Information Retrieval Coursework

Andreas Ioannides

Andreas.Ioannides@city.ac.uk

Introduction: Topic and Evaluation Policy

The topic of this coursework is the Ethical Challenges of Artificial Intelligence in Healthcare, focusing on issues such as bias, transparency, and concern. The evaluation policy employs the PICO framework (Population/Problem, Intervention, Comparison, Outcome) to systematically analyze these challenges. A structured approach using Boolean and Rank search strategies across various online databases and search engines was used to ensure comprehensive and relevant findings. The performance of these search strategies was assessed using precision metrics to provide a robust analysis of the ethical implications of Artificial Intelligence in Healthcare.

Title: Ethical Challenges of Artificial Intelligence in Healthcare.

Description: Identify documents that analyze and evaluate the ethical challenges posed by integrating AI systems in healthcare. It will focus on documents discussing bias, transparency, privacy, and regulatory compliance. Comparative studies between AI-based and traditional healthcare systems, as well as strategies for mitigating ethical issues, are also relevant.

Narrative: The focus of this topic is to compare traditional healthcare practices with AI-based systems and evaluate the ethical challenges associated with AI integration. Ideal documents should discuss such as bias in diagnosis systems, transparency in clinical decision support, privacy concerns, and the role of regulatory bodies. Documents that do not compare AI-based and traditional healthcare systems, or do not address ethical challenges are considered nonrelevant.

Facet Analysis

The final facet analysis which is shown below, is the result of a comprehensive iterative process of “facet development”→ queries→results evaluation→” facet development”→repeat. The facet analysis was constructed using the PICO classification, particularly suitable for healthcare topics.

Population/ Problem (P)

Facet: Stakeholders in Healthcare

Terms: Healthcare providers (doctors, nurses, technicians), Patients, Healthcare administrators, AI developers, Researchers, Regulatory bodies

Healthcare providers are faced with bias and transparency issues, patients are affected by privacy breaches and lack of consent, administrators are responsible for implementing AI systems, developers and researchers must ensure fairness and transparency, and regulatory bodies establish ethical guidelines for AI use in Healthcare.

Intervention (I)

Facet: AI Systems and Applications

Terms: Diagnosis Systems (cardiology, radiology, and dermatology), Treatment recommendation systems, Robotic surgery systems, Clinical decision support systems

Diagnosis systems have a bias from unrepresentative data, treatment systems ensure fairness and integrity, robotic surgery systems face challenges with accountability and consent, and clinical decision support systems struggle with transparency in AI decisions.

Comparison (C)

Facet: Traditional vs AI-based Healthcare Systems

Terms: Non-AI diagnostic methods vs AI diagnosis systems, Human-based recommendations vs Automated treatment recommendations, Deep-learning methods vs rule-based systems, Public data vs Private data, Different regulatory standards

Methods face the dilemma of human expertise and automated performance, recommendations face potential bias and enhanced fairness, systems face complexity and transparency differences, private and public data face privacy and security concerns, and regulatory standards face inconsistencies in addressing ethical challenges.

Outcome (O)

Facet: Impacts and Mitigation of Ethical Challenges

Terms: Improved treatment, Cost reduction, Patient safety, Transparency measures, Regulatory compliance, Mitigating privacy, Bias

Treatment deals with early disease detection and accuracy, cost reduction streamlines processes, patient safety is enhanced through transparency, regulatory compliance ensures adherence to ethical guidelines, and mitigating ethical risks reduces privacy and bias

Boolean Search Strategies

I commenced my investigation by searching on Pro Quest Dialog using a Boolean Search Strategy. The strength of Boolean search strategies lies in their ability to precisely control search results through operators such as AND NOT and OR, allowing for various combinations of keywords to assess the relevance of results. To begin with, my search produced a substantial volume of results. These results allowed me to observe the general trends and types of information available. This preliminary step was essential in identifying common themes and gaps in existing literature. For example, for broad terms such as “Artificial Intelligence”, I retrieved a diverse selection of documents that provided a comprehensive overview of the field. However, most results weren’t relevant to the specific focus of the Ethics of AI in Healthcare. On analyzing the extracted documents, I noticed a plethora of documents addressing broader topics beyond my primary interest, prompting a need for further refinement. To enhance specificity and relevance, I adjusted the search terms. I incorporated more keywords into our PICO framework, to filter out documents focused on unrelated economic and social impacts, thus concentrating more on ethical implications. My final Boolean search strategy using the building blocks method was conducted as follows:

