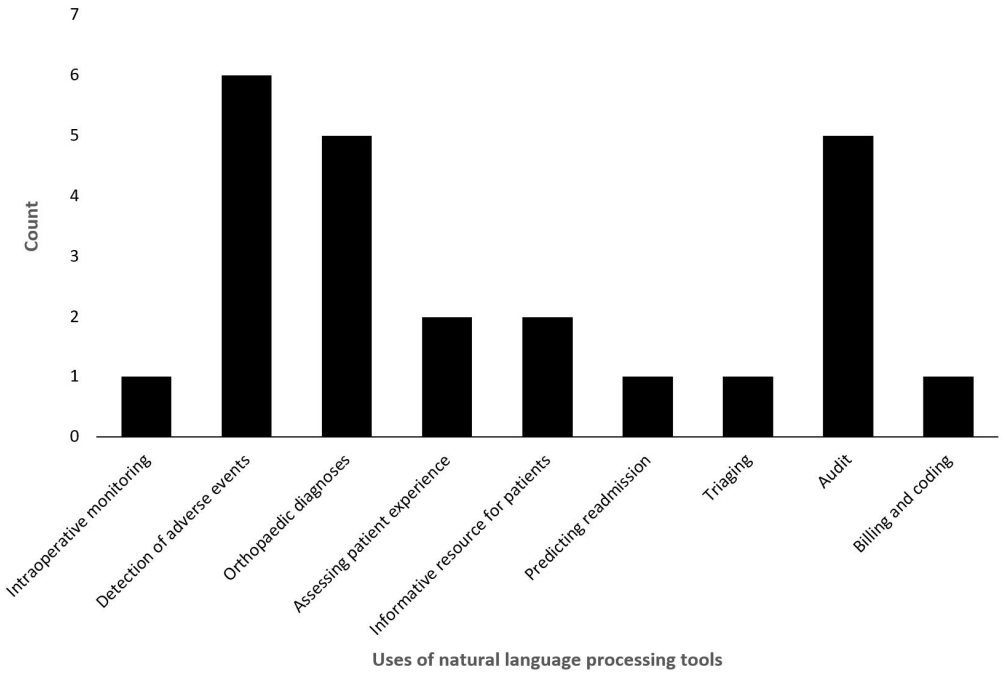

Figure 2. Overview of applications of natural language processing tools of included studies.

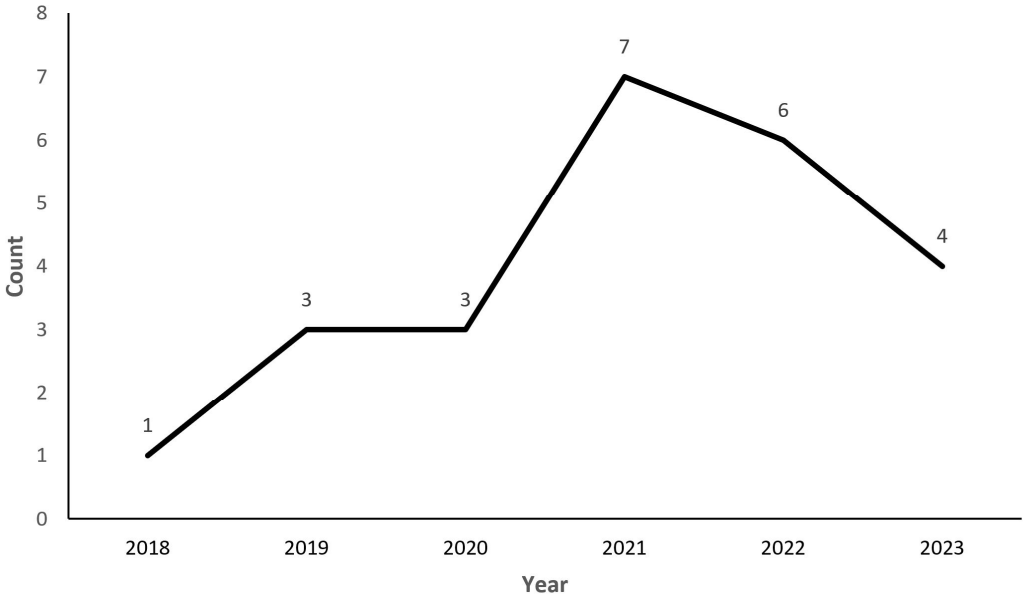

Figure 3. Publication year of included studies.

3.2. Use in Intraoperative Monitoring and Guidance

One study explored the intra-operative uses of NLP in orthopaedic surgery. Agaronnik et al. created a machine learning tool to retrospectively extract neuromonitoring status documentation from operative records of spine surgery [3]. They used an NLP keyword library to identify the relevant operative reports, followed by a deep learning model to characterise the documentation into a change in status, difficulty establishing baseline signals and stable course. Compared to the current gold standard manual chart review, NLP was able to more effectively and efficiently characterise the data. In particular, the NLP keyword library had an F1 score of 1.0, recall of 0.99 and precision of 1.0, compared to 0.64, 0.49, and 0.92, respectively, in current practice (i.e., using the International Classification of Disease (ICD) codes). The deep learning model for change in status of neuromonitoring

signals, difficulty establishing baseline and stable course had an F1 score of 0.8, 0.8 and 0.93, a precision score of 0.67, 0.71, 0.89 and a recall of 1.0, 0.91, 0.98, respectively.

