## Performance on MNLI-MISMATCHED Average accuracy (0-100) baseline auto diff manual K samples per class