1 Quantitative Experiment Results

In the following, we provide an extensive look at the different metrics and their values in our experiment. In general, the metrics are averaged between five independent runs. For each generated action, the reference(s) with the best results regarding a specific metric are marked in **bold**. For the lines of code (LoC) metric, we indicate whether the generation resulted in an in- (↑) or decrease (↓). Additionally, we state the number of generated designators (out of the five generated) that could be compiled successfully in the *Comp.* column. The first two tables (1 & 2) collect the results for the experiment with the `gpt-3.5-turbo-0301` model, the second pair of tables (3 & 4) for the `gpt-3.5-turbo-0613` model and the remaining two tables (5 & 6) describe the results for the `gpt-4-0613` model.

References

[1] C.-Y. Lin. ROUGE: A Package for Automatic Evaluation of Summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain, 2004. Association for Computational Linguistics.

[2] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu. BLEU: A Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics - ACL '02*, page 311, Philadelphia, Pennsylvania, 2001. Association for Computational Linguistics. doi: 10.3115/1073083.1073135.

[3] J. Pennington, R. Socher, and C. Manning. Glove: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar, 2014. Association for Computational Linguistics. doi: 10.3115/v1/D14-1162.

[4] M. Popović. chrF: Character n-gram F-score for automatic MT evaluation. In *Proceedings of the 10th Workshop on Statistical Machine Translation*, pages 392–395, 2015.

[5] C. Wingfield and L. Connell. Sensorimotor distance: A grounded measure of semantic similarity for 800 million concept pairs. *Behav Res*, Sept. 2022. ISSN 1554-3528. doi: 10.3758/s13428-022-01965-7.

[6] Z. Wu and M. Palmer. Verb Semantics and Lexical Selection. In *Proceedings of ACL 94*. arXiv, 1994. doi: 10.48550/ARXIV.CMP-LG/9406033.

[7] S. Zhou, U. Alon, S. Agarwal, and G. Neubig. CodeBERTScore: Evaluating Code Generation with Pretrained Models of Code, 2023.

Table 1: First half of the results for the `gpt-3.5-turbo-0301` model.