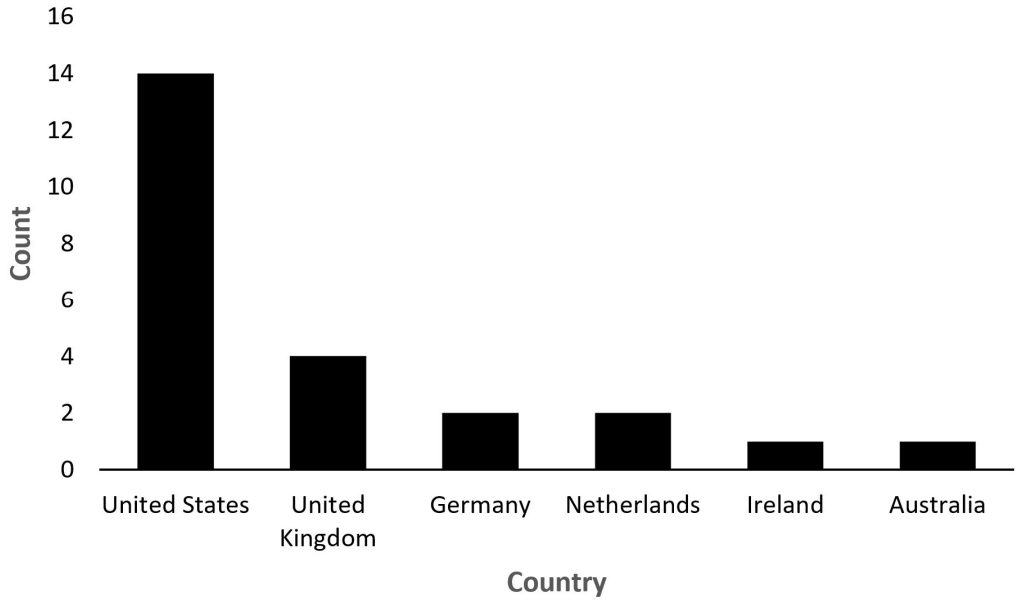

Figure 4. Country of publication of included studies.

3.3. Use in the Detection of Adverse Events

Six studies explored the use of NLP for the detection of adverse events related to orthopaedic surgery. Borjali et al. compared easy-to-access traditional machine learning NLP (ML-NLP) to two newly developed deep learning NLP (DL-NLP) models for retrospective detection of hip dislocations post total hip arthroplasty [4]. The models were used on radiology reports and telephone consult notes. DL-NLP models were highly accurate and performed better than ML-NLP. In particular, DL-NLP models were able to better interpret longer phrases, i.e., "no fracture or dislocation" as opposed to "no dislocation" and they were able to better delineate a dislocated hip from other displaced or dislocated joints or bones. Their convolutional neural network proposed DL-NLP model was the best performing with a kappa of 0.97 (for radiology reports) and 1.00 (for telephone consult notes). Finally, Borjali et al. also showed that 25% of dislocated hip patients identified with NLP did not have valid ICD codes, which infers the potential missed data that occur with traditional coders reviewing documentation [4].

Fu et al. applied NLP models to extract data from operative reports, consultation notes, microbiology and pathology reports related to periprosthetic joint infections (PJI) [5]. The NLP algorithm identified PJI based on the Musculoskeletal Infection Society criteria with an F1 score of 0.911 [5]. Thirukumaran et al. also identified that NLP models were able to correctly perform surgical site infection surveillance, based on their ability to extract information from administrative and medical records [6]. Compared to the gold standard, manual data abstraction, the NLP models precisely identified 97% of surgical site infections. Similarly, Karhade et al. identified that NLP was more accurate in identifying postoperative wound infection requiring reoperation post lumbar discectomy compared to ICD codes [7]. They showed that NLP detected 15 out of 16 cases (sensitivity of 0.94) compared to ICD codes, which detected 12 out of 16 (sensitivity 0.75).

Further, NLP was able to successfully extract data regarding intraoperative vascular injury in anterior lumbar spine surgery, identifying 18 out of 21 cases compared to ICD codes, which identified 6 of 21, with a sensitivity of 0.86 compared to 0.29, respectively [8]. Finally, NLP algorithms were found to be reproducible in geographically diverse populations for the identification of incidental durotomy post-spine surgery [9]. The same algorithm achieved an area under the curve receiver operating characteristic (AUC-ROC)

ranging between 0.95 and 0.99 in three separate geographically diverse cohorts (Australia, Massachusetts, Maryland).

3.4. Use for Orthopaedic Diagnoses

Five studies explored the role of NLP in assisting orthopaedic clinicians with formulating and identifying diagnoses. Li et al. trained NLP algorithms to interpret meniscal tears in magnetic resonance imaging (MRI) scan reports and later applied the same algorithm to arthroscopy reports [10]. For MRI reports, the algorithm had an F1 score of 0.93–0.94 for medial meniscus tears and an F1 score of 0.86–0.88 for lateral meniscus tears. With respect to scope reports, it had an F1 score of 0.97 and 0.99, respectively. Li et al. also combined scope and MRI reports to identify if NLP was able to identify mismatch between the two reports, resulting in a sensitivity of 79% and specificity 87%. Olthof et al. also used machine learning and deep learning NLP models on radiology reports—specifically orthopaedic trauma X-rays [11]. They compared various available models and identified the Bidirectional Encoder Representations from Transformers (BERT) as the most accurate, with an accuracy of 96% and an F1 score of 0.95. The radiology reports were in Dutch and therefore NLP proved to be effective in the non-English written language.

