A CHECKLIST: FRGS GRANT APPLICATION

by ROHAYANTI HASSAN

SOFTWARE ENGINEERING RESEARCH GROUP

FACULTY OF COMPUTING

UNIVERSITI TEKNOLOGI MALAYSIA

TITLE AND KEYWORDS

Assessment Criteria	Score (1-10)
Title and Keywords (5%)	9
Specific in nature reflecting fundamental issues to be resolved/novelty	
Brief and reflects the content of the proposal	

2024	Continuous Learning Code-Mixed Embeddings with Recurrent Neural Networks to Solve Code Mixing in Malay Chatbot
2023	New Malay Polysemy Word Sense Disambiguation Method based on Fuzzy Hypergraph for Malay Text Summarization
2020	A segmented-based representation of the fuzzy relations and tolerance in mixed emotion recognition
2018	A new deductive reasoning of mutant generation in metamorphic testing

2017	A New Case-Based Ensemble Integrator Method in Fuzzy Inference System
2015	A New Template-based Case-Based Reasoning in Fuzzy Logic
2015	A New Lightweight Ontology Complexity Assessment Model
2007	Symbolic data Structure and lumping Techniques Representation of large Markov Models

EXECUTIVE SUMMARY

NEW MALAY POLYSEMY WORD SENSE DISAMBIGUATION METHOD BASED ON FUZZY HYPERGRAPH FOR MALAY TEXT SUMMARIZATION

Executive Summary (10%)	9
Problem statement PS	
Objectives OBJ	
Methodology ME	
Expected output/outcome/implication OUT	
Significance of output SIG	

C(xi). Executive Summary of Research Proposal

(Please include the problem statement, objectives, research methodology, expected output/outcomes/implication, and significance of output from the research project)

Text summarization in the Malay language is a vital process that involves condensing lengthy documents into shorter versions while retaining the essential information and main points. This is particularly important because it allows readers to understand the text's content quickly without having to read the entire document. Word disambiguation, on the other hand, is the process of identifying the correct meaning of a word with multiple possible meanings in a given context. Word disambiguation is important to correctly identify the meaning of ambiguous words and phrases in the original text, in order to generate an accurate summary of a text. However, factors like lack of annotated data and contextual ambiguity make the Malay word sense disambiguation becomes challenging when dealing with the polysemy present in the Malay language. To address this challenge, we propose a new Malay polysemy word sense disambiguation method that is based on fuzzy hypergraph. Our solution involves developing a Malay word sense disambiguation repository that will incorporate a greater scope of domain-specific knowledge to enrich the Malay word context. The innovative fuzzy hypergraph will be designed later to model the word-context-disambiguation annotations and relationships. The fuzzy component in this hypergraph will identify the unannotated data by discovering patterns of co-occurrence with other words, syntactic patterns, and semantic relationships. The advantage of this fuzzy hypergraph is to improve the selection of important words in the text summary. Our vision is that the improvement of this research will lead to the development of Malay summary text applications that facilitate the academic community's love for reading Malay language materials.

PS OBJ ME OUT SIG

OBJECTIVES

NEW MALAY POLYSEMY WORD SENSE DISAMBIGUATION METHOD BASED ON FUZZY HYPERGRAPH FOR MALAY TEXT SUMMARIZATION

PROBLEM

1	The Malay language's polysemy issues pose a significant challenge for word sense disambiguation in Malay text summarization
2	Lack of Annotated Data on Malay Word Sense Disambiguation
3	The limitation of directed graphs for Malay text summarization

RESEARCH QUESTIONS

1	Why is domain-specific knowledge crucial to formulate a new Malay word sense disambiguation ontology framework?
2	How to formulate the Malay polysemy fuzzy hypergraph method and enrich it with entities from the Malay word sense disambiguation ontology framework?
3	How to evaluate the effectiveness of the Malay polysemy fuzzy hypergraph method in generating text summaries that accurately disambiguate word senses?

