Responsible Conversational AI for Equitable Maternal Healthcare

Ggaliwango Marvin
Email: ggaliwango.marvin@mak.ac.ug
Department of Computer Science,
School of Computing and Informatics Technology (SCIT),
College of Computing and Information Science (COCIS)

Maternal healthcare is a sensitive urgent public health issue that requires timely trustworthy and authentic medical responses. Unfortunately, curative healthcare systems of Low Middle-Income Countries (LMICs) are insufficiently responsive to such healthcare needs. Such needs vary among social groups, often founded on social inequalities like income, gender and education. Therefore, health information seekers turn to unregulated online healthcare platforms, social media and Large Language Modes (LLMs) which un-regulated provide unverified healthcare information.

In this work, we systematically examined the philosophical foundations of responsible data and Artificial Intelligence (AI) practices governing data and AI modelling for intelligent systems based on peer reviewed articles, book chapters, technical reports, and studies published between 1973 and 2022. These studies were restricted to the philosophy of AI and Society 5.0 to inform the derivation of over 29 forms of AI philosophies with their fundamental relationships with Society 5.0. This unveiled intrinsic manifestations of algorithmic unfairness arising from inequitable AI and Machine Learning (ML) training datasets besides irresponsible data and AI modelling practices.

We further traced this algorithmic unfairness to the unguided and unregulated AI industry practices propagated by selection of inappropriate research paradigms to inform the creation of specific AI and ML training datasets for building intelligent healthcare systems. Such systems included online platforms and chatbots designed to provide authentic timely responses to inform healthcare decision making among vulnerable online information seekers like teenagers and young women across various social groups. This pointed us to the need for responsible and Inclusive Intersectional AI practices and research approaches to creating ML and AI training datasets for equitable intelligent healthcare systems. Therefore, we inter-sectionally crowd-sourced maternal healthcare advice from over 500 verified practicing healthcare professionals form Lira University teaching hospital, Brac University and Brac Uganda's heath programme Verses their online social acquaintances with in their social networks to create a dataset based on responsible data practices.

We further implemented trustworthy medical sentiment analysis and local interpretable model agonistic explanations as responsible AI principles to distinguish between authentic and non-authentic maternal healthcare advice.

Surprisingly, we obtained a train set accuracy of 93% and a validation set accuracy of 56%, generalisation log loss of 0.259, generalisation balance accuracy of 83% and generalisation Area Under the Curve of 90%. This meant that our model performed perfectly well at everything, except for accurately distinguishing between authentic and non-authentic medical advice. This simply means that AI models cannot certainly distinguish between authentic and non-authentic medical advice. Therefore, there is a need for better online tools for conversationally disseminating authentic medical advice. We embarked on creating conversational AI techniques for leveraging conversational AI tools like ChatGPT by the information seekers through prompt engineering. These have been published and made openly available for the general public. However there is an urgent need for policy, guidelines and regulation of online healthcare practice.

Keywords: Artificial Intelligence (AI), Conversational AI, responsible AI, Large Language Models (LLMs), Maternal Health, Health Equity.

Unregulated Health Advice in Uganda's Digital Healthcare Spaces: The Power of a Social Media Doctor

Executive Summary

In Uganda, as in much of the world, the digital revolution has transformed how health information is accessed and shared. Online and social media platforms have become hotbeds for **"social media doctors"** who, despite their **lack of verifiable credentials** and **absence of online medical practice licenses,** reach millions with health advice and marketed medical products. This unchecked flow of health information poses significant risks due to the dissemination of potentially harmful misinformation and disinformation. With AI systems still unable to reliably distinguish between authentic and fraudulent health content, the situation creates an urgent need for comprehensive regulatory mechanisms to safeguard public health in the digital age.