Groot et al. successfully used NLP algorithms to identify bony metastases in bone scintigraphy reports of patients undergoing surgery for bone metastases [12]. They found that the NLP algorithm had a sensitivity of 0.94 and specificity of 0.82, a positive predictive value of 0.97 and an F1 score of 0.96. Similarly, Tibbo et al. used an NLP algorithm to identify periprosthetic femoral fractures post total hip arthroplasties and compared its results to a diagnosis made by orthopaedic surgeons using chart and radiological review [13]. The algorithm was applied to operative and consult notes and had a sensitivity of 100% and specificity of 99.8% for identifying periprosthetic femoral fractures. With respect to identifying the correct Vancouver classification, it demonstrated a sensitivity of 78.6% and a specificity of 94.8%. Lastly, Tan et al. used an NLP system trained to identify lumbar spine imaging findings from both X-ray and MRI reports obtained from patients with lower back pain presentations [14]. In their study, four spine experts annotated the presence of 26 findings of which the NLP systems were able to achieve a high inter-rater agreement for 25 of 26 findings with a sensitivity of over 0.94 and an AUC of over 0.90, indicating the NLP system performed excellently when benchmarked to reference standard annotation by spinal experts.

3.5. Use in Assessing Patient Experience

Two studies explored the use of NLP in assessing patients' experience with orthopaedic surgery. Bovonratwet et al. used a predeveloped NLP model to extract patient satisfaction data post a total knee replacement and identify its impact on patient outcomes [15]. Whilst a comparison with non-NLP data extraction was not completed, they demonstrated that it is feasible to use NLP to extract relevant data required to create correlations and comparisons. Langerhuizen et al. used NLP to identify themes in online patient reviews of orthopaedic surgeons and their practice [16]. They compared the themes to identify correlations and ultimately found that patient–clinician interactions were the major contributor to patient satisfaction and therefore more effective communication training by surgeons could improve patient satisfaction and reviews.

3.6. Use as an Informative Resource for Patients

Two studies explored the role of NLP as a tool for the generation and provision of orthopaedic information. Dubin et al. compared the use of ChatGPT and Google web search using frequently asked questions by patients undergoing total hip replacements (THR) and total knee replacements (TKR) [17]. ChatGPT proved to use more reliable resources to answer the questions. In total, 15 out of 20 questions were answered with government websites (primarily PubMed) using ChatGPT, whereas 13 out of 20 questions asked on Google Web search were answered using commercial websites. Whilst this does not replace

the importance of education and consent by the treating physician, it has proven to be a potential reliable resource to educate patients on their orthopaedic procedures. Seth et al. prompted ChatGPT with a series of questions related to knee osteoarthritis and surgical management [18]. They found that the information presented was accurate, but relatively superficial and missed key contributing studies or elements of literature, hence proving that whilst ChatGPT could be used for generic patient education (with supplementation from the surgeon), it is not yet at the level where it can be used for orthopaedic research or generation of new ideas.

3.7. Predict Readmission

One study explored the ability of NLP to predict readmission for patients. Karhade et al. used an NLP algorithm on free text discharge summaries, operative notes and multidisciplinary progress notes from the medical record to estimate the risk of readmission within 90 days post lumbar spine fusion [19]. The area under the receiver–operating curve was 0.70, 0.57, 0.57, 0.60, 0.60 and 0.49, respectively. Discharge summaries were most useful at estimating risk, whilst daily progress notes provided little benefit.

3.8. Triaging

One study explored the use of NLP for triaging orthopaedic patients. Krebs et al. used NLP on MRI reports to predict whether patients with lower back pain or neck pain would end up needing a surgical intervention [20]. They found that NLP did not improve the accuracy of prediction to surgery. The three main predictive variables were lower back and leg pain, distal pain and difficulty walking. All three of these variables are best assessed via self-reported assessments as opposed to MRI reports. It is likely that Krebs et al. were not able to validate their current NLP model because they used radiological imaging as opposed to extracting qualitative data from patient reports and notes.

3.9. Audit

Five studies explored the use of NLP for auditing within orthopaedic surgery. Sagheb et al. used NLP algorithms to extract data from TKR operative reports and compared its accuracy to manual chart review [21]. They assessed the category of knee arthroplasty (total, unicompartmental, patellofemoral), laterality, constraint type, patella resurfacing and implant model. These showed an accuracy of 98.3%, 99.5%, 99.2%, 99.4%, and 99.9%, respectively. Wyles et al. completed a similar study using THR operation reports, assessing NLP accuracy in identifying the operative approach (anterolateral, direct anterior, posterior), fixation (uncemented, cemented, hybrid and reverse hybrid) and bearing surface (metal on polyethylene, ceramic on polyethylene, metal on metal, ceramic on ceramic) [22]. They identified an accuracy of 99.2%, 90.7% and 95.8%, respectively. They also validated these results externally using operative reports from a different institution and found an accuracy of 94.4%, 95.6% and 98%, respectively. Wyles et al. later applied it to a further 39 operative THR reports from a private practice "OrthoCarolina" to refine the algorithm and validate its use externally on a broader scale, they identified an accuracy of 100% for operative approach, fixation and bearing surface compared to manual chart review [23].

More broadly, Shah et al. applied NLP algorithms to 1000 randomly selected operative and hospital notes for patients undergoing a primary arthroplasty [24]. They used preoperative, operative and postoperative variables in their NLP algorithm, showing accuracy of 96.3%, sensitivity of 95.2% and specificity of 97.4%. The algorithm was better at detecting structured data, i.e., range of motion, as opposed to unstructured author-dependent information, i.e., written complications. This was compared to the gold standard manual chart review. Finally, Jungmann et al. utilised a pre-trained in-house NLP engine to categorise 5397 radiological reports (hand/wrist, elbow, shoulder, ankle, knee, pelvis/hip) to identify the incidence and age distribution of fractures during the COVID-19 pandemic [25]. In their study, the NLP engine achieved an F1 score of 0.81 when benchmarked against manual human annotation indicating that there was sound evidence to use these technologies for

epidemiological studies, auditing of cases and real-time monitoring of fractures. Overall, these studies indicate that NLP has a promising role in data extraction from joint registries and can contribute to auditing and research processes.

