Supplementary

A. Related work

The concealment of stego determines the success of covert communication to a great extent. The concealment is primarily manifested in three aspects: perceptual, statistical, and semantic.

"Perceptual concealment" focuses on ensuring that the scheme generates stego with complete and fluent. This is the most basic requirement of LS. In pursuit of this objective, Fang et al. (Fang, Jaggi, and Argyraki 2017) proposed an LSTM-based LS scheme. This scheme segments the vocabulary into several sets based on bit blocks. It then selects tokens with the highest probability that corresponds to the secret information from the candidate pool. Ding et al. (Ding et al. 2023a) combined the conditional generation strategy with the replacement technique, using text sequences as auxiliary data in the stego generation process to enhance the embedding capabilities. Yang et al. (Yang et al. 2019) trained a language model to generate stegos. Using HC-based (Huffman coding) and FLC-based (Fixed length coding) to encode the candidate pool, and these stegos have excellent completeness and fluentness.

"Statistical concealment" requires the stegos' distribution to be closely the covers'. To achieve this goal, Yang et al. (Yang et al. 2021) designed an encoder-decoder structure, and VLC-based (Variable length coding) is used to encode the CP. To further reduce the distribution difference between cover and stego, Zhang et al. (Zhang et al. 2021) used adaptive dynamic grouping coding to recursively embed secret. Zhou et al. (Zhou et al. 2021) used a GAN (Generative Adversarial Network) to design an adaptive probability distribution steganography. These ensure statistical concealment.

"Semantic concealment" aims to generate the stego that is coherent and semantically controllable. To this end, Li et al. (Li et al. 2021) put forward an LS method based on the knowledge graph. Yang et al. (Yang et al. 2023) utilized semantic information encoding to embed secret information, realizing the effect of maintaining semantics and increasing the embedding capacity during the translation process. Wang et al. (Wang et al. 2023a) leveraged the relevance of social network context to enhance contextual semantic relevance while maintaining embedding rates. Lu et al. (Lu et al. 2023) used contextual learning and GPT2 guidance to generate stegos.

B. Supplement the effect of the embedding way

Table 8: Text quality, statistical analysis, and discourse-matching comparison of stegos generated by DAIRstega and baselines. The embedding rates of each way include 1Bpw (bit per word) and 2Bpw. "Avg." represents the overall performance. **Bold** is the best overall result. "_*" is the suboptimal overall result. The meanings of "bin" and "bit" are the same in Table 2.