Introduction

In the emerging era of Society 5.0, where digital and artificial intelligence (AI) technologies are becoming deeply integrated into all aspects of our lives, the potential of AI to revolutionize healthcare is immense. This is particularly true in fields such as women's health, maternal and neonatal care, and skincare, where AI has the power to deliver personalized, accessible, and timely medical advice. As the primary consumers of online health information, women, teenagers and young people in Uganda often turn to social media platforms for guidance on these critical health issues.

However, the widespread dissemination of unregulated health information poses significant risks. It is imperative that we deploy AI responsibly, ensuring that the data driving these technologies is accurate, secure, and used ethically to prevent misinformation. For policymakers, healthcare providers, and the public, this calls for a critical understanding and implementation of stringent data practices and AI regulations.

Our commitment to responsible AI in Uganda must involve clear guidelines and robust enforcement to safeguard the integrity of health information. As we navigate this digital transformation, our focus must remain on protecting and empowering all citizens, particularly vulnerable groups, with reliable, scientifically sound healthcare guidance.

This is not just a technological imperative but a moral one, ensuring that the digital advances in healthcare serve the well-being of every Ugandan.

Methodology and Results

Avoid Shallow Philosophies Underpinning Responsible Data and AI Practices.

In this work, we systematically examined the philosophical foundations of responsible data and Artificial Intelligence (AI) practices governing data and AI modelling for intelligent systems based on 80 peer reviewed articles, book chapters, technical reports, and studies published between 1973 and 2022. These studies were restricted to the philosophy of AI and Society 5.0 to inform the derivation of over 29 forms of AI philosophies with their fundamental relationships with Society 5.0. This unveiled intrinsic manifestations of algorithmic unfairness arising from inequitable AI and Machine Learning (ML) training datasets besides irresponsible data and AI modelling practices.

How are The Responsible AI Datasets Created?

We further traced this algorithmic unfairness to the unguided and unregulated AI industry practices propagated by selection of inappropriate research paradigms to inform the creation of specific AI and ML training datasets for building intelligent healthcare systems. Such systems included online platforms and chatbots designed to provide authentic timely responses to inform healthcare decision making among vulnerable online information seekers like teenagers and young women across various social groups. This pointed us to the need for responsible and Inclusive Intersectional AI practices and research approaches to creating ML and AI training datasets for equitable intelligent healthcare systems.

How to Implement Responsible Data Practices.

We intersectionally crowdsourced maternal healthcare advice from over 500 verified practicing healthcare professionals form Lira University teaching hospital, BRAC University and BRAC Uganda's heath program Verses their online social acquaintances with in their social networks to create a dataset based on responsible data practices.

How to Implement Responsible AI Practices.

We further implemented trustworthy medical sentiment analysis and local interpretable model agonistic explanations as responsible AI principles to distinguish between authentic and non-authentic maternal healthcare advice.

Limitations of AI in Digital Health.

We obtained a train set accuracy of 93% and a validation set accuracy of 56%, generalization log loss of 0.259, generalization balance accuracy of 83% and generalization Area Under the Curve of 90% meaning our model performed perfectly well at everything except for accurately distinguishing between authentic and non-authentic medical advice.

This simply means that AI models cannot certainly distinguish between authentic and non-authentic medical advice hence a need for better online tools for conversationally disseminating authentic medical advice. Therefore, we embarked on creating conversational AI techniques for leveraging conversational AI tools like ChatGPT by the information seekers through prompt engineering.

Reference Point for Regulating use of Digital and Social Media Platforms Medical Practice Doctors in Uganda

International guidelines, such as those from the UK's General Medical Council and the American Medical Association, underline the complexities of doctors' online conduct. These standards emphasize maintaining patient confidentiality, ensuring information accuracy, and upholding respect among colleagues. Notable challenges include balancing personal and professional online identities and mitigating potential harms from social media misuse.

Uganda can benefit from adopting similar comprehensive guidelines that are contextually adapted, focusing on professional self-regulation, risk management, and the promotion of online professionalism to safeguard patient interests and enhance the medical profession's integrity globally.