Actions				Semantic Action Similarity			Code Generation Quality					
Gen.	**Ref.**	**LoC**	**Comp.**	**WuP** [6]	**GloVe** [3]	**SMD** [5]	**BLEU** [2]	**R-1** [1]	**R-2** [1]	**R-L** [1]	**CBS** [7]	**chrF** [4]
C	Ha	52 ↑	5/5	20.00	02.69	08.55	79.97	73.90	69.73	73.90	95.52	82.18
C	Ho	23 ↓	0/5	**33.33**	**78.39**	14.36	34.91	56.07	37.53	56.07	94.44	54.49
C	O	46 ↓	0/5	**33.33**	72.08	04.53	**97.57**	**93.55**	**89.34**	**93.55**	98.39	**97.79**
C	P-U	46 ↓	0/5	25.00	51.54	07.03	**97.57**	**93.55**	**89.34**	**93.55**	**99.35**	**97.79**
C	P-D	11 ↓	0/5	25.00	75.76	05.19	04.13	22.37	09.35	18.30	86.55	23.04
C	P	45 ↓	5/5	25.00	20.33	08.16	76.25	81.77	72.73	76.24	95.16	80.58
C	S	40 ↓	5/5	22.22	38.08	08.87	57.10	71.35	64.64	71.35	94.62	66.73
C	W	33 ↓	5/5	25.00	30.31	**15.59**	82.15	67.08	53.61	67.08	95.93	81.54
Ha	C	47 ↓	0/5	20.00	02.69	08.55	85.95	80.65	68.85	80.65	95.38	86.43
Ha	Ho	47 ↓	0/5	**25.00**	18.63	09.78	81.01	69.83	58.68	66.44	95.00	81.19
Ha	O	45 ↓	0/5	**25.00**	13.74	08.27	87.26	80.22	67.39	79.34	95.28	83.47
Ha	P-U	45 ↓	0/5	20.00	16.82	07.10	87.26	80.22	67.22	79.12	95.27	83.79
Ha	P-D	31 ↓	0/5	20.00	11.31	11.42	60.72	67.35	59.14	66.94	93.89	69.02
Ha	P	53 ↓	5/5	20.00	31.58	07.76	89.24	80.85	74.83	75.53	95.22	89.74
Ha	S	55 ↓	5/5	18.18	37.71	09.35	**99.72**	**98.95**	**98.65**	**98.95**	**99.74**	**99.72**
Ha	W	35 ↓	5/5	20.00	**50.98**	**11.83**	87.06	68.75	54.74	68.75	97.79	86.19
Ho	C	66 ↑	0/5	33.33	**78.39**	14.36	10.59	34.08	19.80	30.96	89.85	33.25
Ho	Ha	35 ↓	5/5	25.00	18.63	09.79	39.98	62.46	57.19	62.46	94.15	54.58
Ho	O	10 ↓	0/5	**50.00**	73.97	10.39	01.52	30.51	20.00	30.51	90.09	19.73
Ho	P-U	16 ↓	0/5	33.33	62.85	06.69	09.43	33.13	23.10	33.13	93.15	28.74
Ho	P-D	8 ↓	5/5	33.33	74.81	**23.49**	00.94	30.16	23.81	30.16	88.05	18.52
Ho	P	39 ↓	5/5	33.33	36.53	11.88	**59.64**	**73.37**	**62.17**	**73.37**	**95.03**	**67.06**
Ho	S	28 ↓	5/5	28.57	36.97	05.13	29.95	65.43	56.09	65.17	93.76	48.55
Ho	W	17 ↓	2/5	33.33	43.97	01.58	34.51	36.85	24.31	33.73	89.75	38.42
O	C	36 ↓	0/5	33.33	72.08	04.53	67.12	80.10	75.28	80.10	95.18	77.76
O	Ha	55 ↑	5/5	25.00	13.74	08.27	**86.67**	73.64	69.09	73.64	95.58	**86.16**
O	Ho	33 ↓	0/5	**50.00**	73.97	10.39	57.76	69.82	59.40	68.45	94.95	69.26
O	P-U	36 ↓	0/5	33.33	49.41	03.12	72.47	**87.51**	**85.03**	**87.51**	96.03	82.04
O	P-D	10 ↓	0/5	33.33	**80.67**	12.36	04.38	19.52	07.96	16.28	85.32	23.76
O	P	45 ↓	5/5	33.33	25.42	07.85	76.25	81.77	74.13	81.77	95.68	80.59
O	S	47 ↑	5/5	28.57	31.94	07.71	71.24	71.91	68.59	71.91	**96.16**	76.09
O	W	18 ↓	5/5	33.33	32.29	**13.93**	31.31	50.40	30.17	48.82	92.87	46.18
P-U	C	39 ↓	0/5	25.00	51.54	07.03	73.80	70.66	58.45	70.66	95.22	72.78
P-U	Ha	37 ↓	4/5	20.00	16.82	07.10	52.03	67.94	63.80	67.94	94.50	63.17
P-U	Ho	51 ↑	5/5	**33.33**	62.85	06.69	**98.08**	**96.70**	**95.00**	**96.70**	**98.91**	**98.52**
P-U	O	36 ↓	0/5	**33.33**	49.41	03.12	64.17	71.70	61.24	71.70	94.87	65.55

Table 2: Second half of the results for the `gpt-3.5-turbo-0301` model.