Embedding	Param / Bpw	Text quality PPL ↓	ΔPcs ↓	Statistical analysis CS ↑	JSD ↓	ED ↓	MD ↓	ΔDP ↓	Discourse-matching ΔLDA ↓	Mauve ↑	BLEU ↑	Score ↑
LLMs+HC	bit = 1 / 1.00	13.331	5.639	96.54	45.79	0.263	8.255	3.46	0.014	$70.60_{\pm 27.36}$	$54.26_{\pm 4.25}$	$58.44_{\pm 4.07}$
	bit = 2 / 2.52	24.334	16.641	90.57	47.35	0.350	9.354	9.43	0.009	$64.88_{\pm 30.03}$	$55.00_{\pm 5.77}$	$56.19_{\pm 4.01}$
	Avg.	18.832	11.140	93.55	46.57	0.307	8.805	6.45	**0.011**	67.74	54.63	57.32
LLMs+ADG	τ = 0.3 / 1.34	12.750	5.058	97.20	45.59	0.240	8.316	2.80	0.010	$77.86_{\pm 23.65}$	$71.79_{\pm 3.03}$	$59.18_{\pm 5.84}$
	τ = 0.65 / 2.99	16.175	8.483	95.95	44.74	0.331	8.478	4.05	0.018	$75.60_{\pm 27.02}$	$62.60_{\pm 4.68}$	$58.60_{\pm 5.69}$
	Avg.	14.463*	6.770*	**96.58**	45.17*	0.285	8.397*	**3.42**	0.014	76.73*	67.20*	58.35*
LLMs+AC	CP = 8 / 1.25	73.454	65.762	82.16	51.64	0.746	15.415	17.84	0.014	$35.64_{\pm 22.65}$	$50.92_{\pm 29.30}$	$45.10_{\pm 3.07}$
	CP = 32 / 1.32	86.238	78.545	80.38	51.21	0.829	14.194	19.62	0.012	$37.18_{\pm 24.59}$	$48.01_{\pm 27.45}$	$47.51_{\pm 2.76}$
	Avg.	79.846	72.154	81.27	51.43	0.788	14.805	18.73	0.013*	36.41	49.47	46.31
LLMs+FLC	bit = 1 / 1.34	12.883	5.190	97.04	46.56	0.241	8.280	2.96	0.031	$79.48_{\pm 26.53}$	$54.88_{\pm 6.33}$	$58.44_{\pm 4.64}$
	bit = 2 / 2.67	30.481	22.788	93.86	46.30	0.328	8.939	6.14	0.011	$61.10_{\pm 29.30}$	$58.21_{\pm 9.88}$	$55.20_{\pm 3.39}$
	Avg.	21.682	13.989	95.45	46.43	0.284*	8.610	4.55	0.021	70.29	56.55	56.82
Ours	β=1 α=8 / 1.10	7.686	0.006	97.20	38.86	0.237	6.458	2.80	0.017	$79.91_{\pm 26.75}$	$77.88_{\pm 30.40}$	$65.76_{\pm 4.13}$
	β=1 α=32 / 1.11	8.184	0.492	97.65	38.48	0.217	6.293	2.35	0.022	$76.97_{\pm 26.47}$	$77.85_{\pm 30.42}$	$65.22_{\pm 4.61}$
	β=1 α=48 / 1.13	8.219	0.527	97.45	38.46	0.226	6.273	2.55	0.020	$78.70_{\pm 25.51}$	$76.53_{\pm 31.71}$	$65.25_{\pm 4.78}$
	Avg. (1Bpw)	8.030	0.342	97.43	38.60	0.227	6.341	2.57	0.020	78.53	77.42	65.41
	β=0.5 α=8 / 1.89	13.775	6.082	97.39	37.80	0.229	7.204	2.61	0.018	$80.83_{\pm 25.15}$	$73.19_{\pm 26.93}$	$62.49_{\pm 4.48}$
	β=0.5 α=32 / 2.56	20.876	13.184	92.43	39.69	0.389	8.976	7.57	0.029	$75.58_{\pm 20.11}$	$66.99_{\pm 27.08}$	$60.89_{\pm 3.85}$
	β=0.5 α=48 / 2.58	22.069	14.377	92.44	38.31	0.389	8.592	7.56	0.006	$72.40_{\pm 26.33}$	$79.64_{\pm 12.47}$	$60.29_{\pm 4.00}$
	Avg. (2Bpw)	18.907	11.214	94.09	38.60	0.336	8.258	5.91	0.018	76.27	73.27	61.22
	Avg.	**13.468**	**5.778**	95.76*	**38.60**	**0.281**	**7.299**	4.24*	0.019	**77.40**	**75.35**	**63.32**

Table 9: Anti-steganalysis comparison of stegos generated by DAIRstega and baselines. The embedding rates of each way include 1Bpw (bit per word) and 2Bpw. "Avg." represents the overall performance. **Bold** is the best overall result. "_*" is the suboptimal overall result.

