OccuTriage: An AI Agent Orchestration Framework for Occupational Health Triage Prediction

Alok Kumar Sahu[1] and **Yi Sun**[2] and **Eamonn Swanton**[1]
Farshid Amirabdollahian[2] and **Abi Wren**[1]
[1]Heales Medical, Hitchin, Hertfordshire, England, United Kingdom
[2]University of Hertfordshire, Hatfield, England, United Kingdom
{alok.sahu, eamonn.swanton, abi.wren}@heales.com
{y.2.sun, f.amirabdollahian2}@herts.ac.uk

Abstract

Occupational Health (OH) triage is a systematic process for evaluating and prioritising workplace health concerns to determine appropriate care and interventions. This research addresses critical triage challenges through our novel AI agent orchestration framework, OccuTriage, developed in collaboration with Heales Medical[1] . Our framework simulates healthcare professionals' reasoning using specialized LLM agents, retrieval augmentation with domain-specific knowledge, and a bidirectional decision architecture. Experimental evaluation on 2,589 OH cases demonstrates OccuTriage outperforms single-agent approaches with a 20.16% average discordance rate compared to baseline rates of 43.05%, while matching or exceeding human expert performance (25.11%). The system excels in reducing under-triage rates, achieving 9.84% and 3.1% for appointment and assessor type decisions respectively. These results establish OccuTriage's efficacy in performing complex OH triage while maintaining safety and optimizing resource allocation.

1 Introduction

Triage, the systematic prioritization of cases based on urgency and resource constraints, is essential in occupational healthcare delivery. The Royal College of Occupational Therapists advocates for prioritizing referrals through analysis of need levels and resource optimization (Mandelstam, 2005).

1.1 Triage Frameworks in Occupational Health

Structured frameworks have emerged to standardize triage in occupational healthcare. (CARIBE et al., 2020) developed a questionnaire-based algorithm for occupational health nursing, while (Jones and Greenberg, 2015) implemented the TAG-triage approach, reducing assessment time by 72% while maintaining clinical effectiveness. (Sands et al., 2016) created a seven-tier system with defined urgency time-frames.

For complex cases, (Lalloo et al., 2021) established a comprehensive framework with three domains (health, workplace, and biopsychosocial factors) containing 27 specific elements, representing significant advancement over earlier single-dimension models.

1.2 Triage Implementation and Applications

In practice, (Walker-Bone et al., 2020) deployed an effective three-tier RED/AMBER/GREEN system during COVID-19. The 'telephone first' methodology by (O'Reilly and McDonnell, 2020) and (O'reilly and Carr) demonstrated remarkable efficiency, reducing waiting times by 77% and resolving approximately half of consultations remotely.

For specific conditions, (Green et al., 2024) employed symptom questionnaires for post-COVID syndrome, identifying fatigue as the strongest predictor of work inability. For musculoskeletal disorders, (McCluskey et al., 2006) implemented a biopsychosocial approach that significantly reduced absence duration. Notably, (Gorick et al., 2024) found experienced nurses prioritize visual assessment and clinical judgment over algorithms.

1.3 Machine Learning and AI in Triage

Machine learning has transformed healthcare triage. In emergency departments, (Fernandes et al., 2020) showed logistic regression dominated triage Clinical Decision Support Systems. (Jiang et al., 2021) implemented four machine learning models for cardiovascular triage, with XGBoost achieving highest performance. More sophisticated approaches include (Mutegeki et al., 2023)'s interpretable Histogram-Based Gradient Boosting classifier and (Xie et al., 2021)'s Score for Emergency Risk Prediction. In occupational health specifically, (Weng et al., 2020) developed a surveillance system using NLP and logistic regression.

[1]https://www.heales.com/

Large Language Models (LLMs) have created new triage opportunities. (Uronen et al., 2022) combined supervised BERT-NER and unsupervised query expansion to detect psychosocial risk factors in occupational health checks. (Krastev et al., 2023) proposed a semantic interoperability approach for Occupational Health Assessment Summary. (Kopka et al., 2024)'s RepVig Framework showed LLMs achieved 67.6% accuracy with representative vignettes, performing better on non-emergency cases than emergency cases.

Healthcare-specific LLMs include Med-PaLM Multimodal (Tu et al., 2023), Clinical Camel (Toma et al., 2023), and Asclepius (Kweon et al., 2023). Multi-agent frameworks have emerged for complex triage tasks, with (Lu et al.)'s TRIAGEAGENT utilizing retrieval-augmented generation, achieving up to 18.42% improvement over baselines using GPT-4 (OpenAI et al., 2023). We use LLama[2] and Asclepius[3] to evaluate the performance of our proposed OccuTriage framework against different benchmark tecnhiques.

1.4 Research Gap Addressed

Our review reveals critical gaps in the literature. Traditional triage frameworks remain largely manual, with practitioners preferring clinical judgment over algorithms (Gorick et al., 2024). Current LLM-based systems show variable accuracy depending on case complexity (Kopka et al., 2024). While promising, multi-agent systems like those by (Lu et al.) and (Han and Choi, 2024) focus primarily on emergency departments rather than occupational health settings.