S1: (“Healthcare Providers” OR “Patients “OR “Healthcare Administrators” OR “AI Developers” OR “Researchers” OR “Regulatory Bodies”) (P)

S2: (“Diagnosis Systems” OR “Treatment Recommendation Systems” OR “Robotic Surgery Systems” OR “Clinical Decision Support Systems”) (I)

S3: (“Non-AI Diagnostic Systems” OR “AI Diagnostic Systems” OR “Human Based Recommendations” OR “Automated Treatment Recommendations” OR “Deep-Learning Methods” OR “Rule-Based Systems” OR “Private Data” OR “Public Data” OR “Different Regulatory Standards”) (C)

S4: (“Improved Treatment” OR “Cost Reduction” OR “Patient Safety” OR “Transparency Measures” OR “Regulatory Compliance” OR “Mitigating Privacy” OR “Bias”) (O)

Building Blocks search strategy: S1 AND S2 AND S3 AND S4

The final number of documents retrieved by the building blocks strategy was 523, significantly reducing the document count. However, when reviewing the relevant documents, although some directly addressed “AI systems” impact on healthcare, others were more focused on technical or economic factors, being outside of our scope of interest. Therefore, it was necessary to refine the search strategy more, to capture clinical impacts and regulatory considerations, which were part of our facet analysis more effectively. I used a refined search strategy with exclusion criteria.

S5= (“Technical Difficulties” OR “Economic Effects” OR “Merging Systems” OR “Data Organization”)

New Building Blocks Search: S1 AND S2 AND S3 AND S4 NOT S5

This method reduced the number of overall documents from 523 to 490. Even though this is not a large decrease, the refined search strategy improved the focus on clinical benefits and regulatory issues, which are central to our study. This process highlighted that eliminating terms related to technical and economic aspects reduced quantity but enhanced overall relevance.

Considering the remaining Boolean search strategies that I had to perform, which were visual 2d searches using Google and Bing, and private searches using Start Page, the same search strategy and the same keywords were used, so findings could be consistent and reproducible. In conclusion, this search strategy aligns seamlessly with our facet analysis, addressing key stakeholders (Population) AI applications in healthcare (Intervention) comparisons with traditional methods (Comparison), and examining the impact and mitigation strategies for ethical challenges (Outcome)

Rank Search Strategies

For Rank Strategies, we know from our lectures and labs that more relevant documents occur when we keep our search terms concise. For this reason, I decided to include broad terms that potentially cover a wide range of search output. My final search query for ranking strategies was “AI Ethics in Healthcare”. This search query links well with our facet analysis since it naturally includes all stakeholders, hence the population term since they are directly involved in, or impacted by ethical considerations of AI implementation in healthcare settings. The ethics of AI affect providers, patient care and privacy, researcher studies, and regulatory guidelines. The term AI in the query covers various AI systems and applications used in healthcare. Ethical considerations such as diagnostic tools, treatment recommendations, and robotic surgeries are crucial for each type of system, making the query broadly available. The query implicitly involves

comparing traditional healthcare practices with AI-based ones by focusing on ethical dimensions, especially concerning bias privacy, and regulatory standards. “Ethics” in the query search directly applies to outcomes related to ethical implications of AI in healthcare, such as transparency, compliance, and bias mitigation, all critical for improving patient safety and treatment outcomes. In conclusion, the broad search query “AI Ethics in Healthcare” directly links with our facet analysis and is a good query to tackle rank search strategies. This is because we know engines like YouTube, Google, and Bing images reveal more relevant results when fewer search terms are used.