Actions				Semantic Action Similarity			Code Generation Quality					
Gen.	**Ref.**	**LoC**	**Comp.**	**WuP** [6]	**GloVe** [3]	**SMD** [5]	**BLEU** [2]	**R-1** [1]	**R-2** [1]	**R-L** [1]	**CBS** [7]	**chrF** [4]
P-U	P-D	17 ↓	0/5	25.00	**66.61**	**16.94**	15.17	36.88	19.25	36.88	92.65	35.32
P-U	P	40 ↓	5/5	25.00	16.68	07.00	66.46	79.55	73.91	79.55	96.09	74.55
P-U	S	33 ↓	5/5	22.22	44.73	07.44	46.78	74.85	69.60	74.85	94.39	61.01
P-U	W	32 ↓	5/5	25.00	29.69	09.47	80.31	65.37	51.56	65.37	94.98	80.35
P-D	C	30 ↓	0/5	25.00	75.76	05.19	39.46	43.51	26.15	40.85	94.13	49.50
P-D	Ha	51 ↑	0/5	20.00	11.31	11.42	**73.24**	61.17	**56.33**	61.17	94.90	**75.96**
P-D	Ho	43 ↓	0/5	**33.33**	74.81	**23.49**	70.42	67.00	52.70	62.55	94.93	75.47
P-D	O	31 ↓	0/5	**33.33**	**80.67**	12.36	45.31	48.26	29.86	45.68	94.20	53.79
P-D	P-U	31 ↓	0/5	25.00	66.61	16.94	44.09	48.00	31.14	47.76	94.30	52.45
P-D	P	42 ↓	0/5	25.00	29.18	11.33	66.89	**67.39**	52.07	**66.27**	**94.96**	73.08
P-D	S	31 ↓	0/5	22.22	44.27	13.58	41.14	62.18	50.06	60.91	94.20	57.03
P-D	W	13 ↓	5/5	25.00	32.13	22.02	12.19	34.41	21.05	34.41	91.14	34.38
P	C	56 =	0/5	25.00	20.33	08.16	74.94	80.00	64.15	77.95	**95.81**	93.90
P	Ha	63 ↑	5/5	20.00	31.58	07.76	71.90	53.45	42.49	53.45	94.09	79.02
P	Ho	38 ↓	0/5	**33.33**	36.53	**11.88**	51.34	35.42	20.33	31.40	93.03	57.99
P	O	53 ↓	0/5	**33.33**	25.42	07.85	**78.29**	**81.25**	**66.92**	**81.25**	95.70	**94.54**
P	P-U	55 ↓	0/5	25.00	16.68	07.00	55.04	46.88	31.61	46.88	94.47	78.18
P	P-D	12 ↓	0/5	25.00	29.18	11.33	06.04	21.20	08.48	18.05	87.41	22.84
P	S	55 ↓	5/5	22.22	**54.07**	07.72	70.96	50.46	39.42	50.46	94.26	72.48
P	W	36 ↓	5/5	25.00	50.92	09.70	70.51	61.02	46.63	60.69	94.80	81.12
S	C	48 ↓	0/5	22.22	38.08	08.87	90.40	80.00	67.21	80.00	95.62	85.90
S	Ha	58 ↑	5/5	18.18	37.71	09.35	**93.61**	**85.71**	**82.37**	**85.71**	**98.59**	**93.60**
S	Ho	47 ↓	0/5	**28.57**	36.97	05.13	81.99	74.71	66.17	71.26	95.23	82.73
S	O	47 ↓	0/5	**28.57**	31.94	07.71	85.11	78.42	66.62	78.20	95.20	83.20
S	P-U	38 ↓	0/5	22.22	44.73	07.44	64.46	66.21	48.09	65.02	95.80	66.47
S	P-D	48 ↓	0/5	22.22	44.27	**13.58**	29.31	46.18	38.69	45.34	90.68	45.26
S	P	53 ↓	5/5	22.22	54.07	07.72	89.34	80.35	75.73	80.35	95.70	89.89
S	W	6 ↓	0/5	22.22	**54.18**	05.04	00.03	10.39	03.74	10.39	85.27	08.93
W	C	51 ↑	0/5	25.00	30.31	15.59	**84.52**	78.69	64.74	77.85	**96.53**	85.31
W	Ha	56 ↑	5/5	20.00	50.98	11.83	78.19	57.69	52.37	57.69	94.74	77.73
W	Ho	63 ↑	0/5	**33.33**	43.97	01.58	77.21	80.41	73.20	80.41	96.12	**91.94**
W	O	51 ↑	0/5	**33.33**	32.29	13.93	82.95	78.78	64.81	78.78	95.76	85.35
W	P-U	24 ↓	0/5	25.00	29.69	09.47	28.45	36.73	20.73	36.73	90.25	40.69
W	P-D	18 ↓	1/5	25.00	32.13	**22.02**	20.08	19.26	07.95	15.21	87.03	37.77
W	P	53 ↑	5/5	25.00	50.92	09.70	89.24	**80.68**	**76.16**	**80.68**	95.73	89.54
W	S	42 ↑	5/5	22.22	**54.18**	05.04	61.48	68.73	63.90	68.73	94.99	68.17

Table 3: First half of the results for the `gpt-3.5-turbo-0613` model.

