

		iGSM-med pq						iGSM-med qp						iGSM-hard pq						iGSM-hard qp							row					
		in-dist			out-of-dist (OOD)			in-dist			out-of-dist (OOD)			in-dist			out-of-dist (OOD)			in-dist			out-of-dist (OOD)				avg					
original		99.9	99.1	92.0	88.4	84.5	77.7	91.6	100	99.3	92.4	89.9	84.8	78.3	91.4	100	99.4	94.9	93.7	91.4	88.4	84.0	91.4	100	99.4	94.5	93.3	91.8	88.3	85.3	89.4	91.8
retryrate0.05	retry (finetune, lora qv4e8)	100	98.4	86.9	82.1	75.6	67.5	92.4	99.5	98.7	85.7	79.5	73.5	64.7	91.1	100	99.2	92.6	89.7	86.3	81.0	75.6	95.2	100	98.9	91.4	88.0	84.4	80.6	73.9	92.3	87.5
retryrate0.05	retry (finetune, lora qv4e8)   with mask	100	99.3	91.8	88.6	83.6	75.6	93.8	99.7	99.5	91.5	87.7	81.9	75.0	91.9	100	99.1	93.7	92.0	89.0	84.8	80.6	92.1	100	99.3	94.8	92.5	90.3	85.9	82.8	91.5	90.9
retryrate0.05	retry (finetune, lora qv8e16)	99.9	98.5	88.7	84.3	77.6	70.1	91.6	99.8	98.9	87.2	82.7	76.2	68.3	91.7	100	99.2	93.7	90.2	88.0	83.6	79.3	95.1	99.9	98.9	92.1	89.0	86.0	81.7	76.2	90.9	88.6
retryrate0.05	retry (finetune, lora qv8e16)   with mask	99.9	99.2	91.7	88.9	82.9	77.0	93.4	99.9	99.5	91.7	88.1	82.4	76.0	92.4	100	99.1	94.1	92.5	90.1	86.0	82.1	92.1	100	99.4	94.6	92.5	90.7	86.6	83.1	91.9	91.3
retryrate0.05	retry (finetune, lora qv16e32)	99.9	98.7	89.4	84.7	79.6	72.7	92.5	99.9	98.9	87.8	83.4	77.1	69.8	91.4	100	99.3	93.9	91.3	88.7	84.9	80.9	95.3	99.9	99.0	93.0	89.6	87.0	83.0	76.6	91.7	89.3
retryrate0.05	retry (finetune, lora qv16e32)   with mask	100	99.2	91.9	88.5	83.3	76.6	93.3	99.9	99.4	92.4	88.9	83.5	77.5	92.5	100	99.3	95.3	93.0	91.2	86.7	83.3	92.1	100	99.4	95.2	92.9	91.9	88.2	84.4	91.8	91.7
retryrate0.05	retry (finetune, lora qv32e64)	99.9	98.8	89.9	85.7	80.8	73.8	92.7	99.9	99.1	88.9	83.7	78.1	71.4	91.7	100	99.4	94.5	91.9	89.6	86.0	81.7	95.2	99.9	99.1	93.4	90.6	88.1	85.0	79.5	91.7	90.0
retryrate0.05	retry (finetune, lora qv32e64)   with mask	100	99.3	92.2	88.6	84.0	76.5	93.6	99.9	99.5	92.5	89.0	84.1	78.1	92.4	100	99.4	95.1	93.2	90.8	86.4	83.0	92.1	100	99.5	94.8	93.1	91.8	88.4	84.3	91.7	91.8
retryrate0.05	retry (finetune, lora qv64e128)	100	99.0	89.9	86.3	81.0	74.9	93.0	99.8	99.2	89.4	85.1	79.3	72.7	91.5	100	99.3	95.1	92.5	90.5	87.3	83.3	95.2	100	99.2	94.0	91.5	89.4	86.2	81.1	91.9	90.6
retryrate0.05	retry (finetune, lora qv64e128)   with mask	100	99.3	92.7	89.3	84.5	78.0	93.1	99.8	99.4	92.2	88.3	82.9	76.6	92.1	100	99.4	95.0	92.9	91.2	87.2	83.6	92.4	100	99.5	95.8	93.5	92.5	88.3	85.7	91.9	91.9
retryrate0.05	retry (finetune, lora qv128e256)	100	98.9	90.2	86.7	82.2	75.5	92.7	99.9	99.2	89.5	84.8	79.7	73.3	91.6	100	99.5	94.7	93.0	89.9	86.7	83.8	95.4	99.9	99.2	94.8	92.2	89.7	86.0	81.8	91.5	90.7
retryrate0.05	retry (finetune, lora qv128e256)   with mask	100	99.3	92.7	90.0	85.2	78.2	93.3	99.9	99.5	93.1	89.2	84.4	78.4	92.5	100	99.4	95.5	93.8	91.6	87.6	84.1	92.5	99.9	99.5	95.6	93.4	92.3	88.5	85.4	92.1	92.2
retryrate0.05	retry (finetune, lora qv256e512)	100	98.9	90.5	86.7	81.3	74.5	92.2	99.8	99.2	89.9	85.4	79.3	73.6	91.4	100	99.5	95.2	93.2	91.3	88.2	85.2	95.2	99.9	99.3	94.3	92.1	89.2	86.2	81.9	91.0	90.8
retryrate0.05	retry (finetune, lora qv256e512)   with mask	100	99.3	92.6	89.7	84.7	78.0	93.2	99.9	99.4	92.3	88.9	83.3	77.5	92.4	100	99.4	95.7	93.9	92.0	88.4	85.4	92.5	99.9	99.6	96.2	94.3	93.2	89.7	86.4	91.9	92.3
retryrate0.05	retry (continued pretrain)	100	99.8	96.6	93.5	90.4	85.2	95.1	100	99.8	96.1	93.5	89.1	84.8	92.8	100	100	98.6	97.9	96.6	95.4	93.6	96.9	100	99.9	98.5	97.5	96.9	95.4	93.3	93.0	95.7
retryrate0.05	retry (continued pretrain)   with mask	100	99.9	96.4	93.2	88.9	84.3	94.8	100	99.8	96.3	92.8	89.3	84.5	92.8	100	99.9	98.7	97.8	96.8	95.6	93.2	96.7	100	99.9	98.7	98.0	97.3	96.1	93.8	93.4	95.6
retryrate0.05	retry (continued pretrain x 0.1)	100	99.1	91.3	86.9	82.1	76.2	93.4	99.9	99.0	90.0	87.1	81.2	74.7	90.7	100	99.6	95.1	93.2	92.2	89.2	86.1	94.2	100	99.6	95.4	93.0	91.7	89.7	86.7	91.6	91.6
retryrate0.05	retry (continued pretrain x 0.1)   with mask	100	99.1	90.6	87.2	82.7	76.9	92.6	99.9	99.1	90.3	86.7	81.6	75.0	91.2	100	99.6	95.5	93.4	92.6	89.8	86.1	94.4	100	99.3	95.1	93.3	91.5	89.3	86.7	91.3	91.7
retryrate0.05	retry (pretrain)	100	99.7	94.8	93.0	88.5	83.9	94.5	100	99.4	93.3	89.2	85.7	78.8	92.4	100	99.6	96.6	94.8	93.2	90.7	86.9	94.8	100	99.7	96.2	93.7	92.7	89.4	86.3	92.1	93.3