ATTENTION!! WARNINGS ARISING FROM DIGITAL HEALTH REGULATIONS

Privacy Risks: Overlooked or inadequate data protection measures could expose patient confidential information, violating privacy rights.

Free Speech Limitations: Regulations must balance the enforcement of professionalism without infringing on individuals' rights to free speech.

Overregulation: Excessive controls could stifle the beneficial uses of digital health platforms and social media in healthcare, such as patient support groups and professional development.

One-size-fits-all Approach: Regulations that do not consider the diversity of medical disciplines and digital platforms may be ineffective or unnecessarily restrictive.

Compliance Complexity: Overly complex regulations might be difficult for practitioners to understand and follow, potentially leading to unintended non-compliance.

Technological Lag: Regulations might quickly become outdated due to the rapid evolution of digital technologies, requiring constant updates.

Enforcement Challenges: The global nature of the internet makes enforcement of local or national regulations challenging.

Resource Allocation: Enforcing social media guidelines could require significant resources that might be diverted from other healthcare priorities.

Potential for Misinterpretation: Vague guidelines can lead to varied interpretations, resulting in inconsistent application and effectiveness.

Impact on Innovation: Strict regulations might inhibit technological and methodological innovations in healthcare delivery that could benefit patients.

Key Findings:

Pervasive Misinformation:

A significant proportion of health-related advice and product endorsements on social media by unlicensed practitioners are found to be either misleading or outright false.

Lack of Oversight: There are no adequate regulatory frameworks in place to verify and authenticate the qualifications of individuals providing health advice online, leading to a "wild west" scenario.

AI Limitations: Current AI technologies employed by social media platforms are not yet capable of effectively identifying and flagging unreliable health information, resulting in widespread dissemination of harmful content.

Recommendations:

1. **Establishment of a Digital Health Regulatory Body:** Create a dedicated agency under the Ministry of Health to monitor and regulate online health content, ensuring that all medical advice and product marketing adhere to approved standards.

2. **Licensing System for Online Health Practitioners:** Implement a mandatory licensing system for anyone wishing to offer health advice or market health products online. This system should be linked to existing medical accreditation bodies.
3. **AI Enhancement and Partnership:** Collaborate with tech companies to enhance AI's ability to detect and filter out health misinformation and disinformation. Invest in developing AI models that understand local languages and contexts.
4. **Public Awareness Campaigns:** Launch comprehensive education campaigns aimed at helping the public identify and report unverified health information online.
5. **Strict Penalties for Misinformation:** Enforce stringent penalties for individuals and organizations found spreading health misinformation, to deter such practices.

Conclusion

This policy brief has underscored critical challenges and limitations inherent in utilizing AI to discern authentic from non-authentic healthcare information on online platforms. As digital platforms proliferate, so too has the number of healthcare practitioners utilizing social media and other digital mediums to deliver services. This continuous rise highlights the urgent need for robust digital healthcare regulatory frameworks in Uganda, where currently no dedicated body exists to oversee and mitigate the risks associated with these practices. The absence of such regulation not only exposes the public to potential health misinformation but also compromises the integrity of professional medical practice online. It is crucial that stakeholders including policymakers, implementers, AI engineers, and the general public come together to establish a regulatory body equipped to handle these modern challenges. This body should enforce standards that ensure the responsible use of AI and uphold the accuracy and privacy of healthcare information, ultimately safeguarding patient care and maintaining public trust in the digital age.

References

G. Marvin, M. Tamale, B. Kanagwa, and D. Jjingo, “Philosophical review of artificial intelligence for society 5.0,” in *Algorithms for Intelligent Systems*, Singapore: Springer Nature Singapore, 2023, pp. 1–15.

G. Marvin, N. Hellen, and J. Nakatumba-Nabende, “Research paradigms for health equity in intelligent mobile healthcare technologies: A critical review,” in *Intelligent Sustainable Systems*, Singapore: Springer Nature Singapore, 2023, pp. 365–375.