Our research addresses these limitations through a novel AI agent orchestration framework that bridges clinical judgment and algorithmic approaches with: (1) a multi-agent system with specialized AI agents simulating clinical expertise, (2) retrieval augmentation with external knowledge bases, (3) an iterative discussion protocol with safety-prioritized decision rules, and (4) a bidirectional decision architecture enabling comprehensive coverage across multiple triage conditions.

[2] https://huggingface.co/meta-llama/Llama-3.1-8B-Instruct

[3] https://huggingface.co/starmpcc/Asclepius-Llama3-8B

2 Methodology

2.1 Problem Setup

Occupational Electronic Health Record (EHR) data comprises referral forms (r) and associated attachments (a) such as medical records and job descriptions. Our dataset is represented as $D = \{M_1, M_2, ..., M_n\}$, where $M_i = r_i + a_i$ for the ith medical record. For each case M_i, we predict two triage outcomes:

1. Appointment triage outcome (Y_i^1): Face-to-face or video appointment

2. Assessor triage outcome (Y_i^2): Appropriate Occupational Health Assessor (nurse or doctor)

2.2 Retrieval Augmentation with External Database

To enhance interpreting complex medical terminologies in referral forms, we augment content with information from external sources including job descriptions, medical terminology explanations, and medication details.

Knowledge Base Creation. We incorporate knowledge from diverse external sources into text representation format to enable semantic-based retrieval, represented as $E = \{k_1, k_2, ..., k_m\}$, where m is the total number of text vectors in the corpus. Our knowledge base integrates two specialized resources: the NCI Thesaurus providing comprehensive biomedical terminology with cancer-related clinical and molecular information (Sioutos et al., 2007), and O*NET OnLine (National Center for O*NET Development, 2025) supplying detailed occupational information across multiple dimensions. This integration enables more nuanced semantic understanding and improves domain-specific information retrieval in biomedical and occupational contexts.

Document Anonymization. We employ LLMs to detect and anonymize personal information in unstructured data following recent advances in adversarial anonymization techniques (Staab et al.). We represent the anonymized version of case M_i as M_i'.

Corpus Embedding. Following (Cheng et al., 2023), we use Dragon (Lin et al., 2023), a dual encoder model with strong cross-domain performance, as our retriever. We use the passage encoder E_p to encode passages from E, and the query encoder E_q during runtime to retrieve the relevant results.

Medical Entity Extraction. We leverage LLMs to extract medical entities, as they better understand contextual nuances and recognize specialized terminology in non-standard formats.

Medical Document Summarizer. The Summarizer component (S) processes both anonymized records and retrieved knowledge to produce comprehensive case representations. For each anonymized record M_i', it generates a condensed representation $S_i = LLM(M_i', k_i')$, where k_i' represents relevant knowledge retrieved from E.

Information Retrieval. We encode medical entities using E_q and retrieve the most relevant information (top-k, where k=1) from E_p.

2.3 AI Agent Orchestration Framework

Our framework simulates triage rules practiced by Heales Medical with heterogeneously orchestrated agents divided into two crews, each supervised by dedicated chat managers. Crew 1 is managed by C_M^1 and consists of agents A_1 and A_2, while Crew 2 is managed by C_M^2 and comprises agents A_3 through A_8. Figure 1 illustrates our approach to Occupational Health (OH) Triage using multiple LLM agents.

2.4 System Overview

We constructed our triage agent-based framework following standardized triage protocols developed by expert clinicians at Healthcare Provider. Our framework implements a sequential processing pipeline beginning with LLM-based anonymization of clinical records M_i, followed by a two-stage information enrichment process: (1) extraction of medical entities and occupation-related information, and (2) comprehensive information summarization, producing condensed case representations S_i. These are directed to our dual-channel triage system managed by specialized Chat Managers C_M^1 and C_M^2.

C_M^1 coordinates Crew_1 to analyze communication difficulties and workplace assessment requirements for appointment modality decisions. Concurrently, C_M^2 orchestrates Crew_2 to evaluate specialized case characteristics for healthcare provider assignment. Specifically, Crew_2 identifies critical factors including substance abuse (A_3), job-related safety concerns (A_4), disciplinary action issues (A_5), mental health conditions (A_6), infectious diseases (A_7), and RIDDOR[4]-related cases (A_8).

Table 1: Distribution of Medical Categories in the Dataset

Category	Total Count	Percentage
Mental Health	888	34.1%
Musculoskeletal	770	29.6%
Neurological	174	6.7%
Cardiovascular	133	5.1%
Gastrointestinal	124	4.8%
Genitourinary	109	4.2%
Respiratory	91	3.5%
Oncology	83	3.2%
ENT and Sensory	68	2.6%
Infectious Disease	41	1.6%
Pregnancy	29	1.1%
Other	79	3.0%

Iterative Discussion. Our framework implements five consecutive discussion iterations among specialized agents for each case, employing majority voting to determine the final recommendation.

Decision Rules. We employ a *safety-prioritized protocol* where if any agent in Crew_1 recommends face-to-face consultation, the case defaults to an in-person appointment. Similarly, if any agent in Crew_2 suggests physician consultation, the case is assigned to a doctor rather than an alternative provider.

