

# **Senior Design: Milestones, Timeline, and Matrix Effort**

## **1) Milestones**

- a) Research different Decomilation LLM models methodologies.
- b) Generate source-pair dataset with malware samples
- c) Finetuned pretrained LLM
- d) Demo Ghidra plugin
- e) Integrating finetuned LLM with Ghidra
- f) Demo MalAI

## **2) Timeline**

<b>Task</b>	<b>Start and Completion Dates</b>
<b>Research LLM models that could be utilized as an infrastructure and a starting point of the LLM models to be built.</b>	<b>10/9/2025 - 29/9/2025</b>
<b>Research possible decompilation methods.</b>	<b>22/10/2025 - 1/11/2025</b>
<b>Research how to utilize Ghidra as a plug-in.</b>	<b>1/10/2025 - 15/10/2025</b>
<b>Research and find datasets that will be utilized to train the LLM models.</b>	<b>8/10/2025 - 1/11/2025</b>

<b>Build a Ghidra plugin</b>	<b>8/10/2025 - 29/10/2025</b>
<b>Explored different variations of LLM4Decompile (Original Model, SK2)</b>	<b>5/11/2025 - 12/11/2025</b>
<b>Introduce malware samples into existing binary-source datasets (exebench, decompilebench, idioms</b>	<b>12/11/2025 - 3/12/2025</b>
<b>Explore possible tools to build datasets (Pairs of C/C++ and Binary) like CodableLLM</b>	<b>12/11/2025 - 3/12/2025</b>
<b>Integrate LLM finetuned model</b>	<b>7/1/2026 - 28/1/2026</b>
<b>Train LLM model to analyze executable files and decompile them into C-Code.</b>	<b>10/12/2025 - 21/1/2026</b>

<b>Test the models analysis of C code and Assembly code.</b>	<b>17/12/2025 - 21/1/2026</b>
<b>Document the analyzation performance of the built models in comparison to other recent models that target the same criteria.</b>	<b>21/1/2026 - 28/1/2026</b>
<b>Develop the analyzation page of the application that showcases the analyzation of both C and Assembly codes.</b>	<b>28/1/2026 - 18/2/2026</b>
<b>Develop the malware detection result page</b>	<b>25/2/2026 - 11/3/2026</b>
<b>Demo MalAI</b>	<b>11/3/2026 - 25/3/2026</b>
<b>Deploy Application (If feasible)</b>	<b>1/3/2026 - 5/3/2026</b>
<b>Develop a landing webpage for MalAI</b>	<b>25/3/2026 - 30/3/2026</b>

### 3) Effort Matrix

Task	Start and Completion Dates	Long Nguyen (Hours of Effort)	Luqman Al Hasni (Hours of Effort)
<b>Research LLM models that could be utilized as an infrastructure and a starting point of the LLM models to be built.</b>	<b>10/9/2025 - 29/9/2025</b>	<b>Primary 3 Hours</b>	<b>Secondary 2 Hours</b>
<b>Research possible decompilation methods.</b>	<b>22/10/2025 - 1/11/2025</b>	<b>Primary 4 Hours</b>	<b>Secondary 3 Hours</b>
<b>Research and find datasets that will be utilized to train the LLM models.</b>	<b>8/10/2025 - 1/11/2025</b>	<b>Primary 4 Hours</b>	<b>Secondary 2 Hours</b>

<b>Research how to utilize Ghidra as a plug-in.</b>	<b>1/10/2025 - 15/1/2025</b>	<b>Secondary 2 Hours</b>	<b>Primary 4 Hours</b>
<b>Build a Ghidra plugin</b>	<b>8/10/2025 - 29/10/2025</b>	<b>Secondary 3 Hours</b>	<b>Secondary 5 Hours</b>
<b>Explored different variations of LLM4Decompiler (Original Model, SK2)</b>	<b>5/11/2025 - 12/11/2025</b>	<b>Primary 5 Hours</b>	<b>Secondary 4 Hours</b>
<b>Introduce malware samples into existing binary-source datasets (exebench, decompilebench, idioms</b>	<b>12/11/2025 - 3/12/2025</b>	<b>Primary 5 Hours</b>	<b>Secondary 3</b>

<b>Explore possible tools to build datasets (Pairs of C/C++ and Binary) like CodableLLM</b>	<b>12/11/2025 - 3/12/2025</b>	<b>Secondary 2 Hours</b>	<b>Primary 5 Hours</b>
<b>Integrate LLM fintuned model</b>	<b>7/1/2026 - 28/1/2026</b>	<b>Primary 6 Hours</b>	<b>Secondary 4 Hours</b>
<b>Train LLM model to analyze executable files and decompile them into C-Code.</b>	<b>10/12/2025 - 21/1/2026</b>	<b>Secondary 2 Hours</b>	<b>Primary 6 Hours</b>
<b>Test the models analysis of C code and</b>	<b>17/12/2025 - 21/1/2026</b>	<b>Primary 5 Hours</b>	<b>Secondary 2 Hours</b>

<b>Assembly code.</b>			
<b>Document the analyzation performance of the built models in comparison to other recent models that target the same criteria.</b>	<b>21/1/2026 - 28/1/2026</b>	<b>Secondary 2 Hour</b>	<b>Primary 5 Hours</b>
<b>Develop the analyzation page of the application that showcases the analyzation of both C and Assembly codes.</b>	<b>28/1/2026 - 18/2/2026</b>	<b>Secondary 2 Hours</b>	<b>Primary 3 Hours</b>

<b>Develop the malware detection result page of the MalAI</b>	<b>25/2/2026 - 11/3/2026</b>	<b>Secondary 4 Hours</b>	<b>Primary 6 Hours</b>
<b>Demo MalAI</b>	<b>11/3/2026 - 25/3/2026</b>	<b>Primary 4 Hours</b>	<b>Secondary 3 Hours</b>
<b>Deploy Application (If feasible)</b>	<b>1/3/2026 - 5/3/2026</b>	<b>Secondary 4 Hours</b>	<b>Primary 5 Hours</b>
<b>Develop a landing webpage for MalAI</b>	<b>25/3/2026 - 30/3/2026</b>	<b>Primary 5 Hours</b>	<b>Secondary 4 Hours</b>

**According to the above Estimations the hours for each member are as below:**

- Luqman: 66 Hours
- Long: 62 Hours