G. Marvin, N. Hellen, J. Nakatumba-Nabende, and M. G. R. Alam, “Trustworthy medical sentiment detection for maternal and neonatal healthcare,” pp. 575–580, 2023.

G. Marvin, D. Jjingo, J. Nakatumba-Nabende, and M. G. R. Alam, “Local interpretable model-agnostic explanations for online maternal healthcare,” pp. 1–6, 2023.

G. Marvin, N. Hellen, D. Jjingo, and J. Nakatumba-Nabende, “Prompt Engineering in Large Language

Models," in *Data Intelligence and Cognitive Informatics*, Singapore: Springer Nature Singapore, 2024, pp. 387–402.

Contact:

Ggaliwango Marvin
School of Computing and Information Technology, Makerere University
Email: mggaliwango@gmail.com

Acknowledgements

Guide and Instructions for Using the Responsible Medical Corpus (RMC) for Maternal, Sexual, and Reproductive Health

Introduction

The Responsible Medical Corpus (RMC) dataset is an African-contextualized NLP dataset designed to improve the quality of conversations in systems related to maternal, sexual, and reproductive health (MSRH). This comprehensive guide provides detailed instructions for using the RMC dataset to perform prompt engineering, fine-tune Large Language Models (LLMs), implement Retrieval-Augmented Generation (RAG), and develop responsible software systems for MSRH. The aim is to ensure users can effectively leverage this dataset to build systems that are ethical, effective, and tailored to the African context.

1. Retrieving the Dataset

Objective: Access and prepare the RMC dataset for use.

1. **Download the Dataset:**
   - Obtain the RMC dataset from IEEE Dataport.
   - Verify the integrity of the downloaded dataset.
2. **Explore the Dataset:**
   - Examine the dataset structure, including text samples, annotations, and metadata.
   - Understand the scope and content of the dataset, focusing on its relevance to MSRH.

2. Data Cleaning and Preprocessing

Objective: Prepare the data for modeling by cleaning and preprocessing.

1. **Data Cleaning:**
   - Remove any irrelevant or duplicate entries.
   - Correct any spelling or grammatical errors in the text.
2. **Normalization:**
   - Normalize text by converting it to lowercase.
   - Remove special characters, punctuation, and stop words that do not add value to the analysis.
3. **Tokenization:**
   - Tokenize the text into words or subwords, depending on the requirements of the model.
   - Use libraries like NLTK, SpaCy, or Hugging Face's tokenizers for efficient tokenization.
4. **Lemmatization and Stemming:**
   - Apply lemmatization or stemming to reduce words to their base or root forms.
5. **Text Augmentation:**
   - Use data augmentation techniques to increase dataset variability and robustness.

- Techniques may include synonym replacement, random insertion, or back-translation.

3. Prompt Engineering

Objective: Design effective prompts to leverage the RMC dataset for various NLP tasks.

1. **Understanding Prompt Engineering:**
    - Learn the basics of prompt engineering and its importance in NLP.
    - Understand how to design prompts that elicit the desired response from language models.
2. **Creating Prompts:**
    - Design prompts that are contextually relevant to MSRH.
    - Ensure prompts are clear, concise, and aligned with the intended task (e.g., diagnosis, conversation, information retrieval).
3. **Testing and Refining Prompts:**
    - Test prompts with sample data and refine them based on the model's responses.
    - Iterate on the design to improve the quality and relevance of the outputs.

4. Fine-Tuning Large Language Models (LLMs)

Objective: Fine-tune pre-trained LLMs using the RMC dataset for specific tasks.