Embedding	Param / Bpw	LS_CNN Acc ↓	LS_CNN F1 ↓	TS_CSW Acc ↓	TS_CSW F1 ↓	EILG Acc ↓	EILG F1 ↓	UP4LS Acc ↓	UP4LS F1 ↓
LLMs+HC	bit = 1 / 1.00	77.62$_{\pm 1.38}$	78.41$_{\pm 1.65}$	78.09$_{\pm 1.33}$	78.25$_{\pm 1.21}$	78.58$_{\pm 1.69}$	78.29$_{\pm 2.24}$	86.73$_{\pm 1.32}$	86.13$_{\pm 0.50}$
	bit = 2 / 2.52	77.54$_{\pm 1.50}$	77.57$_{\pm 2.05}$	76.03$_{\pm 1.74}$	76.09$_{\pm 1.33}$	78.29$_{\pm 0.72}$	78.95$_{\pm 1.47}$	88.19$_{\pm 1.00}$	88.43$_{\pm 2.61}$
	Avg.	77.58	77.99	77.06	77.17	78.44	78.62	87.46	87.28
LLMs+ADG	τ = 0.3 / 1.34	70.10$_{\pm 0.66}$	70.17$_{\pm 0.89}$	69.45$_{\pm 2.05}$	69.34$_{\pm 2.53}$	70.99$_{\pm 3.15}$	70.49$_{\pm 1.11}$	84.00$_{\pm 0.83}$	84.23$_{\pm 1.07}$
	τ = 0.65 / 2.99	63.06$_{\pm 2.49}$	63.69$_{\pm 3.61}$	62.22$_{\pm 1.46}$	61.86$_{\pm 1.18}$	66.14$_{\pm 1.38}$	66.15$_{\pm 1.38}$	80.42$_{\pm 0.59}$	80.57$_{\pm 0.58}$
	Avg.	66.58*	66.93*	65.84*	65.60*	68.57*	68.32*	82.21*	82.40*
LLMs+AC	CP = 8 / 1.25	90.15$_{\pm 1.17}$	89.98$_{\pm 1.21}$	89.69$_{\pm 0.93}$	89.37$_{\pm 0.88}$	86.92$_{\pm 1.12}$	86.52$_{\pm 0.96}$	93.62$_{\pm 0.89}$	93.37$_{\pm 0.40}$
	CP = 32 / 1.32	88.45$_{\pm 1.38}$	88.44$_{\pm 1.25}$	90.80$_{\pm 1.97}$	90.52$_{\pm 2.18}$	86.44$_{\pm 2.47}$	85.86$_{\pm 3.10}$	94.03$_{\pm 0.64}$	94.16$_{\pm 0.95}$
	Avg.	89.30	89.21	90.25	89.95	86.68	86.19	93.83	93.77
LLMs+FLC	bit = 1 / 1.34	78.53$_{\pm 1.20}$	80.05$_{\pm 1.39}$	82.72$_{\pm 1.27}$	83.32$_{\pm 1.10}$	76.67$_{\pm 0.85}$	76.15$_{\pm 1.96}$	89.03$_{\pm 0.25}$	89.20$_{\pm 0.80}$
	bit = 2 / 2.67	76.49$_{\pm 1.15}$	76.54$_{\pm 1.13}$	77.08$_{\pm 2.06}$	77.26$_{\pm 1.75}$	76.95$_{\pm 1.69}$	76.46$_{\pm 1.74}$	90.62$_{\pm 0.89}$	90.37$_{\pm 0.40}$
	Avg.	77.51	78.30	79.90	80.29	76.81	76.31	89.83	89.79
Ours	β=1 α=8 / 1.10	59.50$_{\pm 2.77}$	60.15$_{\pm 10.59}$	56.82$_{\pm 4.09}$	64.21$_{\pm 10.03}$	67.25$_{\pm 3.23}$	68.04$_{\pm 1.53}$	75.40$_{\pm 0.37}$	76.26$_{\pm 1.28}$
	β=1 α=32 / 1.11	59.35$_{\pm 1.25}$	59.99$_{\pm 5.51}$	54.05$_{\pm 4.72}$	61.68$_{\pm 11.17}$	62.28$_{\pm 0.50}$	62.66$_{\pm 2.31}$	71.97$_{\pm 0.64}$	72.75$_{\pm 1.21}$
	β=1 α=48 / 1.13	58.00$_{\pm 2.37}$	56.08$_{\pm 13.79}$	54.98$_{\pm 3.14}$	59.17$_{\pm 11.37}$	65.51$_{\pm 2.23}$	66.78$_{\pm 3.91}$	71.40$_{\pm 0.65}$	70.89$_{\pm 2.05}$
	Avg. (1Bpw)	58.95	58.74	55.28	61.69	65.01	65.83	72.92	73.30
	β=0.5 α=8 / 1.89	62.88$_{\pm 1.53}$	62.21$_{\pm 1.83}$	63.13$_{\pm 1.44}$	59.77$_{\pm 5.36}$	65.67$_{\pm 1.56}$	65.84$_{\pm 2.20}$	69.04$_{\pm 0.36}$	68.23$_{\pm 2.50}$
	β=0.5 α=32 / 2.56	72.06$_{\pm 1.17}$	71.78$_{\pm 1.21}$	68.73$_{\pm 1.29}$	68.19$_{\pm 3.44}$	71.71$_{\pm 1.95}$	70.34$_{\pm 2.41}$	76.41$_{\pm 2.56}$	75.90$_{\pm 2.73}$
	β=0.5 α=48 / 2.58	70.17$_{\pm 1.81}$	70.40$_{\pm 3.11}$	69.48$_{\pm 0.68}$	67.80$_{\pm 1.98}$	70.64$_{\pm 1.61}$	71.00$_{\pm 2.23}$	79.46$_{\pm 0.56}$	78.28$_{\pm 1.13}$
	Avg. (2Bpw)	68.37	68.13	67.11	65.25	69.34	69.06	74.97	74.14
	Avg.	**63.66**	**63.44**	**61.20**	**63.47**	**67.18**	**67.44**	**73.95**	**73.72**

C. Complete data for Table 4 and Table 5

In this section, we give the complete data of Table 4 and Table 5, as shown in Table 10 and Table 11. Table 10 contains the standard deviations of many datasets with different discourse, and Table 11 contains the standard deviations of multiple runs.