3.10. Billing and Coding

One study explored the use of NLP for billing and coding practices in orthopaedic surgery. Zaidat et al. showed that NLP can generate Current Procedural Terminology (CPT) codes on operative notes of patients who underwent anterior cervical discectomy and fusion (ACDF), posterior cervical decompression and fusion (PCDF), a combination of the two procedures or cervical disc arthroplasty (CDA) [26]. ACDF operative notes only had an area under the receiver–operator curve (AUROC) of 0.82, an area under the precision–recall curve of 0.81, and an accuracy of 77%. PCDF had an AUROC of 0.82, precision of 0.7 and accuracy of 71%. All operative notes analysed together yielded an AUROC of 0.95, precision of 0.84 and accuracy of 88%. Hence, NLP could generate CPT codes in a comparable manner to the current gold standard code generation by the billing department.

3.11. Quality Assessment

Quality assessment was performed utilising the Newcastle–Ottawa Scale and ranged from low quality (3/9) to high quality (8/9) (Appendix A). The median NOS score achieved was 6 with an interquartile range of 2.5, indicating that despite the high variability, overall the included studies are considered to be of moderate quality.

4. Discussion

As healthcare continues its shift towards digitalisation, the potential for leveraging technology to enhance the orthopaedic surgery experience becomes evident for all involved parties, from surgeons to patients and allied health staff. Yet, the realm of orthopaedics lags in a structured, guideline-based approach to integrating these digital tools, with limited research exploring their specific application into streamlining and augmenting processes in orthopaedics. As orthopaedics becomes further digitalised, with advancements like electronic medical records, the capabilities of NLP and generative AI tools stand out. These technologies, with their vast computational power and adaptability, hold the promise of optimising efficiency, task streamlining, and auditing processes. This review aims to answer this research question pertaining to how we can best utilise these AI tools by presenting a comprehensive overview of existing research on the use of NLP and generative AI within orthopaedic surgery. Furthermore, it sheds light on potential future research trajectories, aiming to inform the creation of evidence-based strategies and frameworks for NLP technology adoption in this field.

Our analysis encompassed and synthesised the current traits, evidence, and potential of NLP and generative AI tools from 24 publications, focusing on their influence in the realm of orthopaedics. A key finding is the current absence of evidence-based guidelines for these tools' application within orthopaedics. There has been an emerging body of research since 2018 that started to explore the strategies by which implementation of NLP could improve processes within orthopaedic surgery. Notably, the majority of this research hails from the United States, the birthplace of popular tools like ChatGPT and Google Bard.

For implementation, our scoping review identified one study by Agaronnik et al. that assessed the use of NLP for intraoperative monitoring and guidance in spinal surgery [3]. They found that NLP could with high accuracy, as compared to manual chart review, retrospectively identify important events including changes in patient status and baseline signals, stability and change in neuromonitoring signals. Practically these early findings could establish a role for NLP in retrospectively identifying and evaluating events that occur during surgery for learning and auditing processes. However, the true potential lies in augmenting these NLP systems by integrating them into modern monitoring tools to potentially develop a real-time system that can identify important intraoperative changes to support more efficient and rapid decision making for orthopaedic surgeons. This same

concept could be applied to other orthopaedic surgeries, such as hip arthroplasties where monitoring the sciatic or femoral nerves can have significant importance to patient outcomes.

Our review pinpointed burgeoning research into NLP's applications for diagnostic purposes, specifically for identifying adverse events linked to orthopaedic surgery and assisting in pinpointing orthopaedic diagnoses. Six studies probed the deployment of NLP for these tasks, leveraging its capability to extract data from diverse digital documents such as radiology reports, consultation notes, and pathology reports [4–9]. In all cases, NLP is able to correctly identify relevant adverse outcomes including hip dislocations, periprosthetic joint and surgical site infection and intraoperative vascular injury. Furthermore, five studies explored the use of NLP in assisting with the diagnostic process for orthopaedic patients [10–14]. In these studies, NLP was able to extract data from radiology reports of multiple modalities to accurately define orthopaedic diagnoses including meniscal tears, bone metastases, periprosthetic fractures and traumatic orthopaedic injuries. Interestingly, one study could perform this task effectively from non-English radiology reports [11].

Historically, orthopaedic clinicians have depended on manual chart reviews, data accumulation, and synthesis as the gold standard for gathering adverse outcomes or diagnostic data. Efforts to enhance the efficiency of these processes have included the transition of healthcare systems from paper-based to electronic medical records, allowing all the information to be found on one interface. This has allowed clinicians to streamline these processes by allowing computer-generated searches of adverse outcomes or diagnoses from relevant investigations as required.

However, this digital transition has its set of challenges, primarily due to the proliferation of unstructured text, vast data quantities, and user interfaces that might not effectively capture pertinent data. There is clear emerging evidence that NLP offers a solution for evaluating vast swathes of digital healthcare data in a highly accurate and efficient manner. Integration of these technologies could theoretically improve the time orthopaedic surgeons and trainees spend in auditing and data collection for research. This is supported by five additional studies which demonstrate NLP is able to extract with high accuracy various outcome measures and important operative details [21–25]. In almost all of these cases, hundreds to thousands of separate documents were evaluated with these tools, a task that would be laborious and highly prone to human error, particularly if factors including fatigue, burnout and staff shortages are considered. The allure of automated, swift data extraction, processing, and synthesis through NLP is undeniable. Nevertheless, it is pivotal to acknowledge the need for further refinement and training of these tools to hone their precision.