Our framework employs a multi-dimensional approach: distributed parallel assessment (horizontal dimension) where specialized agents concurrently evaluate distinct clinical aspects, and temporal iterative refinement (vertical dimension) consisting of five sequential deliberation cycles.

Early Stopping Mechanism. We terminate agent deliberation after three consistent decisions from an individual agent, as the majority outcome in a five-iteration sequence is determined after three identical decisions.

3 Experiments

3.1 Experiment Setups

Dataset. We conducted experiments using a comprehensive private occupational healthcare dataset from Heales Medical, comprising 2,589 clinically diverse cases. The distribution of medical categories is detailed in Table 1. Our preliminary investigation employed a transformer-based model with

[4] https://www.hse.gov.uk/riddor/

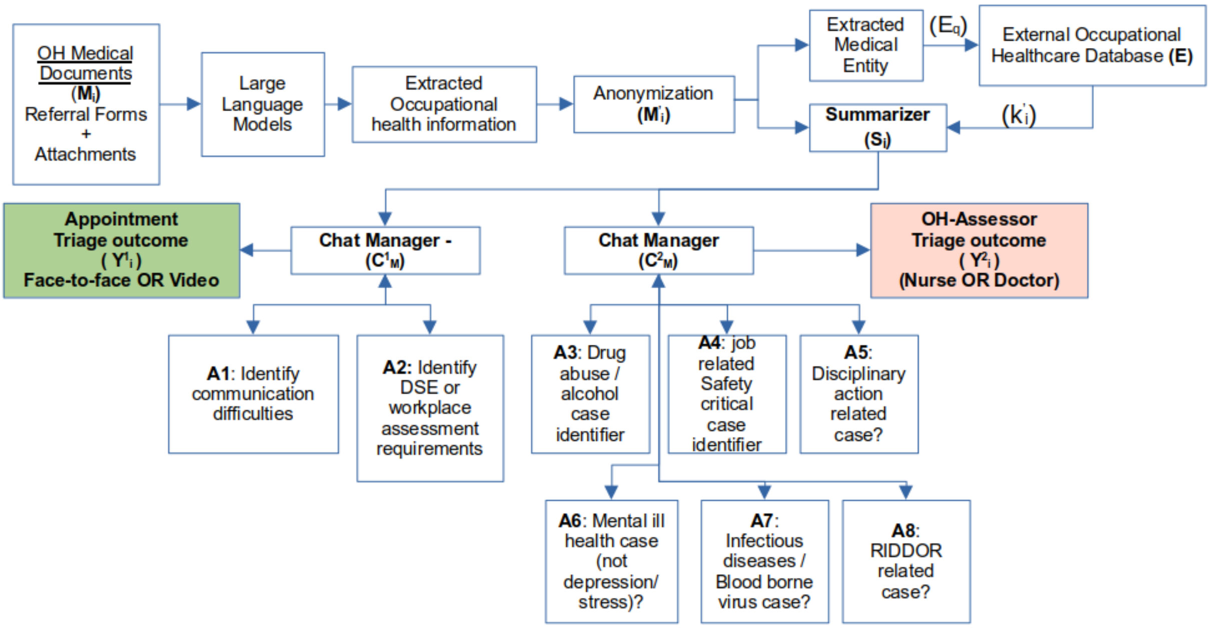

Figure 1: Overview of our **OccuTriage** orchestration framework for occupational healthcare multi triage prediction, developed in collaboration with Heales Medical . The framework integrates referral forms, medical records, and external knowledge bases, utilizes multiple specialized LLM agents to perform comprehensive analysis and generate accurate triage recommendations.

a standard data partitioning protocol, yielding moderate F1-scores of 63% for assessor type prediction and 55% for appointment modality classification. These limitations stemmed from insufficient training data volume and architectural constraints in learning from sparse, unstructured clinical information.

Implementation Details. We implement Llama3.1 8B and Llama3.2 13B vision models by (Team and Meta, 2024) deployed using Text Generation Inference engine on a Linux server with four Nvidia H100 GPUs. Llama3.2 13B was utilised to extract information from case related pdf documents. We use temperature 0.7, top_p 0.95, and repetition_penalty 1.0 for inference. Our agent framework uses Microsoft's Autogen[5] for multi-agent interactions.

Evaluation Metrics. Following (Lu et al.), we evaluate performance using discordance rate as our primary metric, supplemented by under-triage and over-triage rates (Table 2). Under-triage occurs when patients receive insufficient care, creating potential safety risks. Over-triage represents resource inefficiency through unnecessary allocation of higher care levels. While total discordance measures overall triage accuracy, under-triage poses the greater clinical risk.

[5]https://microsoft.github.io/autogen/

Table 2: Triage discordance metrics.