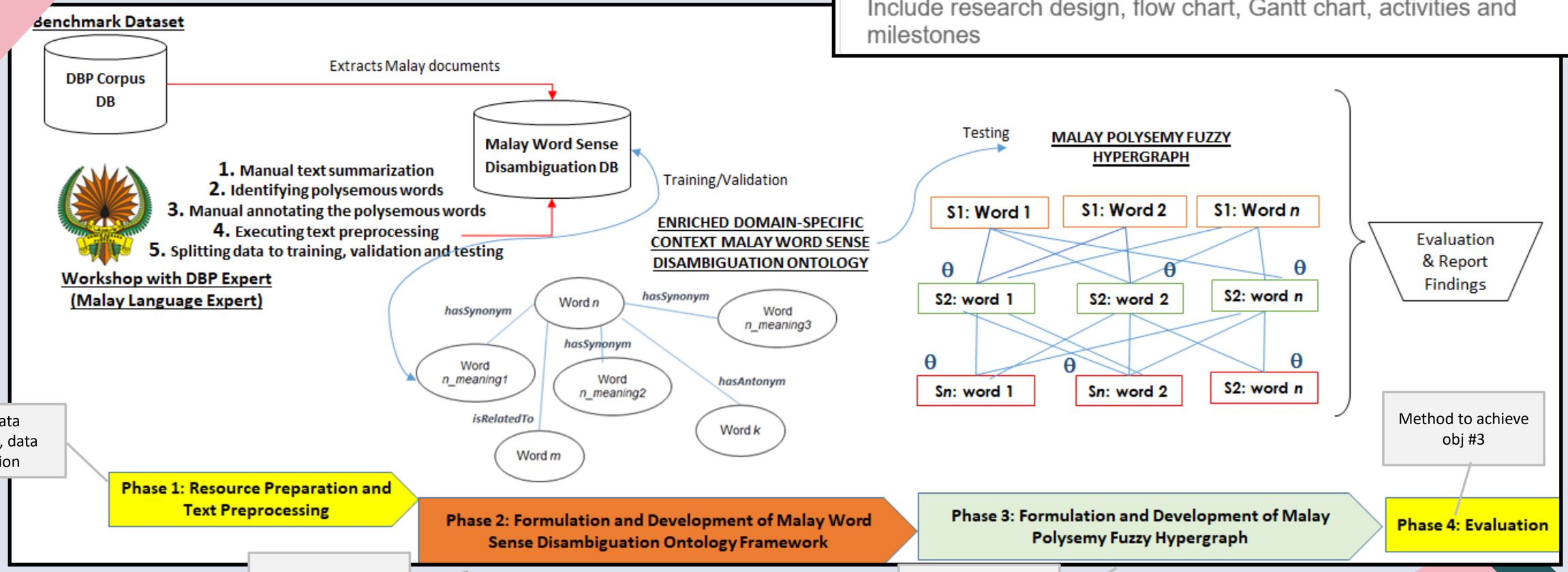
Objectives (10%)
Specific, Measurable, Achievable, Realistic and within Time- frame (SMART)
Relate to problem statement/research question

OBJECTIVES

1	To formulate a new Malay word sense disambiguation ontology framework by leveraging crucial domain-specific knowledge.
2	To formulate a modified Malay polysemy fuzzy hypergraph method using enriched entities from Malay word sense disambiguation in (1).
3	To evaluate the effectiveness of the Malay polysemy fuzzy hypergraph method in (2) in generating text summaries that accurately disambiguate word senses.