Table 10: Complete data for Table 4. The meanings of "bin" and "bit" are the same in Table 2. "$a_{\pm b}$" represents "Avg.$_{\pm \text{Std}}$". **Bold** is the best result. "_*" is the suboptimal result. The "Std"s of Ours Mauve and BLEU values are bigger. This is not due to the instability of 5-time runs, but is obtained from the effect on dozens of different discourse datasets, and represents the overall performance.

Schemes		Param / Bpw	Mauve ↑	BLEU ↑	Score ↑
Fang		bin = 1 / 1.00	2.15$_{\pm 0.87}$	0.90$_{\pm 1.58}$	42.58$_{\pm 1.30}$
		bin = 3 / 3.00	1.96$_{\pm 0.68}$	0.90$_{\pm 1.59}$	41.65$_{\pm 1.33}$
RNNstega		bit = 1 / 1.00	12.73$_{\pm 10.56}$	1.87$_{\pm 3.28}$	52.57$_{\pm 2.20}$
		bit = 3 / 2.63	11.97$_{\pm 10.56}$	1.83$_{\pm 3.21}$	52.36$_{\pm 1.80}$
ADG		τ = 0.5 / 4.38	18.54$_{\pm 13.89}$	2.72$_{\pm 4.79}$	47.55$_{\pm 2.25}$
LSCS		- / 1.12	2.40$_{\pm 1.78}$	31.95$_{\pm 12.89}$	59.22$_{\pm 3.14}$
PLMmark		- / -	7.22$_{\pm 4.88}$	0.15$_{\pm 0.27}$	43.50$_{\pm 0.79}$
Ours	β = 1	α = 8 / 1.10	79.91$_{\pm 26.75}$	76.37$_{\pm 31.85}$	**65.76$_{\pm 4.13}$**
		α = 32 / 1.11	76.97$_{\pm 26.47}$	77.85$_{\pm 30.42}$*	65.22$_{\pm 4.61}$
		α = 48 / 1.13	78.70$_{\pm 25.51}$	76.53$_{\pm 31.71}$	65.25$_{\pm 4.78}$*
	β = 0.5	α = 8 / 1.89	**80.83$_{\pm 25.12}$**	73.19$_{\pm 26.93}$	62.49$_{\pm 4.48}$
		α = 32 / 2.56	75.58$_{\pm 20.11}$	66.99$_{\pm 27.08}$	60.89$_{\pm 3.85}$
		α = 48 / 2.58	72.40$_{\pm 26.33}$	**79.64$_{\pm 12.47}$**	60.29$_{\pm 4.00}$

Table 11: Complete data for Table 5. The meanings of "bin" and "bit"are the same in Table 2. "$a_{\pm b}$" represents "$\text{Avg.}_{\pm \text{Std}}$". "Avg" and "Std" in this table are obtained by running 5 times. **Bold** is the best result. "_*" is the suboptimal result.