Term	**Definition**	**Formula**
Undertriage	Lower level of care than clinically needed	$\frac{\text{Undertriage cases}}{\text{Total cases}} \times 100\%$
Overtriage	Higher level of care than clinically needed	$\frac{\text{Overtriage cases}}{\text{Total cases}} \times 100\%$
Discordance	Total incorrect triage decisions	$\frac{\text{Under + Over}}{\text{Total cases}} \times 100\%$

Baselines. We compared our proposed OccuTriage framework against several baseline configurations: a single LLM agent without enhancements, progressively adding Chain of Thought (CoT) reasoning and Retrieval-Augmented Generation (RAG).

3.2 Main Experimental Results

Table 3 presents a comprehensive comparison of our OccuTriage framework against baseline configurations and human expert performance.

The single-agent LLM baseline without enhancements demonstrates substantial discordance rates, with Llama and Asclepius models achieving average discordance rates of 45.38% and 43.05% re-

Table 3: Performance comparison of different experimental configurations for occupational health triage prediction using Llama3.1 and Asclepius LLM models. Results show discordance metrics (%) for both appointment type and OH assessor type prediction tasks. Lower values indicate better performance.

Configuration	Model	Appointment Type			OH Assessor Type			Average
		Under	Over	Disc.	Under	Over	Disc.	Disc.
1-Agent LLM (No RAG, few shot or CoT)	Llama	22.54	26.21	48.75	7.0	35.0	42.0	45.38
	Asclepius	19.82	27.18	47.00	6.8	32.3	39.1	43.05
1-Agent LLM + RAG	Llama	18.65	25.10	43.75	5.9	30.2	36.1	39.93
	Asclepius	15.40	22.10	37.50	6.1	26.4	32.5	35.00
1-Agent LLM + Few-shot (3)	Llama	16.32	27.43	43.75	8.5	31.0	39.5	41.63
	Asclepius	16.95	24.05	41.00	7.2	32.8	40.0	40.50
1-Agent LLM + CoT	Llama	14.85	19.65	34.50	5.3	28.7	34.0	34.25
	Asclepius	14.25	20.75	35.00	5.7	24.8	30.5	32.75
OccuTriage (our framework)	Llama	**9.52**	15.91	25.43	**2.9**	**14.2**	**17.1**	21.27
	Asclepius	9.84	**12.48**	**22.32**	3.1	14.9	18.0	**20.16**
Human Expert		11.84	14.38	26.22	9.0	15.0	24.0	25.11

spectively. This indicates that unaugmented LLMs struggle with the complex decision-making required for occupational health triage.

When incorporating retrieval augmentation (RAG), performance improves significantly, reducing average discordance to 39.93% (Llama) and 35.00% (Asclepius). This improvement highlights the importance of domain-specific knowledge integration.

Few-shot learning (3 examples) yields modest improvements over the baseline, with average discordance rates of 41.63% (Llama) and 40.50% (Asclepius). Chain of Thought (CoT) reasoning demonstrates substantial performance gains, reducing average discordance to 34.25% (Llama) and 32.75% (Asclepius).

Our proposed OccuTriage framework significantly outperforms all baseline configurations, achieving an average discordance rate of 21.27% with Llama and 20.16% with Asclepius. Notably, OccuTriage exceeds human expert performance (25.11% average discordance).

The most clinically significant finding relates to under-triage rates, where OccuTriage achieves 9.52% (Llama) and 9.84% (Asclepius) for appointment type decisions, and 2.9% (Llama) and 3.1% (Asclepius) for assessor type decisions. These results are particularly important as under-triage represents potential safety risks.

When analyzed by triage decision type, assessor type prediction demonstrates consistently lower discordance rates than appointment type prediction across all configurations. This superior performance can be attributed to our comprehensive six-agent architecture in Crew 2, which effectively captures the multifaceted clinical factors influencing provider selection.

The consistent performance advantage of Asclepius over Llama3.1 across most configurations confirms the value of domain-specific model training as established by (Kweon et al., 2023).

4 Case Study

We evaluated OccuTriage on 2,589 occupational health cases from Heales Medical, comparing its performance against single-agent LLM baselines and human experts. The framework demonstrated superior triage accuracy across all metrics, achieving an average discordance rate of 20.16% with the Asclepius model, compared to 25.11% for human experts.

The progression from baseline configurations through our multi-agent approach showed steady improvement in triage accuracy. Most significantly, OccuTriage reduced under-triage rates for assessor type prediction to 2.9% (Llama3.1) and 3.1% (Asclepius), substantially outperforming human experts' 9.0% rate.

Our safety-efficiency tradeoff analysis demonstrates OccuTriage's optimal balance between under-triage (safety risk) and over-triage (efficiency risk). Configuration progression consistently moved toward the ideal performance region, with the final framework achieving both lower under-triage and over-triage rates than human experts.

Statistical analysis revealed that OccuTriage per-forms better on assessor type prediction than appointment type prediction across all configurations. The framework achieved discordance rates of 22.32% and 18.0% for appointment and assessor type predictions respectively using Asclepius, compared to 26.22% and 24.0% for human experts.

While domain-specific Asclepius models generally outperformed Llama3.1, the performance gap varied across configurations. The most substantial improvement occurred with RAG integration (4.93% average discordance reduction), suggesting domain-specific models significantly enhance knowledge-intensive operations.