METHODOLOGY

Methodology (20%)
Clear and detailed description of methodology (may consist of field work, sampling techniques, interview session, analysis, lab work of different phases, experimental protocol, statistical analysis)
Able to achieve research objectives
Include research design, flow chart, Gantt chart, activities and milestones



Clear data collection, data extraction

Phase 1: Resource Preparation and Text Preprocessing

Phase 2: Formulation and Development of Malay Word Sense Disambiguation Ontology Framework

Phase 3: Formulation and Development of Malay Polysemy Fuzzy Hypergraph

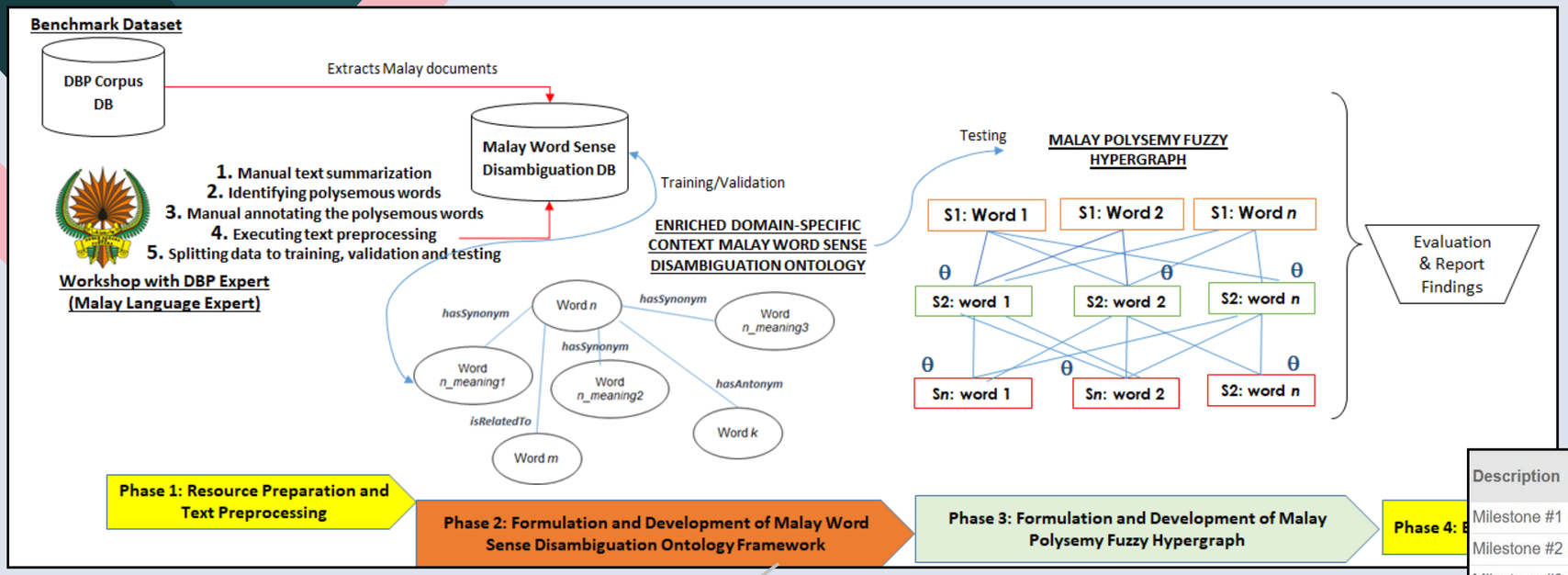
Phase 4: Evaluation

Method to achieve obj #3

Method to achieve obj #1

Method to achieve obj #2

METHODOLOGY



Phases in methodology workflow tally with the list of activities

Activity	Start Date	End Date
Phase 1: Resource Preparation	01/10/2023	29/02/2024
Phase 2: Formulation and Development of Malay Word Sense Disambiguation Ontology Framework and Text Pre-processing	01/03/2024	30/06/2024
Phase 3: Formulation and Development of Malay Polysemy Fuzzy Hypergraph Method	01/07/2024	28/02/2025
Phase 4: Evaluation and Report Findings	01/03/2025	30/09/2025
Report Writing	01/09/2025	30/09/2025