Our scoping review identified various preliminary studies that explored unique applications of NLP in orthopaedic surgery including in the assessment of the patient experience, as an informative resource for patients, in the prediction of readmission, in assistance with billing/coding procedures and in triaging of patients [15–20,26]. With the exception of predicting readmission and triage, NLP was adequately and efficiently used for these purposes. The ability of NLP to assess the patient experience and provide advice may represent the beginnings of a new paradigm shift in the way we harness digital tools outside of the acute care setting. For the former, sentiment analysis of patient experience can be streamlined and better represented with NLPs [15,16]. This may allow not only orthopaedic departments but also various quality and safety units within hospital systems to best understand how our health services can improve the delivery of healthcare. Furthermore, the ability of ChatGPT to provide answers to frequently asked questions associated with hip and knee arthroplasty may be a useful adjunct for surgeons to present information in more accessible and palatable ways for patients [17,18]. This is particularly relevant in an environment where diverse caseloads of patients exist. More streamlined generation of CPT codes from operative notes by NLP tools is of great benefit to health services, particularly as this would improve the workload and efficiency of clinical coders [26]. Importantly, allowing surgeons to be prompted with specific CPT codes based on their operation notes, may yield more accurate and comprehensive documentation. The downstream impacts of this

include improved healthcare funding, more appropriate remuneration of clinicians and better identification of areas for resource allocation. Notably, when it came to predicting readmission and triage, Karhade et al. and Krebs et al., respectively, found that NLP was unable to assist with these processes with high accuracy [19,20]. Specifically, Karhade et al. demonstrated low to moderate AUC scores when using NLP to predict 90-day readmission and Krebs et al. demonstrated NLP was unable to be validated for triaging patients with neck or lower back pain in operative or nonoperative populations [19,20]. These studies highlight that we are far from close to utilising NLP for decision making or prognostication within orthopaedic surgery. Specifically, these clinical decisions still require the dynamic and well-trained insights of orthopaedic surgeons.

A strength of this scoping review is the diverse range of studies with a wide variety of methods, objectives and outcomes evaluating the use of NLPs within the field of orthopaedic surgery. In fact, to our knowledge, this review is the first to systematically search for and synthesise the literature in this space to gain a comprehensive understanding of the currently studied applications of NLPs within orthopaedics as well as the potential avenues to explore with respect to future directions. Another strength of the review is that it explores both quantitative and qualitative outcomes related to the implementation of NLPs within orthopaedics, thereby allowing a better understanding of the current benefits, downsides and challenges of integrating this technology at this point in time.

Limitations of this review relate to the paucity of research within this space, particularly with many of the applications of NLPs proposed remaining largely theoretical or poorly characterised. Given that NLPs are a new technology and efforts to integrate these technologies within orthopaedic surgery are still in their infancy, the current literature exploring their applications is also of poor methodological quality, largely theoretical or aimed at assessing feasibility. Consequently, it remains difficult to gain a deep understanding of the true effect of implementing NLPs in the various stages of orthopaedic surgery. Furthermore, from our quality assessment of the included studies using the NOS, the quality of papers was highly variable but overall of moderate quality. Given this, the results of this review should be considered with caution but nonetheless provide exciting insights into the beginnings of further research and applications of NLPs within orthopaedic surgery.

Regardless, in all cases, NLP tools were used to extract a significant variety of unstructured and structured free-form text information. The general theme was that NLP tools did so with moderate to high accuracy, sensitivity and specificity when compared with current manual approaches. Understandably, given these findings and the tremendous computational capacity of NLP algorithms, it is expected that the implementation of NLP confers significant potential in improving efficiency and streamlining processes within all stages of the orthopaedic journey.

Given that NLP is a new technology and efforts to integrate these technologies within orthopaedic surgery are still in their infancy, the current literature exploring their applications is also of poor methodological quality, largely theoretical or is aimed at assessing feasibility. Consequently, it remains difficult to gain a deep understanding of the true effect of implementing NLP in orthopaedic surgery. Furthermore, from our quality assessment of the included studies using the NOS, the quality of papers was highly variable but overall of moderate quality. Given this, the results of this review should be considered with caution.

The prospect of integrating NLP into orthopaedics is highly appealing. However, as for all new technologies, some significant barriers and challenges must be considered before seamless integration. Specific to healthcare in general is the potential for error that may arise by blindly trusting the outputs from NLP algorithms. Zhu et al. demonstrated that when five different NLP systems were subjected to a set of 22 questions from a prostate cancer community, 90% of answers were able to be answered appropriately however these systems lacked the ability to navigate queries that resulted from further clarification, nor could they provide empathic comfort to patients [27]. Similarly, in a study by Haemmerli et al. whereby ChatGPT was used to provide advice based on analysing glioma diagnoses from 10 patients, the algorithm demonstrated poor ability to classify glioma by type and

subsequently lacked nuance when medical advice for these tumours was requested [28]. These studies suggest that these tools remain imperfect, with a very low likelihood that they can replace any role that requires specialist medical input.

Furthermore, without the right safeguards, these tools might pose harm. For instance, ChatGPT has demonstrated the ability to produce persuasive yet incorrect information when queried about ophthalmological diagnoses [29]. This underscores the need to assess what human oversight is necessary to monitor these tools and facilitate their integration. In an era marked by meticulous healthcare budgeting and spending, there is a demand for compelling evidence before fully endorsing the integration of generative AI technologies. A deeper dive into research is imperative to discern the specific scenarios within orthopaedics where these tools would be most beneficial.