Clinician feedback confirms that OccuTriage's improved accuracy justifies its modest computational overhead, particularly as reduced under-triage directly impacts patient safety while decreased over-triage optimizes resource allocation. These findings demonstrate OccuTriage's potential for improving occupational health triage through its specialized agent architecture and safety-prioritized decision protocols.

5 Conclusion

This paper presents OccuTriage, a novel AI agent orchestration framework for occupational health triage prediction. Our approach employs specialized LLM agents, retrieval augmentation, and a bidirectional decision architecture to simulate clinical reasoning. Experimental evaluation on 2,589 occupational health cases demonstrates that OccuTriage outperforms single-agent approaches with a 20.16% average discordance rate compared to baseline rates of 43.05%, while matching or exceeding human expert performance (25.11%).

The most significant finding is OccuTriage's ability to reduce under-triage rates to 9.84% and 3.1% for appointment and assessor type decisions respectively, substantially outperforming human experts (11.84% and 9.0%). This improvement is critical for patient safety, as under-triage represents inadequate care allocation.

Our multi-agent architecture demonstrates particular efficacy in assessor type prediction, with each agent focusing on distinct clinical domains—substance abuse, safety concerns, disciplinary issues, mental health, infectious diseases, and RIDDOR-related cases. This specialized focus enables robust consensus formation and precise decision-making, establishing OccuTriage as an effective tool for complex healthcare triage tasks.

The framework's safety-prioritized protocol ensures that high-risk cases default to face-to-face consultations and physician evaluations, aligning with clinical safety practices. The early stopping mechanism optimizes computational efficiency without compromising decision integrity.

In comparison with existing approaches, OccuTriage addresses the limitations identified in previous work by bridging clinical judgment and algorithmic approaches, incorporating domain-specific knowledge, and implementing a multi-dimensional decision framework specifically designed for occupational health settings.

These results establish OccuTriage's efficacy in performing complex occupational health triage while maintaining safety and optimizing resource allocation, with potential applications across diverse healthcare settings.

6 Extended Analysis and System Evaluation

6.1 Error Analysis and Performance Patterns

Analysis of the remaining 20.16% discordance cases reveals specific patterns that inform system optimization strategies. The residual discordance cases primarily cluster around complex multi-comorbidity scenarios where manual clinical judgment traditionally varies among practitioners. The specialized Mental Health Agent (A_6) systematically applies consistent diagnostic criteria across cases, with musculoskeletal cases (29.6% of dataset) showing improved consistency through structured decision protocols. Category-specific analysis reveals no systematic classification failures in any diagnostic domain.

Complex cases involving rare medical conditions or non-standard terminology usage in referral documentation present ongoing challenges that contribute to remaining discordance cases. Knowledge base retrieval with NCI Thesaurus and O*NET integration enables nuanced interpretation of medical terminology and occupational context, though these edge cases highlight areas for knowledge base expansion.

6.2 Computational Architecture Analysis

Model-specific analysis reveals distinct output formatting characteristics that impact system integration. Asclepius consistently generates responses in paragraph format with reasoning rather than structured decision outputs, necessitating additional processing overhead through a secondary Llama-based sentiment analysis layer to extract binary triage decisions. This architectural requirement contrasts with Llama models that directly produce structured classifications without requiring post-processing.

The sentiment analysis overhead adds processing complexity to Asclepius-based implementations, requiring additional model invocations per case to convert paragraph-format clinical reasoning into structured binary classifications. Despite this computational trade-off, the clinical accuracy benefits of the domain-specialized Asclepius model justify the additional processing requirements.

Runtime performance metrics demonstrate practical efficiency for clinical deployment. Processing time per case averages approximately 12 seconds, representing acceptable computational overhead for non-emergency occupational health triage. The early stopping mechanism optimizes efficiency by terminating agent discussions after achieving consensus, while the dual-crew architecture enables concurrent evaluation, maximizing resource utilization through parallel processing.

6.3 Clinical Workflow Integration

The framework demonstrates robust integration capabilities with existing healthcare information systems. Structured JSON-formatted outputs maintain compatibility with Electronic Health Record systems, while comprehensive audit trails preserve complete decision reasoning for clinical governance compliance. The system successfully processes typical occupational health referral volumes without performance degradation.

Clinical workflow compatibility extends to professional oversight capabilities, with complete reasoning chains available for practitioner review and quality assurance processes. The safety-prioritized protocol preserves clinical discretion, allowing healthcare providers to override system recommendations when clinical judgment necessitates alternative decisions.

6.4 Multi-Agent Discussion Protocol Effectiveness

Multi-agent discussion protocols prove essential for complex case resolution, with iterative consensus mechanisms resolving borderline cases that challenge single-agent approaches. The six-agent architecture in Crew_2 demonstrates particular effectiveness for assessor type predictions, achieving 18.0% discordance compared to human expert performance of 24.0%. Analysis reveals that simple cases maintain high accuracy matching human expert performance, while complex multi-comorbidity cases represent the primary source of remaining discordance, where the framework's structured approach provides more consistent results than traditional manual assessment methods.