Description	Date	Cumu Perce
Milestone #1 (Phase 1): Technical Analysis report of Malay Word Sense Disambiguation dataset	29/02/2024	20
Milestone #2 (Phase 2): Malay Word Sense Disambiguation Repository	30/04/2024	40
Milestone #3 (Phase 2): Publication #1, To be submitted in International Conference 2024 Local/International, Indexed Journal	30/06/2024	55
Milestone #4 (Phase 3): Malay Polysemy fuzzy hypergraph method	30/11/2024	75
Milestone #5 (Phase 3): Publication #2, To be submitted in International Conference 2025 Local/International, Indexed Journal	28/02/2025	85
Milestone #6 (Phase 4): Publication #3, To be submitted in Q PLOS ONE Journal 2025	31/05/2025	95
Milestone #7: Project Closing Report	30/09/2025	100

EXPECTED RESULTS

Expected Results (5%)	10
New theory or new findings/knowledge **FINDINGS**	
Publication in indexed journals (top tier)/Intellectual property **PUBLICATION**	
Talents - Masters or PhD **TALENT**	

3. Research Publications (Each proposal must produce at least two (2) papers in indexed journals, one of which should be in Web of Science (WoS))

Indexing Body	Indexed Journal	
	Number of Publication	Name of Journal
WoS	1	PLOS ONE
SCOPUS	2	Potential to be published in: 1. International Journal on Informatics Visualization 2. Indexed journal via conference 4th International Conference on Social Science, Teaching and Learning (ICSSTL-2025)
ERA		
MyCITE		
	Total 3	

STATE TARGETED PUBLICATION

4. Specific or Potential Applications of the Research Findings **FINDINGS**

The outcome of this study contributes to a better automatic text summarization of Malay documents that can be potentially commercialized as a tool in education.

Total Number of Applications: 1

5. Number of PhD and Masters (by research) Students

Total Number of PhD (by research) Student(s):

Total Number of Masters (by research) Student(s):
1

Remark (if any):
Research that focus on advanced knowledge in the Malay word sense disambiguation

NO OF TALENT WITH POTENTIAL TOPIC

6. Intellectual Properties (IPs)

Malay Polysemy Fuzzy Hypergraph Method

Total Number of IP: 1

TRACK RECORD

Track Record of Project Leader and Composition of Team (5%)
Evidence of previous successful research projects
Qualification and rank of researchers
Well balanced team

INCLUDE MEMBERS WHO CAN BENEFIT YOU:

- ✓ **Expert in the chosen topic/title**
- ✓ **Experience in the chosen topic/title (has previous grant/publication)**

OTHERS

Translational Potential (10%)
Impact on Quintuple Helix (Society, Academia, Industry, Government and Environment)

Robustness of Proposal (10%)
Risk Assessment
Meticulous and proper use of language (grammar, spelling, sentence construction)
Good formatting and presentation

Elements of FRGS Criteria (5%)
Novel, cutting edge, high impact

2. Impact Statement on Quintuple Helix (please delineate/describe expected research deliverables on Society, Academia, Government, Industry and Environment)

By utilizing the new advanced knowledge, we envision that the quality of Malay text summarization tools can be significantly improved, that can benefit:

Academia – benefits academia by reducing the time and effort required to read and comprehend large volumes of information, allowing them to focus on the most critical aspects of the text.

Society – help the young generation to improve their proficiency in Malay language.

Government – contribute to MyDigital policy, developing digital talent and innovation in Malaysia.

Industry – In industries where customer service is crucial, such as hospitality or retail, a text summarization tool can help customer service representatives quickly understand customer complaints or inquiries and respond in a timely and efficient manner.

Environment - By condensing lengthy documents into shorter summaries, a text summarization tool can reduce the amount of paper needed to print out these documents. This can help reduce deforestation and save trees.

