

Explainable Framework for Interpreting Multimodal Transformer-based Neural Networks

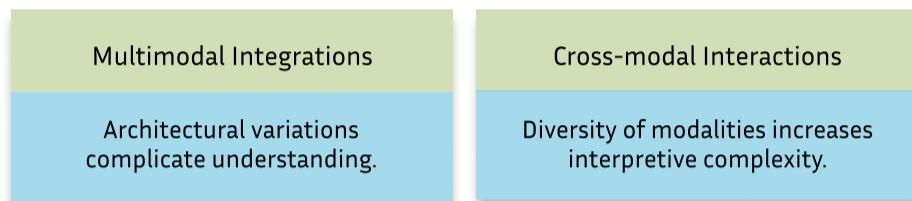
Md Raisul Kibria, Supervisor: Sebastien Lafond

Faculty of Science and Engineering, Information Technology, Åbo Akademi University



PROBLEM

- Transformers excel in scalability and multimodal modeling, driving their widespread adoption in multimodal applications.
- Lack of robust frameworks that provide comprehensive interpretations across diverse multimodal systems.
- Key challenges in explaining these models include:



OBJECTIVE

1 Extensible and interpretable framework

for analyzing various transformer-based multimodal integration techniques.

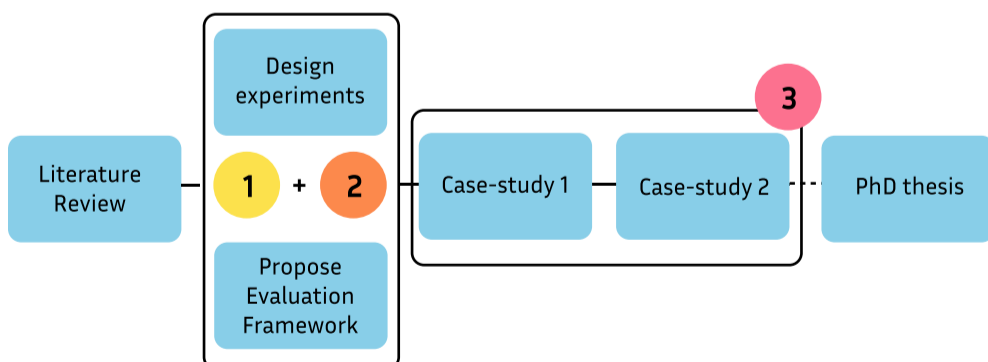
2 Evaluation Benchmark

for cross-algorithm explainable AI (XAI) assessment.

3 Validation of the framework

across multiple multimodal domains to provide more accessible results for all stakeholders (users and developers).

ROADMAP



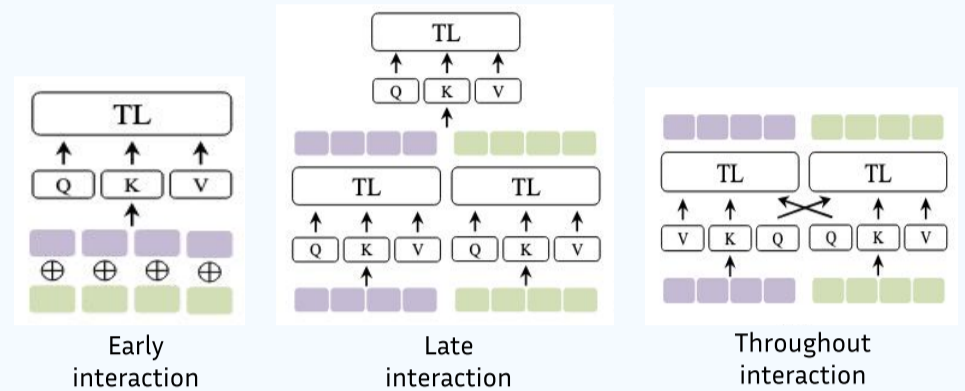
REFERENCES

- Xu, Peng, et al. "Multimodal learning with transformers: A survey." IEEE Transactions on Pattern Analysis and Machine Intelligence 45, no. 10 (2023): 12113-12132.
- Ben Abacha, Asma, et al. "Vqa-med: Overview of the medical visual question answering task at imageclef 2019." In Proceedings of CLEF (Conference and Labs of the Evaluation Forum) 2019 Working Notes. 9-12 September 2019, 2019.
- Farahnakian, Fahimeh, et al. "Deep learning based multi-modal fusion architectures for maritime vessel detection." Remote Sensing 12, no. 16 (2020): 2509.

