ValueAlignment-Dataset-Alpaca-zh：Alpaca格式的中文价值观微调数据集，共有72000条数据，每6000条数据对应中国社会主义价值观中的一个价值观维度（例如自由、平等）。每条数据由4个部分构成，分别为instruction、input、output与system。其中，input包含一个根据中国社会主义价值观中的某个特定维度构建的道德困境，并分别提供一个支持该特定价值观维度与一个违背该特定价值观维度的选项；output分别提供选择两个选项时，不同的理由与解释。

ValueAlignment-Dataset-Alpaca-en：Alpaca格式的英文价值观微调数据集，内容与上述中文微调数据集一致。

ValueAlignment-Dataset-DPO：支持DPO方法的偏好数据集，由72000条数据构成，每6000条数据对应中国社会主义价值观中的一个价值观维度。每条数据由5个部分构成，分别为instruction、input、chosen、rejected与system。其中，instruction包含一个基于某个价值观维度构造的道德困境，并提供支持与违背该价值观维度的选项；chosen提供选择支持该价值观维度的选项的理由与解释；rejected提供选择违背该价值观维度的选项的理由与解释。

ValueAlignment-Embadding：基于K-means方法进行语义聚类，共提供12个数据集。每个数据集包含75条数据，分别对应中国社会主义核心价值观的一个价值观维度。这些数据均从ValueAlignment-Dataset-DPO的12个数据集中随机选取。

embdding.py：用于进行语义聚类的代码，自动读取12个数据集中的文本内容，并根据文本内容将语义上相近的价值观维度归为一个簇，总共生成三个簇。

result：提供一个测试集，由2400条数据构成，每200条数据对应中国社会主义价值观中的一个价值观维度。同时提供微调前后的Llama、Qwen、DeepSeek等三个模型在该数据集上的表现。

mask-generation.py：用于控制模型回答测试集中的2400条问题，并同时提供支持价值观的选项的支持率与理由以及违背价值观的选项的支持率与理由。

acc\_precent.py：用于计算模型生成的回答中支持价值观的选项的平均支持率，以此作为评估模型微调效果的重要指标。