Lastly, the integration of NLP technology comes with strong ethical considerations. Given the tremendous computational demands of NLP, it is unlikely, nor cost-effective, to develop in-house NLP algorithms with the same capacity as current open-access tools. It is likely that integration of these would leverage the current infrastructure available on the market, which raises the question about the confidentiality and privacy of healthcare data. In an environment where data can be used for harm and cybersecurity measures are at the forefront of digital health, clinicians and policymakers must consider the nature of the personal data that these tools may receive, where it is stored, how secure it is, data ownership and how data are used by larger corporations. Perhaps for a single specialty like orthopaedics, the benefit of this is far outweighed by the costs. However, when it comes to overall healthcare, health-centric sub-brands of these current NLP tools, with a specific focus on evidence-based training from healthcare databases, stronger security and transparent processes behind data storage, may provide a more reassuring avenue for this digital transition. For this to take place, significant multi-sectoral partnerships would need to occur and the cost of this process would need to be strongly considered.

5. Conclusions

The advent of widely accessible natural language processing (NLP) tools has spurred their adoption across various sectors, including healthcare. Despite this, there is currently a lack of evidence-based understanding of how to best integrate these tools, as well as a lack of best-practice methodologies to optimise the implementation of these tools within orthopaedic surgery. This scoping review is the first to provide broad insight into the potential applications of contemporary NLP technologies in the field of orthopaedics. Our review has demonstrated the significant potential of these tools from the foundational literature to rapidly and efficiently extract and repurpose digital data for a multitude of tasks through the orthopaedic journey. There is a pressing need for more in-depth studies to ascertain how to refine, integrate, and commercialise these tools in orthopaedic surgery. Such advancements could catalyse a transformative shift in harnessing digital data within the discipline.

Author Contributions: Conceptualization, K.D.R.L., F.S. and J.W.V.; methodology, K.D.R.L., F.S. and S.B.P.T.; validation, F.S. and K.D.R.L.; formal analysis, F.S. and K.D.R.L.; investigation, F.S. and K.D.R.L.; data curation, K.D.R.L. and F.S.; writing—original draft preparation, K.D.R.L., F.S. and S.B.P.T.; writing—review and editing, K.D.R.L. and F.S.; visualization, K.D.R.L. and F.S.; supervision, P.T. and J.W.V.; project administration, K.D.R.L. and F.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data can be re-generated by replicating the search query found in Appendix B.

Conflicts of Interest: There are no conflict of interest, disclaimers or financial and grant support.

Appendix A

Table A1. Quality analysis of included studies using Newcastle–Ottawa Score.

Study	Representative of the Exposed Cohort	Selection of External Control/ Non-Exposed Cohort	Ascertainment of Exposure	Outcome of Interest Not Present at the Start of the Study	Study Controls for Intervention of Natural Language Processing Tool	Study Control for External Confounders	Assessment of Outcomes/ Ascertainment of Exposure	Sufficient Follow-Up/Same Method of Ascertainment for Cases and Controls	Adequacy of Follow-up/ Non-Response Rate	Total Score (/9)
Agaronnik et al., 2022 [3]	+	-	+	+	+	+	+	+	-	7
Borjali et al., 2021 [4]	+	-	+	+	+	+	+	+	-	7
Bovonratwet et al., 2021 [15]	-	-	+	+	-	-	+	-	-	3
Dubin et al., 2023 [17]	-	-	+	+	-	-	+	-	-	3
Fu et al., 2021 [5]	+	-	+	+	-	-	+	-	-	4
Groot et al., 2020 [12]	+	-	+	+	+	+	+	-	-	6
Junmann et al., 2022 [25]	+	-	+	+	+	+	+	+	-	7
Karhade et al., 2020 [7]	+	-	+	+	+	+	+	+	-	7
Karhade et al., 2021 [8]	+	-	+	+	+	+	+	+	-	7
Karhade et al., 2022 [19]	+	-	+	+	+	+	+	+	-	7
Karhade et al., 2022 [9]	+	-	+	+	+	+	+	+	-	7
Krebs et al., 2023 [20]	-	-	+	+	+	+	+	+	-	6
Langerhuizen et al., 2021 [16]	+	-	+	+	-	-	+	-	-	4
Li et al., 2022 [10]	+	-	+	+	-	-	+	-	-	4
Olthof et al., 2021 [11]	+	-	+	+	+	+	+	+	-	7
Sagheb et al., 2021 [21]	+	-	+	+	+	+	+	+	-	7
Seth et al., 2023 [18]	-	-	+	+	-	-	+	=	=	3
Shah et al., 2020 [24]	+	-	+	+	+	-	+	+	-	6
Tan et al., 2018 [14]	+	+	+	+	+	+	+	+	-	8
Thirukumaran et al., 2019 [6]	-	+	+	+	+	+	+	+	-	7
Tibbo et al., 2019 [13]	-	-	+	+	+	+	+	-	-	5
Wyles et al., 2022 [23]	-	-	+	+	+	-	+	+	-	5
Wyles et al., 2019 [22]	-	-	+	+	+	-	+	+	-	5
Zaidat et al., 2023 [26]	+	-	+	+	+	-	+	+	-	6