Actions				Semantic Action Similarity			Code Generation Quality					
Gen.	**Ref.**	**LoC**	**Comp.**	**WuP** [6]	**GloVe** [3]	**SMD** [5]	**BLEU** [2]	**R-1** [1]	**R-2** [1]	**R-L** [1]	**CBS** [7]	**chrF** [4]
C	Ha	13 ↓	0/5	20.00	02.69	08.55	00.84	35.71	23.59	35.71	91.24	17.86
C	Ho	41 ↓	0/5	**33.33**	**78.39**	14.36	75.75	83.83	75.20	83.83	96.29	80.43
C	O	46 ↓	0/5	**33.33**	72.08	04.53	**98.61**	**96.77**	**94.26**	**96.77**	**98.73**	**98.73**
C	P-U	46 ↓	0/5	25.00	51.54	07.03	**98.61**	**96.77**	**94.26**	**96.77**	**98.73**	**98.73**
C	P-D	15 ↓	1/5	25.00	75.76	05.19	18.29	30.36	19.72	30.36	87.84	34.65
C	P	4 ↓	0/5	25.00	20.33	08.16	00.00	09.52	02.45	09.52	83.35	04.74
C	S	13 ↓	0/5	22.22	38.08	08.87	01.20	40.98	26.44	40.98	91.77	19.12
C	W	35 ↓	5/5	25.00	30.31	**15.59**	86.74	68.75	54.74	68.75	98.72	86.27
Ha	C	47 ↓	0/5	20.00	02.69	08.55	85.95	80.65	68.85	80.65	95.73	86.43
Ha	Ho	41 ↓	0/5	**25.00**	18.63	09.78	69.89	67.47	56.00	67.47	96.40	74.06
Ha	O	46 ↓	0/5	**25.00**	13.74	08.27	85.91	80.65	68.85	80.65	95.85	86.44
Ha	P-U	44 ↓	0/5	20.00	16.82	07.10	78.05	62.98	46.41	62.98	95.26	77.62
Ha	P-D	32 ↓	0/5	20.00	11.31	11.42	59.77	75.28	70.04	75.28	94.73	67.14
Ha	P	56 ↓	5/5	20.00	31.58	07.76	**93.79**	**82.72**	**79.35**	**82.72**	**98.08**	**94.68**
Ha	S	13 ↓	0/5	18.18	37.71	09.35	00.39	38.66	19.54	36.97	91.07	15.67
Ha	W	35 ↓	5/5	20.00	**50.98**	**11.83**	87.06	69.29	54.74	69.29	97.63	86.15
Ho	C	39 ↓	0/5	33.33	**78.39**	14.36	71.20	**70.74**	**59.10**	**70.74**	95.24	70.33
Ho	Ha	9 ↓	0/5	25.00	18.63	09.79	00.01	25.76	14.97	25.76	88.86	10.00
Ho	O	24 ↓	0/5	**50.00**	73.97	10.39	66.66	65.38	51.91	64.42	90.71	70.82
Ho	P-U	28 ↓	4/5	33.33	62.85	06.69	37.19	44.32	31.08	42.94	94.53	44.58
Ho	P-D	20 ↓	0/5	33.33	74.81	**23.49**	28.46	61.15	56.87	61.15	93.51	46.24
Ho	P	12 ↓	0/5	33.33	36.53	11.88	00.18	33.90	13.41	30.51	90.38	13.40
Ho	S	9 ↓	0/5	28.57	36.97	05.13	00.01	29.82	16.87	29.82	89.09	10.67
Ho	W	35 ↓	5/5	33.33	43.97	01.58	**87.06**	68.75	54.74	68.75	**97.66**	**85.82**
O	C	47 ↑	0/5	33.33	72.08	04.53	98.62	96.77	94.26	96.77	98.75	98.63
O	Ha	13 ↓	0/5	25.00	13.74	08.27	00.84	35.71	23.59	35.71	91.24	17.99
O	Ho	33 ↓	0/5	**50.00**	73.97	10.39	48.86	41.56	25.55	41.11	93.33	57.96
O	P-U	46 =	0/5	33.33	49.41	03.12	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
O	P-D	9 ↓	0/5	33.33	**80.67**	12.36	02.12	13.38	05.55	13.38	85.18	20.28
O	P	41 ↓	5/5	33.33	25.42	07.85	68.47	78.59	70.36	78.59	96.35	75.56
O	S	13 ↓	0/5	28.57	31.94	07.71	01.21	40.98	26.44	40.98	91.24	19.26
O	W	35 ↓	5/5	33.33	32.29	**13.93**	86.74	68.75	54.74	68.75	97.99	86.45
P-U	C	47 ↑	0/5	25.00	51.54	07.03	90.29	79.74	67.27	79.74	95.59	86.00
P-U	Ha	9 ↓	0/5	20.00	16.82	07.10	00.01	25.76	14.97	25.76	89.14	09.94
P-U	Ho	52 ↑	5/5	**33.33**	62.85	06.69	**99.09**	**97.83**	**97.51**	**97.83**	94.98	**99.38**
P-U	O	46 =	0/5	**33.33**	49.41	03.12	91.14	80.65	68.85	80.65	95.96	86.91