Evaluation

Command Line ProQuest Dialog searches using Boolean queries revealed a nuanced performance characterized by moderate precision, as indicated by the metrics P@5 at 0.40 and P@10 at 0.50, with an average precision (AveP) of 0.31. This suggests that the service can identify relevant documents within the top results. For example, those documents focus on AI applications in healthcare and ethical considerations. However, it struggles to maintain relevance consistently across all results. The topics of top documents ranged from “AI-based medical device technologies implementation strategies” to “Ethics in AI and its impact on leadership styles”, where leadership styles are irrelevant to our topic. A noticeable entry is a report by the European parliament discussing the ethical risks of AI in healthcare. This highlights the ability of the service to extract documents that are relevant and also varied across the broader search theme. On the other hand, a notable concern was the occurrence of 3 spam entries. This shows that the search engine retrieves irrelevant, and low-quality documents that detract from user experience. The Command Line ProQuest Dialog shows a competent level of performance in fetching relevant results; however, the issues of spam indicate the need for improved filtering processes to enhance precision and user trust in the search results provided.

For visual 2d Boolean searches using Google, demonstrated proficiency in navigating results. The service achieved a P@5 of 0.80 a P@10 of 0.60 and an (AveP) of 0,52 indicating a strong ability to retrieve highly relevant results at the top of the search output, but decreasing relevance as more results are considered. Searches range from academic articles to general discussions of AI. Notable entries such as “Ethical implications of AI and robotics in healthcare” outline the capability of the service to source content that is both deeply informative and broad in scope. Another example was an article from the National Institute of Health discussing clinicians' views on artificial intelligence in healthcare. The lack of spam and irrelevant links showcases Google’s robust algorithm in filtering relevant searches. This shows that Google’s 2d Visual Search is well-optimized and seeks detailed and specific information within the healthcare industry.

Visual 2d Boolean searches using Bing demonstrated a proficient but not optimal performance in retrieving relevant documents. P@5 was at 0.60, P@10 at 0,50, and average precision at 0.33 indicating that while Bing can successfully identify relevant documents within the top 5 results, relevancy decreases in the top 10. Search outcomes include articles, industry reports, and regulatory perspectives reflecting Bing's ability to display highly credible and significant information. An interesting example is a narrative from Science Direct discussing the ethical and regulatory challenges of AI in healthcare. Despite this strength, a lower AveP than Google suggests that Bing can benefit from enhanced fine-tuning to improve the relevancy and

consistency of results. Just like Google, Bing had no spam or duplicate results indicating robustness and quality outcomes.

For the ranking web search using Google, Google exhibited an almost perfect precision with P@5 of 1.00 and P@10 of 0.80 alongside an AveP of 0.78 indicating its superior ability to extract highly relevant documents at the top and bottom of the results. The search output included scholarly articles, comprehensive reviews, and industry insights, showcasing Google's ability to display relevant and reliable content. A documented example is an article about the ethical implications of AI in healthcare by research gate. Conversely, Bing displayed a lower but still respectable performance of P@5 of 0.60, P@10 of 0.70, and AveP of 0.58. This shows that Bing is effective at retrieving relevant documents, but does not match Google in precision and ranking efficiency of top-tier results. Both platforms returned minimal spam, with one instance of each, however, it is evident that Google has the advantage of utilizing complex algorithms to prioritize and present the most authoritative content about ethical issues of AI in healthcare.

Considering rank image searches using Google, it demonstrated exceptional ranking precision with P@5 of 1.00, and P@10 of 0.70, along with an average precision of 0.65. This suggests that the top results from Google Images were highly relevant and maintained a strong relevance slightly deeper into the search results. Images included well-organized and informative visuals such as diagrams, and professional slides from reputable sources like Frontiers and Springer Link, directly related to the search query. In contrast, Bing images presented a moderate performance with a P@5 of 0.60, and a P@10 of 0.60, reflecting a consistent search, but lower in precision since it had an AveP of 0.50. Results indicated educational materials and graphical representations with visuals from sources like the Medical College of Wisconsin and various health blogs. Both services showed zero spam and irrelevant documents, however, Google's ranking mechanism retrieved images that were highly relevant and deeply integrated with the ethical issues of AI in healthcare. This is because Google images were more aligned with the highly scholarly standards when compared with Bing.