Appendix B. Search Query

1. * natural language processing
2. (chatgpt * or chat gpt * or "GPT-3" or "GPT-4" or language model * or natural language processing).ti,kf.
3. ((Gopher or Chinchilla or Google Bard or Perplexity or SpaCy or Stanford Core NLP or NTLK) adj6 (AI or language model *)).mp.
4. 1 or 2 or 3
5. exp surgery/or exp perioperative complication/or exp perioperative care/or exp surgical training/or "surg *".jw, ti, kw.
6. 4 and 5
7. ((chatgpt * or chat gpt* or "GPT-3" or "GPT-4" or language model* or natural language processing) adj6 (surg* or neurosurg * or orthop?edic * or operative * or preoperative * or postoperative * or intraoperative *)).mp.
8. 6 or 7
9. limit 8 to english language

References

1. Esteva, A.; Robicquet, A.; Ramsundar, B.; Kuleshov, V.; DePristo, M.; Chou, K.; Cui, C.; Corrado, G.; Thrun, S.; Dean, J. A guide to deep learning in healthcare. *Nat. Med.* **2019**, *25*, 24–29. [CrossRef] [PubMed]
2. Wu, T.; He, S.; Liu, J.; Sun, S.; Liu, K.; Han, Q.-L.; Tang, Y. A brief overview of ChatGPT: The history, status quo and potential future development. *IEEE/CAA J. Autom. Sin.* **2023**, *10*, 1122–1136. [CrossRef]
3. Agaronnik, N.D.; Kwok, A.; Schoenfeld, A.J.; Lindvall, C. Natural language processing for automated surveillance of intraoperative neuromonitoring in spine surgery. *J. Clin. Neurosci.* **2022**, *97*, 121–126. [CrossRef] [PubMed]
4. Borjali, A.; Magnéli, M.; Shin, D.; Malchau, H.; Muratoglu, O.K.; Varadarajan, K.M. Natural language processing with deep learning for medical adverse event detection from free-text medical narratives: A case study of detecting total hip replacement dislocation. *Comput. Biol. Med.* **2020**, *129*, 104140. [CrossRef] [PubMed]
5. Fu, S.; Wyles, C.C.; Osmon, D.R.; Carvour, M.L.; Sagheb, E.; Ramazanian, T.; Kremers, W.K.; Lewallen, D.G.; Berry, D.J.; Sohn, S.; et al. Automated Detection of Periprosthetic Joint Infections and Data Elements Using Natural Language Processing. *J. Arthroplast.* **2021**, *36*, 688–692. [CrossRef] [PubMed]
6. Thirukumaran, C.P.; Zaman, A.; Rubery, P.T.; Calabria, C.; Li, Y.; Ricciardi, B.F.; Bakhsh, W.R.; Kautz, H. Natural Language Processing for the Identification of Surgical Site Infections in Orthopaedics. *J. Bone Jt. Surg. Am.* **2019**, *101*, 2167–2174. [CrossRef]
7. Karhade, A.V.; Bongers, M.E.; Groot, O.Q.; Cha, T.D.; Doorly, T.P.; Fogel, H.A.; Hershman, S.H.; Tobert, D.G.; Schoenfeld, A.J.; Kang, J.D.; et al. Can natural language processing provide accurate, automated reporting of wound infection requiring reoperation after lumbar discectomy? *Spine J.* **2020**, *20*, 1602–1609. [CrossRef]
8. Karhade, A.V.; Bongers, M.E.; Groot, O.Q.; Cha, T.D.; Doorly, T.P.; Fogel, H.A.; Hershman, S.H.; Tobert, D.G.; Srivastava, S.D.; Bono, C.M.; et al. Development of machine learning and natural language processing algorithms for preoperative prediction and automated identification of intraoperative vascular injury in anterior lumbar spine surgery. *Spine J.* **2021**, *21*, 1635–1642. [CrossRef]
9. Karhade, A.V.; Oosterhoff, J.H.; Groot, O.Q.; Agaronnik, N.; Ehresman, J.; Bongers, M.E.; Jaarsma, R.L.; Poonnoose, S.I.; Sciubba, D.M.; Tobert, D.G.; et al. Can We Geographically Validate a Natural Language Processing Algorithm for Automated Detection of Incidental Durotomy Across Three Independent Cohorts From Two Continents? *Clin. Orthop. Relat. Res.* **2022**, *480*, 1766–1775. [CrossRef]
10. Li, M.D.; Deng, F.; Chang, K.; Kalpathy-Cramer, J.; Huang, A.J. Automated Radiology-Arthroscopy Correlation of Knee Meniscal Tears Using Natural Language Processing Algorithms. *Acad. Radiol.* **2022**, *29*, 479–487. [CrossRef]
11. Olthof, A.W.; Shouche, P.; Fennema, E.M.; IJpma, F.F.; Koolstra, R.C.; Stirler, V.M.; van Ooijen, P.M.; Cornelissen, L.J. Machine learning based natural language processing of radiology reports in orthopaedic trauma. *Comput. Methods Programs Biomed.* **2021**, *208*, 106304. [CrossRef]
12. Groot, O.Q.; Bongers, M.E.; Karhade, A.V.; Kapoor, N.D.; Fenn, B.P.; Kim, J.; Verlaan, J.J.; Schwab, J.H. Natural language processing for automated quantification of bone metastases reported in free-text bone scintigraphy reports. *Acta Oncol.* **2020**, *59*, 1455–1460. [CrossRef]
13. Tibbo, M.E.; Wyles, C.C.; Fu, S.; Sohn, S.; Lewallen, D.G.; Berry, D.J.; Kremers, H.M. Use of Natural Language Processing Tools to Identify and Classify Periprosthetic Femur Fractures. *J. Arthroplast.* **2019**, *34*, 2216–2219. [CrossRef]
14. Tan, W.K.; Hassanpour, S.; Heagerty, P.J.; Rundell, S.D.; Suri, P.; Huhdanpaa, H.T.; James, K.; Carrell, D.S.; Langlotz, C.P.; Organ, N.L.; et al. Comparison of Natural Language Processing Rules-based and Machine-learning Systems to Identify Lumbar Spine Imaging Findings Related to Low Back Pain. *Acad. Radiol.* **2018**, *25*, 1422–1432. [CrossRef] [PubMed]
15. Bovonratwet, P.; Shen, T.S.; Islam, W.; Ast, M.P.; Haas, S.B.; Su, E.P. Natural Language Processing of Patient-Experience Comments After Primary Total Knee Arthroplasty. *J. Arthroplast.* **2021**, *36*, 927–934. [CrossRef] [PubMed]