Table 4: Second half of the results for the `gpt-3.5-turbo-0613` model.

Actions				Semantic Action Similarity			Code Generation Quality					
Gen.	**Ref.**	**LoC**	**Comp.**	**WuP** [6]	**GloVe** [3]	**SMD** [5]	**BLEU** [2]	**R-1** [1]	**R-2** [1]	**R-L** [1]	**CBS** [7]	**chrF** [4]
P-U	P-D	43 ↓	0/5	25.00	**66.61**	**16.94**	91.96	89.58	87.57	89.58	96.68	94.08
P-U	P	40 ↓	5/5	25.00	16.68	07.00	61.58	78.42	72.02	78.42	96.02	70.87
P-U	S	9 ↓	0/5	22.22	44.73	07.44	00.01	29.82	16.87	29.82	89.46	10.55
P-U	W	35 ↓	5/5	25.00	29.69	09.47	86.74	68.75	54.74	68.75	**98.01**	86.45
P-D	C	55 ↑	0/5	25.00	75.76	05.19	72.88	72.08	58.11	72.08	96.19	84.98
P-D	Ha	55 ↑	5/5	20.00	11.31	11.42	85.43	71.82	67.48	71.82	95.42	85.89
P-D	Ho	52 ↑	0/5	**33.33**	74.81	**23.49**	**94.49**	**86.65**	**84.82**	**86.65**	**99.11**	**94.70**
P-D	O	46 =	0/5	**33.33**	**80.67**	12.36	90.54	79.57	67.21	79.57	96.64	86.54
P-D	P-U	46 =	0/5	25.00	66.61	16.94	86.25	79.57	67.21	79.57	97.29	86.16
P-D	P	32 ↓	0/5	25.00	29.18	11.33	46.66	64.15	46.15	62.89	95.23	59.99
P-D	S	10 ↓	0/5	22.22	44.27	13.58	00.02	29.31	16.67	29.31	89.95	11.12
P-D	W	37 ↓	5/5	25.00	32.13	22.02	76.90	61.65	48.34	61.65	95.24	83.54
P	C	61 ↑	0/5	25.00	20.33	08.16	66.18	**72.03**	**58.12**	**72.03**	**95.99**	**91.13**
P	Ha	63 ↑	5/5	20.00	31.58	07.76	71.68	52.59	41.61	52.59	94.17	79.54
P	Ho	41 ↓	0/5	**33.33**	36.53	**11.88**	47.26	30.17	15.64	30.17	92.27	53.28
P	O	4 ↓	5/5	**33.33**	25.42	07.85	00.04	09.43	05.84	09.43	79.45	13.45
P	P-U	4 ↓	5/5	25.00	16.68	07.00	00.04	09.43	05.84	09.43	79.45	13.45
P	P-D	38 ↓	0/5	25.00	29.18	11.33	56.93	47.37	30.25	44.75	93.29	66.56
P	S	18 ↓	1/5	22.22	**54.07**	07.72	13.94	25.22	15.46	25.22	90.32	22.68
P	W	39 ↓	5/5	25.00	50.92	09.70	**76.14**	64.71	50.98	64.71	94.81	87.15
S	C	47 ↓	0/5	22.22	38.08	08.87	85.95	80.65	68.85	80.65	95.79	86.42
S	Ha	58 ↑	5/5	18.18	37.71	09.35	94.51	87.50	84.27	87.50	98.43	95.32
S	Ho	41 ↓	0/5	**28.57**	36.97	05.13	67.49	64.29	54.03	64.29	95.87	71.71
S	O	46 ↓	0/5	**28.57**	31.94	07.71	85.91	80.65	68.85	80.65	95.93	86.40
S	P-U	17 ↓	0/5	22.22	44.73	07.44	45.55	49.48	33.97	48.18	89.13	54.88
S	P-D	46 ↓	0/5	22.22	44.27	**13.58**	**98.75**	**96.19**	**94.37**	**96.19**	**99.71**	**99.05**
S	P	56 ↑	5/5	22.22	54.07	07.72	93.79	82.29	79.35	82.29	99.01	94.37
S	W	35 ↓	5/5	22.22	**54.18**	05.04	87.06	68.75	54.74	68.75	98.44	87.16
W	C	47 ↑	0/5	25.00	30.31	15.59	91.13	80.65	68.85	80.615	95.58	86.87
W	Ha	12 ↓	0/5	20.00	50.98	11.83	00.06	18.84	10.36	18.84	89.64	10.40
W	Ho	52 ↑	0/5	**33.33**	43.97	01.58	88.13	83.24	76.50	83.24	96.92	89.86
W	O	21 ↓	0/5	**33.33**	32.29	13.93	31.63	52.00	37.56	52.00	90.51	45.28
W	P-U	46 ↑	0/5	25.00	29.69	09.47	91.11	80.65	68.85	80.65	95.63	86.88
W	P-D	46 ↑	0/5	25.00	32.13	**22.02**	**96.47**	**90.48**	**84.51**	**89.52**	**99.02**	**95.86**
W	P	41 ↑	5/5	25.00	50.92	09.70	68.86	77.71	68.59	77.71	95.16	74.83
W	S	12 ↓	0/5	22.22	**54.18**	05.04	00.11	25.00	11.63	25.00	90.13	11.66