EXPECTED RESULTS

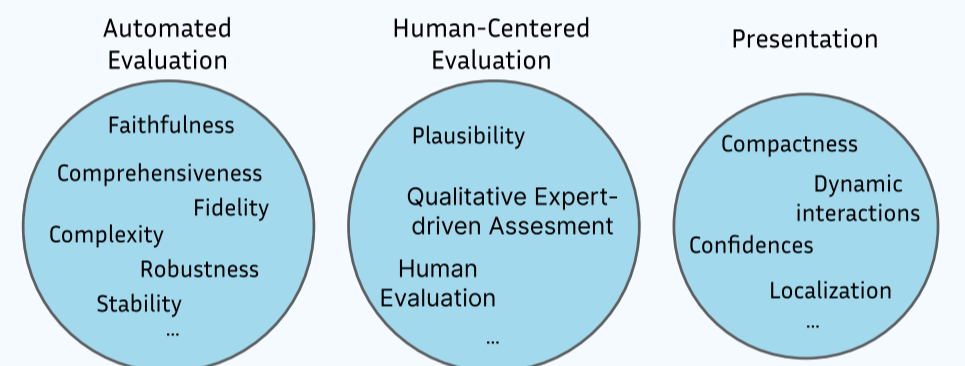
1

Unified XAI framework for most common architectural variants¹



2

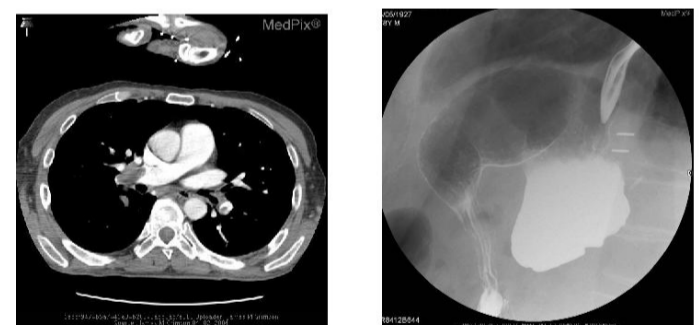
Multi-criteria XAI Evaluation Benchmark



3

Validation Studies

Healthcare: Medical Question-Answering²

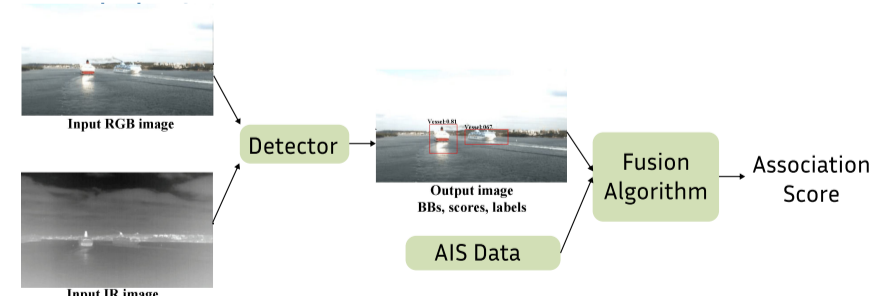


Q: which organ system is shown in the ct scan? A: lung, mediastinum, pleura

Q: what is abnormal in the gastrointestinal image? A: gastric volvulus (organoaxial)

Expected explanation: Highlight significant regions in the image corresponding to key terms in the question, leading to pathological insights.

Maritime: Multimodal Sensor Tracking Association³



Expected explanation: Highlight how visual cues interact with key AIS data that leads to a match or mismatch.

LEARN MORE:

Email: raisul.kibria@abo.fi