Processing efficiency considerations support integration into existing healthcare information systems, where occupational health decisions occur within consultation scheduling timeframes rather than emergency response requirements. The computational overhead remains justified by the substantial accuracy improvements demonstrated across all experimental configurations.

Ethical Considerations

This research was conducted with a strong commitment to ethical standards and data protection regulations. All personal data collected and processed during this study adhered to the principles outlined in the General Data Protection Regulation (GDPR) of the European Union. The study utilized data from the National Cancer Institute (NCI) Thesaurus, and all usage complied with the terms specified in the NCI Thesaurus Data Use Agreement[6]. Data from ONET OnLine were incorporated into the research, following the guidelines set forth in the ONET Privacy[7] and Security Statement.

Acknowledgements

The authors would like to acknowledge the support of the Knowledge Transfer Partnership (KTP) programme, which facilitated this research through collaborative partnership between the University of Hertfordshire and Heales Medical. This work was conducted under KTP project number 12121, with partial funding provided by Innovate UK[8]. The KTP programme's objectives of facilitating knowledge transfer, enhancing technical and business skills, and strengthening industry-academia collaboration were instrumental in the successful completion of this research. We extend our gratitude to all personnel from both the academic and industrial partners who provided supervision and guidance throughout the project duration.

[6] https://evs.nci.nih.gov/ftp1/NCI_Thesaurus/archive/deprecated_terms_of_use/June2006_Aug2018_ThesaurusTermsofUse.htm

[7] https://www.onetonline.org/help/privacy/

[8] https://iuk-ktp.org.uk/

References

Janaina S. CARIBE, Lilian M.F. VITERBO, Diogo G. VIDAL, and Katia N. SA. 2020. Development and validation of a nursing instrument for triage in occupational health services. *Espacios*, 41(45):261–271.

Hao Cheng, Hao Fang, Xiaodong Liu, and Jianfeng Gao. 2023. Task-Aware Specialization for Efficient and Robust Dense Retrieval for Open-Domain Question Answering. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1864–1875, Toronto, Canada. Association for Computational Linguistics.

Marta Fernandes, Susana M. Vieira, Francisca Leite, Carlos Palos, Stan Finkelstein, and João M.C. Sousa. 2020. Clinical Decision Support Systems for Triage in the Emergency Department using Intelligent Systems: a Review.

Hugh Gorick, Marie McGee, and Toby Smith. 2024. Understanding the demographics, training experiences and decision-making practices of UK triage nurses. *Emergency Nurse*, 32(3):1–9.

C. E. Green, J. S. Leeds, and C. M. Leeds. 2024. Occupational effects in patients with post-COVID-19 syndrome. *Occupational Medicine*, 74(1):86–92.

Seungjun Han and Wongyung Choi. 2024. Development of a Large Language Model-based Multi-Agent Clinical Decision Support System for Korean Triage and Acuity Scale (KTAS)-Based Triage and Treatment Planning in Emergency Departments.

Huilin Jiang, Haifeng Mao, Huimin Lu, Peiyi Lin, Wei Garry, Huijing Lu, Guangqian Yang, Timothy H. Rainer, and Xiaohui Chen. 2021. Machine learning-based models to support decision-making in emergency department triage for patients with suspected cardiovascular disease. *International Journal of Medical Informatics*, 145.

Norman Jones and N. Greenberg. 2015. The use of Threshold Assessment Grid triage (TAG-triage) in mental health assessment. *Journal of the Royal Army Medical Corps*, 161:i46–i51.

Marvin Kopka, Hendrik Napierala, Martin Privoznik, Desislava Sapunova, Sizhuo Zhang, and Markus A. Feufel. 2024. Evaluating self-triage accuracy of laypeople, symptom-assessment apps, and large language models: A framework for case vignette development using a representative design approach (RepVig).

Evgeniy Krastev, Dimitar Tcharaktchiev, Petko Kovachev, and Simeon Abanos. 2023. Occupational health assessment summary designed for semantic interoperability. *International Journal of Medical Informatics*, 178.

Sunjun Kweon, Junu Kim, Jiyoun Kim, Sujeong Im, Eunbyeol Cho, Seongsu Bae, Jungwoo Oh, Gyubok Lee, Jong Hak Moon, Seng Chan You, Seungjin Baek, Chang Hoon Han, Yoon Bin Jung, Yohan Jo, and Edward Choi. 2023. Publicly Shareable Clinical Large Language Model Built on Synthetic Clinical Notes.

Drushca Lalloo, John Gallagher, Ewan Macdonald, and Conor McDonnell. 2021. Clinical Case Complexity in Occupational Health: Contributing Factors and a Proposed Conceptual Framework Model. *Journal of Occupational and Environmental Medicine*, 63(6):E352–E361.

Sheng-Chieh Lin, Akari Asai, Minghan Li, Barlas Oguz, Jimmy Lin, Yashar Mehdad, Wen-tau Yih, and Xilun Chen. 2023. How to Train Your DRAGON: Diverse Augmentation Towards Generalizable Dense Retrieval.

Meng Lu, Brandon Ho, Dennis Ren, and Xuan Wang. TRIAGEAGENT: Towards Better Multi-Agents Collaborations for Large Language Model-Based Clinical Triage. Technical report.