Table 5: First half of the results for the `gpt-4-0613` model.

Actions				Semantic Action Similarity			Code Generation Quality					
Gen.	**Ref.**	**LoC**	**Comp.**	**WuP** [6]	**GloVe** [3]	**SMD** [5]	**BLEU** [2]	**R-1** [1]	**R-2** [1]	**R-L** [1]	**CBS** [7]	**chrF** [4]
C	Ha	39 ↓	4/5	20.00	02.69	08.55	55.87	64.88	59.77	64.88	93.88	62.48
C	Ho	48 ↑	0/5	**33.33**	**78.39**	14.36	87.33	83.49	76.25	80.77	95.78	87.86
C	O	46 ↓	0/5	**33.33**	72.08	04.53	**97.78**	**94.19**	**90.33**	**94.19**	**99.44**	**97.98**
C	P-U	41 ↓	0/5	25.00	51.54	07.03	81.69	83.45	76.75	83.45	96.75	87.39
C	P-D	19 ↓	0/5	25.00	75.76	05.19	23.79	39.37	24.58	39.37	91.38	39.31
C	P	47 =	5/5	25.00	20.33	08.16	79.69	81.96	75.17	81.96	96.36	83.23
C	S	28 ↓	4/5	22.22	38.08	08.87	37.03	66.19	59.52	66.19	93.44	50.90
C	W	33 ↓	5/5	25.00	30.31	**15.59**	80.94	67.19	51.84	65.89	95.72	81.33
Ha	C	44 ↓	0/5	20.00	02.69	08.55	80.58	74.55	61.55	73.68	95.49	77.51
Ha	Ho	45 ↓	0/5	**25.00**	18.63	9.78	75.45	69.41	58.70	67.12	95.55	78.52
Ha	O	46 ↓	0/5	**25.00**	13.74	8.27	86.86	78.20	65.60	77.33	96.02	81.52
Ha	P-U	43 ↓	0/5	20.00	16.82	07.10	78.19	72.93	59.65	72.05	94.91	77.03
Ha	P-D	32 ↓	0/5	20.00	11.31	11.42	63.57	66.17	58.03	64.97	94.18	70.96
Ha	P	54 ↓	5/5	20.00	31.58	07.76	91.06	81.51	76.64	78.32	96.48	91.63
Ha	S	55 ↓	5/5	18.18	37.71	09.35	**97.99**	**94.11**	**93.11**	**94.11**	**99.53**	**98.07**
Ha	W	35 ↓	5/5	20.00	**50.98**	**11.83**	87.06	68.75	54.74	68.75	97.92	86.29
Ho	C	13 ↓	4/5	33.33	**78.39**	14.36	06.05	35.04	25.57	35.04	90.62	24.50
