E Appendix: Unoptimized α

This appendix contains results for an unoptimized α hyperparameter, arbitrarily set at .10 and .01. We see that for most data sets, there is no need to ask a clarification question as the model already achieves the desired coverage. Much higher coverages (as in Table 2) are achievable for these data sets. For some more challenging data sets such as C150, HWU64 and IND, CICC yields small clarification questions while retaining a reasonably large number of clarification questions of size 1.

Setting	$1-\alpha$	th		Cov↑	Single↑	|CQ| ↓	Amb
ACID	.90	7	CICC	<u>.90</u>	.92	–	0
			B1	<u>.97</u>	.93	5	0
			B2	<u>.95</u>	**1**	–	0
			B3	<u>.99</u>	0	5	0
ATIS	.90	7	CICC	.88	.89	–	0
			B1	<u>.99</u>	.93	5	0
			B2	<u>.98</u>	**1**	–	0
			B3	<u>1</u>	0	5	0
B77/BERT	.90	7	CICC	<u>.98</u>	.79	**2.90**	.04
			B1	<u>.97</u>	.90	5	0
			B2	<u>.93</u>	**1**	–	0
			B3	<u>.99</u>	0	5	0
B77/DFCX	.90	4	CICC	<u>.91</u>	.66	2.63	.02
			B1	<u>.95</u>	.71	4.79	.27
			B2	.90	**.98**	**2.26**	0
			B3	<u>.97</u>	0	5	1
C150	.90	7	CICC	<u>.99</u>	.97	**2.66**	0
			B1	<u>.99</u>	.82	5	0
			B2	<u>.98</u>	**1**	–	0
			B3	<u>1</u>	0	5	0
HWU64	.90	7	CICC	<u>.90</u>	.97	**2.00**	0
			B1	<u>.96</u>	.79	5	0
			B2	<u>.90</u>	**1**	–	0
			B3	<u>.98</u>	0	5	0
IND	.90	7	CICC	<u>.91</u>	**.25**	**3.46**	.11
			B1	.88	.42	5	0
			B2	.70	1	–	0
			B3	<u>.91</u>	0	5	0
MTOD	.90	7	CICC	<u>.90</u>	.90	–	0
			B1	<u>.99</u>	.99	5	0
			B2	<u>.99</u>	**1**	–	0
			B3	<u>1</u>	0	5	0

Table 8: Test set results for $1-\alpha=.90$ where <u>underline</u> indicates meeting coverage requirement. **Bold** denotes best when meeting this requirement, omitted for last column due to missing ground truth for ambiguous.

Setting	$1-\alpha$	th		Cov↑	Single↑	|CQ| ↓	Amb
ACID	.99	7	CICC	1	**.77**	**3.00**	.10
			B1	.98	.85	5	0
			B2	.95	1	–	0
			B3	.99	0	5	0
ATIS	.99	7	CICC	.99	**.98**	**2.54**	0
			B1	.99	.73	5	0
			B2	.98	1	–	0
			B3	1	0	5	0
B77/BERT	.99	7	CICC	.98	**.79**	**2.90**	.04
			B1	.97	.90	5	0
			B2	.93	1	–	0
			B3	.99	0	5	0
B77/DFCX	.99	4	CICC	.97	0	5	1
			B1	.97	.05	5	.95
			B2	.90	1	–	0
			B3	.97	0	5	1
C150	.99	7	CICC	.99	**.97**	**2.66**	0
			B1	.99	.82	5	0
			B2	.98	1	–	0
			B3	1	0	5	0
HWU64	.99	7	CICC	.99	**.25**	**3.39**	.28
			B1	.98	.05	5	0
			B2	.90	1	–	0
			B3	.98	0	5	0
MTOD	.99	7	CICC	.99	**1**	–	0
			B1	1	.98	5	0
			B2	.99	**1**	–	0
			B3	1	0	5	0

Table 9: Test set results for $1-\alpha=.99$ where underline indicates meeting coverage requirement. **Bold** denotes best when meeting this requirement, omitted for last column due to missing ground truth for ambiguous.

F Appendix: Comparison results OOS detection

We here compare the results of OOS detection as reported by baselines. Note that these results were generated on different splits of the data and (where applicable), possibly using different open-domain samples, and that a direct comparison between results is invalid.

Dataset	Algorithm	F1↑	Accuracy↑
C150	CICC-OOS	.91	.68
	Zhan et al. (2021) 25%	.81	.88
	Zhan et al. (2021) 50%	.87	.88
	Zhan et al. (2021) 75%	.89	.88
	Cavalin et al. (2020)	.76	.73
B77	CICC-OOS	.90	.89
	Zhan et al. (2021) 25%	.74	.70
	Zhan et al. (2021) 50%	.80	.73
	Zhan et al. (2021) 75%	.87	.81

Table 10: Results for the OOS detection task.