Michael Mandelstam. 2005. Occupational therapy: law and good practice. Technical report.

Serena McCluskey, A. Kim Burton, and Chris J. Main. 2006. The implementation of occupational health guidelines principles for reducing sickness absence due to musculoskeletal disorders. *Occupational Medicine*, 56(4):237–242.

Henry Mutegeki, Alvin Nahabwe, Joyce Nakatumba-Nabende, and Ggaliwango Marvin. 2023. Interpretable Machine Learning-Based Triage For Decision Support in Emergency Care. In *7th International Conference on Trends in Electronics and Informatics, ICOEI 2023 - Proceedings*, pages 983–990. Institute of Electrical and Electronics Engineers Inc.

National Center for O*NET Development. 2025. O*net online. Accessed 22 March 2025.

OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, Irwan Bello, Jake Berdine, Gabriel Bernadett-Shapiro, Christopher Berner, Lenny Bogdonoff, Oleg Boiko, Madelaine Boyd, Anna-Luisa Brakman, Greg Brockman, Tim Brooks, Miles Brundage, Kevin Button, Trevor Cai, Rosie Campbell, Andrew Cann, Brittany Carey, Chelsea Carlson, Rory Carmichael, Brooke Chan, Che Chang, Fotis Chantzis, Derek Chen, Sully Chen, Ruby Chen, Jason Chen, Mark Chen, Ben Chess, Chester Cho, Casey Chu, Hyung Won Chung, Dave Cummings, Jeremiah Currier, Yunxing Dai, Cory Decareaux, Thomas Degry, Noah Deutsch, Damien Deville, Arka Dhar, David Dohan, Steve Dowling, Sheila Dunning, Adrien Ecoffet, Atty Eleti, Tyna Eloundou, David Farhi, Liam Fedus, Niko Felix, Simón Posada Fishman, Juston Forte, Isabella Fulford, Leo Gao, Elie Georges, Christian Gibson, Vik Goel, Tarun Gogineni, Gabriel Goh, Rapha Gontijo-Lopes, Jonathan Gordon, Morgan Grafstein, Scott

Gray, Ryan Greene, Joshua Gross, Shixiang Shane Gu, Yufei Guo, Chris Hallacy, Jesse Han, Jeff Harris, Yuchen He, Mike Heaton, Johannes Heidecke, Chris Hesse, Alan Hickey, Wade Hickey, Peter Hoeschele, Brandon Houghton, Kenny Hsu, Shengli Hu, Xin Hu, Joost Huizinga, Shantanu Jain, Shawn Jain, Joanne Jang, Angela Jiang, Roger Jiang, Haozhun Jin, Denny Jin, Shino Jomoto, Billie Jonn, Heewoo Jun, Tomer Kaftan, Łukasz Kaiser, Ali Kamali, Ingmar Kanitscheider, Nitish Shirish Keskar, Tabarak Khan, Logan Kilpatrick, Jong Wook Kim, Christina Kim, Yongjik Kim, Jan Hendrik Kirchner, Jamie Kiros, Matt Knight, Daniel Kokotajlo, Łukasz Kondraciuk, Andrew Kondrich, Aris Konstantinidis, Kyle Kosic, Gretchen Krueger, Vishal Kuo, Michael Lampe, Ikai Lan, Teddy Lee, Jan Leike, Jade Leung, Daniel Levy, Chak Ming Li, Rachel Lim, Molly Lin, Stephanie Lin, Mateusz Litwin, Theresa Lopez, Ryan Lowe, Patricia Lue, Anna Makanju, Kim Malfacini, Sam Manning, Todor Markov, Yaniv Markovski, Bianca Martin, Katie Mayer, Andrew Mayne, Bob McGrew, Scott Mayer McKinney, Christine McLeavey, Paul McMillan, Jake McNeil, David Medina, Aalok Mehta, Jacob Menick, Luke Metz, Andrey Mishchenko, Pamela Mishkin, Vinnie Monaco, Evan Morikawa, Daniel Mossing, Tong Mu, Mira Murati, Oleg Murk, David Mély, Ashvin Nair, Reiichiro Nakano, Rajeev Nayak, Arvind Neelakantan, Richard Ngo, Hyeonwoo Noh, Long Ouyang, Cullen O'Keefe, Jakub Pachocki, Alex Paino, Joe Palermo, Ashley Pantuliano, Giambattista Parascandolo, Joel Parish, Emy Parparita, Alex Passos, Mikhail Pavlov, Andrew Peng, Adam Perelman, Filipe de Avila Belbute Peres, Michael Petrov, Henrique Ponde de Oliveira Pinto, Michael, Pokorny, Michelle Pokrass, Vitchyr H. Pong, Tolly Powell, Alethea Power, Boris Power, Elizabeth Proehl, Raul Puri, Alec Radford, Jack Rae, Aditya Ramesh, Cameron Raymond, Francis Real, Kendra Rimbach, Carl Ross, Bob Rotsted, Henri Roussez, Nick Ryder, Mario Saltarelli, Ted Sanders, Shibani Santurkar, Girish Sastry, Heather Schmidt, David Schnurr, John Schulman, Daniel Selsam, Kyla Sheppard, Toki Sherbakov, Jessica Shieh, Sarah Shoker, Pranav Shyam, Szymon Sidor, Eric Sigler, Maddie Simens, Jordan Sitkin, Katarina Slama, Ian Sohl, Benjamin Sokolowsky, Yang Song, Natalie Staudacher, Felipe Petroski Such, Natalie Summers, Ilya Sutskever, Jie Tang, Nikolas Tezak, Madeleine B. Thompson, Phil Tillet, Amin Tootoonchian, Elizabeth Tseng, Preston Tuggle, Nick Turley, Jerry Tworek, Juan Felipe Cerón Uribe, Andrea Vallone, Arun Vijayvergiya, Chelsea Voss, Carroll Wainwright, Justin Jay Wang, Alvin Wang, Ben Wang, Jonathan Ward, Jason Wei, CJ Weinmann, Akila Welihinda, Peter Welinder, Jiayi Weng, Lilian Weng, Matt Wiethoff, Dave Willner, Clemens Winter, Samuel Wolrich, Hannah Wong, Lauren Workman, Sherwin Wu, Jeff Wu, Michael Wu, Kai Xiao, Tao Xu, Sarah Yoo, Kevin Yu, Qiming Yuan, Wojciech Zaremba, Rowan Zellers, Chong Zhang, Marvin Zhang, Shengjia Zhao, Tianhao Zheng, Juntang Zhuang, William Zhuk, and Barret Zoph. 2023. GPT-4 Technical Report.

