

---

# MMMModal - Multi-Images Multi-Audio Multi-turn Multi-Modal

---

Aisyah Razak<sup>\*</sup>

Husein Zolkepli<sup>†</sup>

Kamarul Adha<sup>‡</sup>

Ariff Nazhan<sup>§</sup>

## 1 Acknowledgement

Special thanks to Malaysia-AI volunteers especially [Wan Adzhar Faiq Adzlan](#), [Ammar Azman](#), [M. Amzar](#), [Muhammad Farhan](#) and [Syafie Nizam](#) for contributing dataset to train MaLLaM.

We would like to express our gratitude to NVIDIA Inception for generously providing us with the opportunity to train our model on the Azure cloud. Their support has played a crucial role in the success of our research, enabling us to leverage advanced technologies and computational resources.

We extend our thanks to the wider research community for their valuable insights and collaborative discussions, which have greatly influenced our work. This paper reflects the collective efforts and contributions from both NVIDIA Inception and the broader research community.

## References

---

<sup>\*</sup>aisyahrazak171@gmail.com

<sup>†</sup>husein@mesolitica.com

<sup>‡</sup>kamarul.adha360@gmail.com

<sup>§</sup>ariffnzhn@gmail.com