The Computer Vision Group at TUM (https://cvg.cit.tum.de/) is seeking one student assistant (m/f/x). In this role, you will assist with implementing and evaluating a multimodal agent for autonomous driving data [1]. Conditioned on video frames, the agent will automatically generate interesting question-answer pairs for the current driving situation, which will lead to better planning for future trajectories. We are looking for student assistants to support this line of research flexibly for 10-20 hours per week.

We offer:

- Insight into cutting-edge research and opportunity to collaborate with domain experts - Practical exposure to applied computer vision
- Possibility to publish in top-tier venues as the lead author or co-author - Flexible working hours & TUM-standard compensation

Responsibilities:

- Implement and run a multimodal agent for frame-conditioned question-answer generation.
- Evaluate the agent on DriveLM [1].

Requirements:

- Strong knowledge of PyTorch and experience implementing neural networks.
- Strong communication skills and ability to work independently.
- Commitment to high code quality and maintainability.

Optional:

- Experience with vision-language models, agent frameworks (smolagents, LangGraph), or DriveLM.

The initial employment contract will be for 6 months, with the possibility of extension. The relevant project is highly research-oriented, and we encourage motivated students interested in gaining more research experience and publishing to directly contact us via email. Excellent outcomes can easily result in a publication in top-tier venues. If you're interested, please send your CV and grade reports to Dominik Schnaus at dominik.schnaus@tum.de and Xi Wang at xi.wang@inf.ethz.ch.

[1] Sima, Chonghao, et al. "Drivelm: Driving with graph visual question answering." European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024.