EXECUTIVE SUMMARY

CONTINUOUS LEARNING CODE-MIXED EMBEDDINGS WITH RECURRENT NEURAL NETWORKS TO SOLVE CODE MIXING IN MALAY CHATBOT

Executive Summary (10%)	9
Problem statement	
Objectives	
Methodology	
Expected output/outcome/implication	
Significance of output	

D(i). Executive Summary of Research Proposal

(Please include the problem statement, objectives, research methodology, expected output/outcomes/implication, and significance of output from the research project)

In the academic context, chatbots are pivotal tools for fostering dynamic learning environments, yet the domain of Malay chatbots lacks thorough research. Code-mixing poses a significant hurdle, constraining their efficiency as Standard Recurrent Neural Networks (RNNs) struggle to interpret such phrases accurately, hindering responses to prevalent multilingual queries. This study proposes a comprehensive approach centered on augmenting Natural Language Understanding (NLU) in Malay chatbots. By employing continuous evaluation and improvement strategies for NLU components and developing specialized Code-Mixed Embeddings tailored for English-Malay language pairs, this research aims to address these limitations. The continuous evaluation method ensures iterative enhancement of NLU components through feedback integration and retraining with updated data, adapting chatbot language comprehension to real-world code-mixed conversations. Simultaneously, the focus on Code-Mixed Embedding aims to capture nuanced nuances and contextual relevance of both languages, empowering chatbots to interpret code-mixed inputs more accurately. Ultimately, leveraging these strategies within the NLU framework aligns with the intention to enhance chatbot capabilities to respond effectively to diverse code-mixed queries in academic settings. The envisioned solution aligns with the United Nations' Sustainable Development Goal 4 (Quality Education) by facilitating accessible and inclusive academic resources for students from diverse linguistic backgrounds.