Ho	Ha	29 ↓	4/5	25.00	18.63	09.78	29.51	55.44	48.15	55.44	93.06	44.19
Ho	O	13 ↓	4/5	**50.00**	73.97	10.39	07.89	32.70	23.25	32.70	90.58	24.93
Ho	P-U	18 ↓	0/5	33.33	62.85	06.69	12.07	33.64	23.24	33.64	93.12	30.53
Ho	P-D	12 ↓	4/5	33.33	74.81	**23.49**	09.26	37.86	31.78	37.86	89.24	25.96
Ho	P	41 ↓	5/5	33.33	36.53	11.88	**61.35**	**72.80**	**64.62**	**72.80**	**95.44**	**68.55**
Ho	S	26 ↓	4/5	28.57	36.97	05.13	28.03	60.48	51.02	60.48	92.94	43.90
Ho	W	12 ↓	1/5	33.33	43.97	01.58	17.42	26.22	13.94	22.06	87.36	22.99
O	C	47 ↑	0/5	33.33	72.08	04.53	**98.62**	**96.77**	**94.26**	**96.77**	**98.00**	**98.63**
O	Ha	54 ↑	1/5	25.00	13.74	08.27	81.85	71.78	67.59	71.78	96.31	82.99
O	Ho	47 ↑	0/5	**50.00**	73.97	10.39	87.76	84.80	78.19	82.02	96.27	88.52
O	P-U	37 ↓	0/5	33.33	49.41	03.12	75.76	88.77	86.77	88.77	96.79	84.24
O	P-D	18 ↓	0/5	33.33	**80.67**	12.36	22.02	31.79	19.25	26.92	88.03	38.97
O	P	47 ↑	5/5	33.33	25.42	07.85	80.29	83.75	76.85	83.75	96.27	83.79
O	S	34 ↓	4/5	28.57	31.94	07.71	47.68	65.05	60.11	65.05	93.89	57.67
O	W	23 ↓	3/5	33.33	32.29	**13.93**	49.20	46.60	32.86	43.93	92.09	53.15
P-U	C	41 ↓	0/5	25.00	51.54	07.03	77.11	72.44	60.20	72.44	96.11	75.45
P-U	Ha	37 ↓	4/5	20.00	16.82	07.10	53.59	63.18	57.16	63.18	93.66	60.52
P-U	Ho	43 ↓	4/5	**33.33**	62.85	06.69	78.64	**85.91**	**81.33**	**85.23**	**97.44**	82.41
P-U	O	38 ↓	0/5	**33.33**	49.41	03.12	71.48	74.38	62.42	74.38	95.95	71.12

Table 6: Second half of the results for the `gpt-4-0613` model.