16. Langerhuizen, D.W.G.; Brown, L.E.; Doornberg, J.N.; Ring, D.; Kerkhoffs, G.M.M.J.; Janssen, S.J. Analysis of Online Reviews of Orthopaedic Surgeons and Orthopaedic Practices Using Natural Language Processing. *J. Am. Acad. Orthop. Surg.* **2021**, *29*, 337–344. [CrossRef]
17. Dubin, J.A.; Bains, S.S.; Chen, Z.; Hameed, D.; Nace, J.; Mont, M.A.; Delanois, R.E. Using a Google Web Search Analysis to Assess the Utility of ChatGPT in Total Joint Arthroplasty. *J. Arthroplast.* **2023**, *38*, 1195–1202. [CrossRef] [PubMed]
18. Seth, I.; Rodwell, A.; Bulloch, G.; Seth, N. Exploring the role of open artificial intelligence platform on surgical management of knee osteoarthritis: A case study of ChatGPT. *J. Clin. Cases Rep.* **2023**, *13*, 6. [CrossRef]
19. Karhade, A.V.; Lavoie-Gagne, O.; Agaronnik, N.; Ghaednia, H.; Collins, A.K.; Shin, D.; Schwab, J.H. Natural language processing for prediction of readmission in posterior lumbar fusion patients: Which free-text notes have the most utility? *Spine J.* **2022**, *22*, 272–277. [CrossRef]
20. Krebs, B.; Nataraj, A.; McCabe, E.; Clark, S.; Sufiyan, Z.; Yamamoto, S.S.; Zaïane, O.; Gross, D.P. Developing a triage predictive model for access to a spinal surgeon using clinical variables and natural language processing of radiology reports. *Eur. Spine J.* **2023**. [CrossRef]
21. Sagheb, E.; Ramazanian, T.; Tafti, A.P.; Fu, S.; Kremers, W.K.; Berry, D.J.; Lewallen, D.G.; Sohn, S.; Kremers, H.M. Use of Natural Language Processing Algorithms to Identify Common Data Elements in Operative Notes for Knee Arthroplasty. *J. Arthroplast.* **2020**, *36*, 922–926. [CrossRef]
22. Wyles, C.C.; Tibbo, M.E.; Fu, S.; Wang, Y.; Sohn, S.; Kremers, W.K.; Berry, D.J.; Lewallen, D.G.; Maradit-Kremers, H. Use of Natural Language Processing Algorithms to Identify Common Data Elements in Operative Notes for Total Hip Arthroplasty. *J. Bone Jt. Surg. Am.* **2019**, *101*, 1931–1938. [CrossRef]
23. Wyles, C.C.; Fu, S.; Odum, S.L.; Rowe, T.; Habet, N.A.; Berry, D.J.; Lewallen, D.G.; Maradit-Kremers, H.; Sohn, S.; Springer, B.D. External Validation of Natural Language Processing Algorithms to Extract Common Data Elements in THA Operative Notes. *J. Arthroplast.* **2023**, *38*, 2081–2084. [CrossRef]
24. Shah, R.F.; Bini, S.; Vail, T. Data for registry and quality review can be retrospectively collected using natural language processing from unstructured charts of arthroplasty patients. *Bone Jt. J.* **2020**, *102-B*, 99–104. [CrossRef]
25. Jungmann, F.; Kämpgen, B.; Hahn, F.; Wagner, D.; Mildenberger, P.; Düber, C.; Kloeckner, R. Natural language processing of radiology reports to investigate the effects of the COVID-19 pandemic on the incidence and age distribution of fractures. *Skelet. Radiol.* **2021**, *51*, 375–380. [CrossRef] [PubMed]
26. Zaidat, B.; Tang, J.; Arvind, V.; Geng, E.A.; Cho, B.; Duey, A.H.; Dominy, C.; Riew, K.D.; Cho, S.K.; Kim, J.S. Can a Novel Natural Language Processing Model and Artificial Intelligence Automatically Generate Billing Codes From Spine Surgical Operative Notes? *Glob. Spine J.* **2023**. [CrossRef]
27. Zhu, L.; Mou, W.; Chen, R. Can the ChatGPT and other large language models with internet-connected database solve the questions and concerns of patient with prostate cancer and help democratize medical knowledge? *J. Transl. Med.* **2023**, *21*, 269. [CrossRef] [PubMed]
28. Haemmerli, J.; Sveikata, L.; Nouri, A.; May, A.; Egervari, K.; Freyschlag, C.; Lobrinus, J.A.; Migliorini, D.; Momjian, S.; Sanda, N.; et al. ChatGPT in glioma adjuvant therapy decision making: Ready to assume the role of a doctor in the tumour board? *BMJ Health Care Inform.* **2023**, *30*, e100775. [PubMed]
29. Balas, M.; Ing, E.B. Conversational AI Models for ophthalmic diagnosis: Comparison of ChatGPT and the Isabel Pro Differential Diagnosis Generator. *JFO Open Ophthalmol.* **2023**, *1*, 100005. [CrossRef]