For video-ranked searches, YouTube achieved a P@5 of 1.00, a P@10 of 0,80, and an AveP of 0.78. This indicates that all top 5 records closely matched the search intent. In the latter stages, results diminish but are still strong deeper into the search results. Videos included discussions, panels, and educational webinars, showcasing YouTube's effectiveness in sourcing complex content related to ethical AI and healthcare. An interesting video that popped up was about the principles for ethics in health AI from WHO. Bing also retrieved relevant documents with a P@5 of 0.80, a P@10 of 0.70, and an AveP of 0.58 indicating commendable performance, but with a drop compared to YouTube. Videos from Bing included presentations and talks from various healthcare conferences, however, doesn't consistently reach high relevancy and viewer engagement levels when compared with YouTube. A unique video from Bing was a webinar from Stanford university discussing how to bring AI into healthcare safely and ethically.

The social search social searchers service ranked search results and achieved a P@5 of 0.80 and a P@10 of 0.50 indicating initial relevant content, which slowly diminished as more results were considered. The AveP score was 0.41 further supporting a moderate consistency in relevance across the top 10 results. The source provided articles covering critical issues such as patient rights, data privacy, and implications of AI reflecting its utility for users who want to understand the ethical landscape of AI in healthcare. Even though one broken link existed, which can affect user experience, Social Searcher demonstrated strong capabilities in extracting relevant information.

Private Search StartPage for Boolean searches achieved a P@5 of 0.60, a P@10 of 0.40, and an average precision of 0.27. This suggests that initial results were relevant, but precision declined later on. The lack of broken links, spam, and irrelevant results indicates a clean but not highly precise retrieval for specific searches. In contrast, the ranking query achieved perfect scores in all metrics, showing exceptionally relevant results without any noise or irrelevant information. This level of performance makes the search engine the best search engine for users seeking highly relevant scholarly articles and authoritative content that aligns closely with user inquiries, particularly specializing in ethical considerations of AI in healthcare.

Finally, considering Chat GPT and rank searches, we have a P@5 of 0.60, a P@10 of 0,40, and an AveP of 0.20, reflecting the model's ability to initially capture relevant information about the impact of AI on stakeholders, but struggles to maintain this relevancy consistently across a broader search of results. Chat GPT discussed various perspectives on the impact of AI in Healthcare. The low AveP indicates room for improvement in refining the depth and breadth of the coverage since the 5 duplicate references provided reduce the accuracy of findings.

1	Search Service	Query Type	P@5	P@10	AveP	RT	LB	NT	SPAM
2	Command Line ProQuest Dialog	Boolean	0.40	0.50	0.31	0	0	0	3
3	Visual 2d Search Google	Boolean	0.80	0.60	0.52	0	0	0	0
4	Visual 2D Search Bing	Boolean	0.60	0.50	0.33	0	0	0	0
5	Web Search Google	Ranking	1.00	0.80	0.78	0	0	0	1
6	Web Search Bing	Ranking	0.60	0.70	0.58	0	0	0	0
7	Images Google	Ranking	1.00	0.70	0.65	0	0	0	0
8	Images Bing	Ranking	0.60	0.60	0.50	0	0	0	0
9	Video YouTube	Ranking	0.80	0.60	0.52	0	0	0	0
10	Video Bing	Ranking	0.80	0.70	0.55	0	0	0	0
11	Social Search SocialSearcher	Ranking	0.80	0.50	0.41	0	1	0	0
12	Private Search StartPage	Boolean	0.60	0.40	0.27	0	0	0	0
13	Private Search StartPage	Ranking	1.00	1.00	1.00	0	0	0	0
14	ChatGPT	Ranking	0.60	0.40	0.20	5	0	0	1

Table 1. Search results for our queries

Summary In this coursework, we initially performed the PICO facet analysis to structure an inquiry into the ethical challenges of AI in healthcare. The document outlines the diverse perspectives of stakeholders, stakeholder issues, and various facets of AI applications in diagnosis and treatment. The exploration is methodologically grounded in Boolean and Rank searches across various search engines aiming to provide a comprehensive overview of the current ethical landscapes in AI-enhanced healthcare.