A O'reilly and P Carr. Telephone Consultation in the Covid-19 Era: An Occupational Health Perspective. Technical Report 5.

A. O'Reilly and C. McDonnell. 2020. Management referral triaging process pilot study: A 'telephone first' approach. *Occupational Medicine*, 70(9):656–664.

Natisha Sands, Stephen Elsom, Robert Colgate, Helen Haylor, and Roshani Prematunga. 2016. Development and interrater reliability of the UK Mental Health Triage Scale. *International journal of mental health nursing*, 25(4):330–336.

N. Sioutos, S. de Coronado, M.W. Haber, F.W. Hartel, W.L. Shaiu, and L.W. Wright. 2007. Nci thesaurus: a semantic model integrating cancer-related clinical and molecular information. *Journal of Biomedical Informatics*, 40(1):30–43.

Robin Staab, Mark Vero, Balunovi ́c Balunovi ́c, and Martin Vechev. LANGUAGE MODELS ARE ADVANCED ANONYMIZERS. Technical report.

Llama Team and Ai @ Meta. 2024. The Llama 3 Herd of Models. Technical report.

Augustin Toma, Patrick R. Lawler, Jimmy Ba, Rahul G. Krishnan, Barry B. Rubin, and Bo Wang. 2023. Clinical Camel: An Open Expert-Level Medical Language Model with Dialogue-Based Knowledge Encoding.

Tao Tu, Shekoofeh Azizi, Danny Driess, Mike Schaekermann, Mohamed Amin, Pi-Chuan Chang, Andrew Carroll, Chuck Lau, Ryutaro Tanno, Ira Ktena, Basil Mustafa, Aakanksha Chowdhery, Yun Liu, Simon Kornblith, David Fleet, Philip Mansfield, Sushant Prakash, Renee Wong, Sunny Virmani, Christopher Semturs, S Sara Mahdavi, Bradley Green, Ewa Dominowska, Blaise Aguera y Arcas, Joelle Barral, Dale Webster, Greg S. Corrado, Yossi Matias, Karan Singhal, Pete Florence, Alan Karthikesalingam, and Vivek Natarajan. 2023. Towards Generalist Biomedical AI.

L. Uronen, S. Salanterä, K. Hakala, J. Hartiala, and H. Moen. 2022. Combining supervised and unsupervised named entity recognition to detect psychosocial risk factors in occupational health checks. *International Journal of Medical Informatics*, 160.

Karen Walker-Bone, Samreen Channa, Jenny Leeser, Juliane Kause, Angela Skidmore, and Julia Smedley. 2020. Occupational health: The thin line protecting the front line.

Ting-Chia Weng, Yu-Ting Wei, Tsung-Yu Chan, Wen-Chau Chen, Ming-Hong Chen, Jung-Der Wang, Chung-I Li, and Yau-Chang Kuo. 2020. Novel Surveillance of Occupational Injury Powered by Machine Learning Using Chief Complaint at Emergency Triage.

Feng Xie, Marcus Eng Hock Ong, Johannes Nathaniel Min Hui Liew, Kenneth Boon Kiat Tan, Andrew Fu Wah Ho, Gayathri Devi Nadarajan, Lian Leng

Low, Yu Heng Kwan, Benjamin Alan Goldstein, David Bruce Matchar, Bibhas Chakraborty, and Nan Liu. 2021. Development and Assessment of an Interpretable Machine Learning Triage Tool for Estimating Mortality after Emergency Admissions. *JAMA Network Open*, 4(8).