Actions				Semantic Action Similarity			Code Generation Quality					
Gen.	**Ref.**	**LoC**	**Comp.**	**WuP** [6]	**GloVe** [3]	**SMD** [5]	**BLEU** [2]	**R-1** [1]	**R-2** [1]	**R-L** [1]	**CBS** [7]	**chrF** [4]
P-U	P-D	23 ↓	0/5	25.00	**66.61**	**16.94**	31.64	48.36	33.85	48.36	94.03	47.86
P-U	P	48 ↑	5/5	25.00	16.68	07.00	**79.81**	82.22	76.95	82.22	96.83	**84.36**
P-U	S	28 ↓	4/5	22.22	44.73	07.44	37.43	66.19	59.29	66.19	93.35	50.93
P-U	W	33 ↓	5/5	25.00	29.69	09.47	82.02	67.59	53.58	67.59	96.42	83.23
P-D	C	28 ↓	4/5	25.00	75.76	05.19	36.33	47.27	32.94	46.10	94.45	48.29
P-D	Ha	42 ↓	0/5	20.00	11.31	11.42	58.59	52.68	47.03	52.68	93.37	62.29
P-D	Ho	47 ↑	0/5	**33.33**	74.81	**23.49**	69.60	67.45	54.64	63.91	94.78	77.55
P-D	O	33 ↓	0/5	**33.33**	**80.67**	12.36	52.17	51.96	34.95	49.86	95.11	59.29
P-D	P-U	32 ↓	0/5	25.00	66.61	16.94	48.09	49.69	35.35	49.15	94.40	56.76
P-D	P	46 =	1/5	25.00	29.18	11.33	**75.64**	**72.72**	**59.66**	**71.84**	**95.88**	**80.10**
P-D	S	27 ↓	0/5	22.22	44.27	13.58	37.54	55.20	43.52	54.21	93.19	50.12
P-D	W	17 ↓	5/5	25.00	32.13	22.02	27.10	41.28	27.79	41.28	92.54	44.94
P	C	66 ↑	0/5	25.00	20.33	08.16	62.95	69.93	57.13	69.93	96.17	91.18
P	Ha	62 ↑	4/5	20.00	31.58	07.76	71.58	52.00	41.13	52.00	94.30	77.89
P	Ho	46 ↓	0/5	**33.33**	36.53	**11.88**	58.80	42.58	29.15	40.52	93.73	66.95
P	O	53 ↓	0/5	**33.33**	25.42	07.85	**79.65**	**83.35**	**71.85**	**83.35**	**96.55**	**95.18**
P	P-U	50 ↓	0/5	25.00	16.68	07.00	52.95	46.26	31.01	45.01	93.88	74.30
P	P-D	19 ↓	0/5	25.00	29.18	11.33	21.76	32.76	20.13	29.28	89.32	37.32
P	S	51 ↓	4/5	22.22	**54.07**	07.72	64.09	47.59	36.50	47.59	93.90	67.51
P	W	32 ↓	5/5	25.00	50.92	09.70	58.69	53.60	37.27	52.61	94.60	69.82
S	C	44 ↓	0/5	22.22	38.08	08.87	73.27	69.15	55.78	69.15	95.19	74.60
S	Ha	49 ↓	4/5	18.18	37.71	09.35	77.03	80.29	74.98	80.29	**97.25**	79.94
S	Ho	48 ↓	0/5	**28.57**	36.97	05.13	84.36	76.73	69.39	73.97	95.97	84.93
S	O	42 ↓	0/5	**28.57**	31.94	07.71	75.78	72.02	60.42	72.02	95.25	74.41
S	P-U	36 ↓	0/5	22.22	44.73	07.44	56.79	61.92	45.35	60.96	95.50	61.66
S	P-D	30 ↓	0/5	22.22	44.27	**13.58**	47.96	43.23	27.81	41.26	93.01	58.46
S	P	54 ↓	5/5	22.22	54.07	07.72	**90.15**	**81.14**	**76.80**	**81.14**	96.25	**90.85**
S	W	13 ↓	1/5	22.22	**54.18**	05.04	17.78	22.74	13.87	21.82	88.78	25.81
W	C	47 ↑	0/5	25.00	30.31	15.59	**90.35**	79.74	67.38	79.74	**98.08**	85.94
W	Ha	47 ↑	3/5	20.00	50.98	11.83	62.71	51.30	44.97	51.30	93.89	65.35
W	Ho	61 ↑	0/5	**33.33**	43.97	01.58	77.37	74.33	65.80	74.33	96.41	88.83
W	O	47 ↑	0/5	**33.33**	32.29	13.93	88.71	78.53	65.63	78.53	96.90	84.28
W	P-U	25 ↓	0/5	25.00	29.69	09.47	30.97	38.30	22.62	38.30	90.94	42.35
W	P-D	23 ↓	0/5	25.00	32.13	**22.02**	34.93	33.60	23.62	30.85	89.17	49.98
W	P	54 ↑	5/5	25.00	50.92	09.70	90.15	**80.97**	**76.80**	**80.97**	96.40	**90.48**
W	S	36 ↑	4/5	22.22	**54.18**	05.04	47.96	60.17	53.52	60.17	94.04	56.69