OBJECTIVES

CONTINUOUS LEARNING CODE-MIXED EMBEDDINGS WITH RECURRENT NEURAL NETWORKS TO SOLVE CODE MIXING IN MALAY CHATBOT

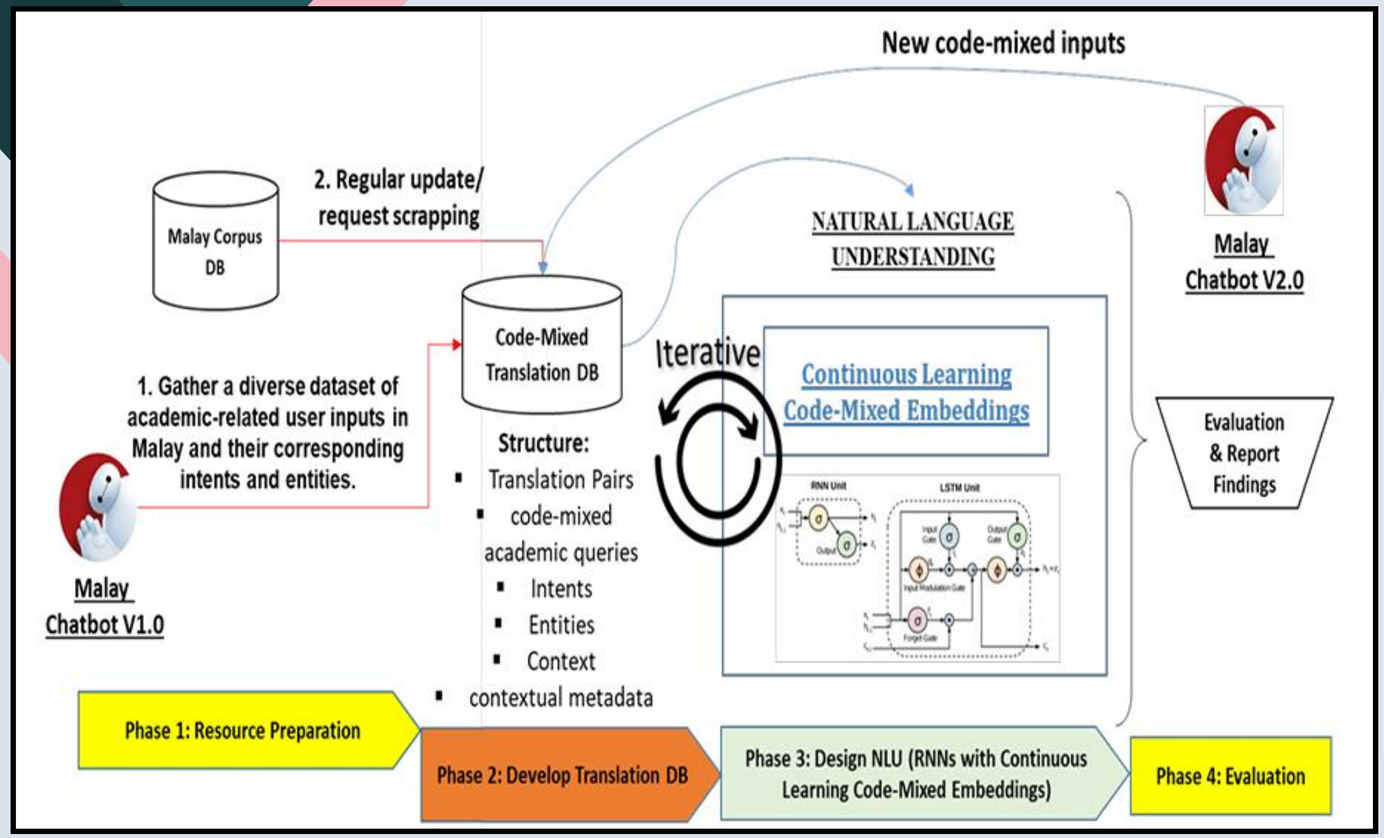
PROBLEM

1	Challenges in Malay Chatbot
2	Limitation when using standard Recurrent Neural Network
3	Limitation of code-mixed phrases might be considered out-of-vocabulary

OBJECTIVES

1	To formulate the Malay chatbot's Natural Language Understanding technique by integrating the Recurrent Neural Networks and specialized Code-Mixed Embeddings for more accurate context, intents, and entity recognition in code-mixing Malay chatbots.
2	To formulate a Continuous Learning Code-Mixed Embeddings within the Malay chatbot Recurrent Neural Networks by learning from new code-mixed inputs, translation corrections, and emerging language variations to adapt and improve its language handling over time.
3	To evaluate the effectiveness of Malay chatbot Recurrent Neural Networks with Continuous Learning Code-Mixed Embeddings using translation accuracy, intent recognition accuracy, contextual accuracy, and adaptability of continuous learning.

METHODOLOGY



Description	Date	Cumul Percen
Publication #1, to be submitted in non-indexed journal (IJIC), Unstructured code-mixed Malay chats	31/05/2025	30
Malay Translation Database	31/10/2025	60
Publication #2, to be submitted in WOS Journal, Continuous Example-based Machine Translation in RNNs	31/10/2025	70
Natural Language Understanding of Malay Chatbot	28/02/2026	80
Publication #3, to be submitted in International Conference, Natural Language Understanding of Malay Chatbot	30/04/2026	90
Project Closing Report	31/07/2026	100

Activity	Start Date	End Date
Phase 1: Resource Preparation	01/08/2024	31/12/2024
Phase 2: Develop a Code-Mixed Translation Database	01/01/2025	31/05/2025
Phase 3: Formulate Learning Code-Mixed Embeddings with Recurrent Neural Networks	01/06/2025	31/10/2025
Phase 3: Develop Natural Language Understanding	01/11/2025	31/01/2026
Phase 4: Evaluation and Report Finding	01/02/2026	30/06/2026
Report Writing	01/07/2026	31/07/2026

THANK YOU

Good luck!