1. **Selecting Pre-trained Models:**
    - Choose suitable pre-trained models from Hugging Face Model Hub (e.g., BERT, GPT-3, T5).
    - Ensure the selected model aligns with the task requirements and dataset characteristics.
2. **Preparing Data for Fine-Tuning:**
    - Format the dataset to match the input requirements of the chosen model.
    - Create training and validation splits to evaluate model performance.
3. **Fine-Tuning Process:**
    - Use frameworks like Hugging Face's Transformers to fine-tune the model.
    - Set appropriate hyperparameters (e.g., learning rate, batch size, number of epochs).
    - Monitor training progress and adjust parameters as necessary.
4. **Evaluation:**
    - Evaluate the fine-tuned model on validation data using relevant metrics (e.g., accuracy, F1-score).
    - Conduct qualitative analysis by reviewing the model's outputs for specific prompts.

5. Retrieval-Augmented Generation (RAG)

Objective: Implement RAG to enhance the quality and accuracy of generated responses.

1. **Understanding RAG:**

- Learn about RAG and its benefits in combining retrieval mechanisms with generation capabilities.

2. **Setting Up Retrieval Mechanisms:**
   - Use tools like ElasticSearch or Faiss to index and retrieve relevant documents or passages from the dataset.
   - Implement retrieval strategies that ensure high recall and precision.
3. **Integrating Retrieval with Generation:**
   - Combine retrieved information with generative models to produce informed and accurate responses.
   - Ensure the retrieval component is optimized to fetch the most relevant data.
4. **Evaluation:**
   - Assess the performance of the RAG system using both automated metrics and human evaluation.
   - Focus on relevance, coherence, and informativeness of the generated responses.

6. Building and Evaluating Models

Objective: Develop models for various NLP tasks and evaluate their performance.

1. **Model Selection:**
   - Choose appropriate model architectures for the intended tasks (e.g., classification, question-answering, dialogue systems).
2. **Model Training:**
   - Train models using the prepared dataset and fine-tuned LLMs.
   - Implement techniques like transfer learning to leverage pre-trained knowledge.
3. **Evaluation:**
   - Evaluate models on test data using relevant metrics (e.g., accuracy, precision, recall, F1-score).
   - Conduct error analysis to identify areas for improvement.

7. Model Fusion and Ensembling

Objective: Improve model robustness and accuracy through ensembling techniques.

1. **Ensemble Methods:**
   - Combine predictions from multiple models using techniques like voting, averaging, or stacking.
   - Experiment with different ensemble strategies to find the best-performing combination.
2. **Evaluation:**
   - Evaluate the ensemble model's performance on the test set and compare it with individual models.

8. Machine Learning Operations (MLOps)

Objective: Streamline the development, deployment, and monitoring of models.

1. **Version Control:**
    - Use version control systems (e.g., Git) to track changes in code, data, and models.
2. **Continuous Integration and Continuous Deployment (CI/CD):**
    - Set up CI/CD pipelines to automate the training, testing, and deployment of models.
    - Use tools like Jenkins, GitLab CI, or GitHub Actions for pipeline automation.
3. **Model Monitoring:**
    - Implement monitoring tools to track model performance in real-time.
    - Set up alert systems to detect and respond to performance degradation or biases.

9. Deployment of Models

Objective: Deploy models in various Responsible Software Systems across different fields.

1. **Healthcare:**
    - Deploy models in telemedicine applications to assist healthcare professionals in diagnosing patients based on symptom text.
    - Use chatbots to provide information and support for maternal, sexual, and reproductive health.
2. **Education:**
    - Integrate models into educational platforms to provide accurate information on MSRH topics.
    - Develop conversational screening protocols for educational purposes.
3. **Public Health:**
    - Implement models in public health systems to improve the dissemination of health information.
    - Use RAG systems to support health workers in retrieving up-to-date medical information.
4. **Customer Service:**
    - Use conversational AI to provide accurate and empathetic responses in customer service applications related to health.

Conclusion

The Responsible Medical Corpus (RMC) dataset provides a unique opportunity to develop responsible software systems for maternal, sexual, and reproductive health. By following these detailed instructions, users can leverage the RMC dataset to create AI systems that are accurate, ethical, and contextually relevant. This guide aims to foster innovation in MSRH, ensuring that AI systems are developed with a focus on responsibility, inclusivity, and effectiveness.