Schemes		Param / Bpw	LS_CNN Acc ↓	LS_CNN F1 ↓	TS_CSW Acc ↓	TS_CSW F1 ↓	EILG Acc ↓	EILG F1 ↓	UP4LS Acc ↓	UP4LS F1 ↓
Fang		bin = 1 / 1.00	$99.70_{\pm 0.29}$	$99.68_{\pm 0.31}$	$88.34_{\pm 21.22}$	$91.01_{\pm 15.69}$	$99.50_{\pm 0.25}$	$99.51_{\pm 0.25}$	$99.75_{\pm 0.06}$	$99.60_{\pm 0.09}$
		bin = 3 / 3.00	$99.75_{\pm 0.16}$	$99.76_{\pm 0.15}$	$89.03_{\pm 20.58}$	$92.39_{\pm 13.82}$	$99.59_{\pm 0.17}$	$99.52_{\pm 0.66}$	$99.88_{\pm 0.10}$	$99.80_{\pm 0.15}$
RNNstega		bit = 1 / 1.00	$85.81_{\pm 1.93}$	$86.03_{\pm 1.40}$	$81.39_{\pm 1.70}$	$82.59_{\pm 1.62}$	$85.86_{\pm 0.50}$	$86.29_{\pm 0.89}$	$97.42_{\pm 0.44}$	$95.97_{\pm 0.67}$
		bit = 3 / 2.63	$82.93_{\pm 2.18}$	$84.48_{\pm 2.20}$	$80.35_{\pm 2.38}$	$77.94_{\pm 3.04}$	$82.71_{\pm 0.17}$	$83.10_{\pm 1.02}$	$96.78_{\pm 0.82}$	$94.74_{\pm 1.34}$
ADG		τ = 0.5 / 4.38	$85.86_{\pm 1.73}$	$86.35_{\pm 1.82}$	$82.23_{\pm 2.40}$	$82.08_{\pm 2.96}$	$84.86_{\pm 1.74}$	$84.10_{\pm 1.80}$	$98.34_{\pm 0.39}$	$97.41_{\pm 0.58}$
LSCS		- / 1.12	$78.86_{\pm 3.07}$	$78.98_{\pm 2.83}$	$81.19_{\pm 6.62}$	$80.11_{\pm 8.26}$	$84.15_{\pm 0.41}$	$85.11_{\pm 0.79}$	$96.22_{\pm 2.08}$	$95.48_{\pm 3.11}$
PLMmark		- / -	$98.64_{\pm 0.92}$	$98.60_{\pm 0.95}$	$98.81_{\pm 1.83}$	$98.75_{\pm 1.93}$	$99.92_{\pm 0.17}$	$99.92_{\pm 0.11}$	$99.85_{\pm 0.30}$	$99.85_{\pm 0.30}$
Ours	$\beta = 1$	α = 8 / 1.10	$59.50_{\pm 2.77}$	$60.15_{\pm 10.59}$	$56.82_{\pm 4.09}$	$64.21_{\pm 10.03}$	$67.25_{\pm 3.23}$	$68.04_{\pm 1.53}$	$75.40_{\pm 0.37}$	$76.26_{\pm 1.28}$
		α = 32 / 1.11	$59.35_{\pm 1.25}$	$59.99_{\pm 5.51}$*	$\mathbf{54.05_{\pm 4.72}}$	$61.68_{\pm 11.17}$	$\mathbf{62.28_{\pm 0.50}}$	$62.66_{\pm 2.31}$*	$71.97_{\pm 0.64}$	$72.75_{\pm 1.21}$
		α = 48 / 1.13	$\mathbf{58.00_{\pm 2.37}}$	$\mathbf{56.08_{\pm 13.79}}$	$54.98_{\pm 3.14}$*	$\mathbf{59.17_{\pm 11.37}}$	$65.51_{\pm 2.23}$	$66.78_{\pm 3.91}$	$71.40_{\pm 0.65}$*	$70.89_{\pm 2.05}$*
	$\beta = 0.5$	α = 8 / 1.89	$62.88_{\pm 1.53}$	$62.21_{\pm 1.83}$	$63.13_{\pm 1.44}$	$59.77_{\pm 5.36}$*	$65.67_{\pm 1.56}$	$65.84_{\pm 2.20}$	$\mathbf{69.04_{\pm 0.36}}$	$\mathbf{68.23_{\pm 2.50}}$
		α = 32 / 2.56	$72.06_{\pm 1.17}$	$71.78_{\pm 1.21}$	$68.73_{\pm 1.29}$	$68.19_{\pm 3.44}$	$71.71_{\pm 1.95}$	$70.34_{\pm 2.41}$	$76.41_{\pm 2.56}$	$75.90_{\pm 2.73}$
		α = 48 / 2.58	$70.17_{\pm 1.81}$	$70.40_{\pm 3.11}$	$69.48_{\pm 0.68}$	$67.80_{\pm 1.98}$	$70.64_{\pm 1.61}$	$71.00_{\pm 2.23}$	$79.46_{\pm 0.56}$	$78.28_{\pm 1.13}$

D. Supplement data for anti-steganalysis

We also perform other 7 steganalysis methods to detect the stegos we generated, as shown in Table 12. These include non-BERT-based, BERT-based, and LLMs-based steganalysis methods.

Table 12: The supplemental anti-steganalysis capability of the DAIRstega from other 7 steganalysis works. **Bold** represents the best result and the best Avg. result. "_*" represents the suboptimal result.

Supplemental anti-steganalysis			VAE-Stega [1] AC Movie	AC Twitter	AC News	HC Movie	HC Twitter	HC News	Avg.	Ours $\alpha = 8$ $\beta = 1$	$\alpha = 8$ $\beta = 0.5$	$\alpha = 16$ $\beta = 1$	$\alpha = 16$ $\beta = 0.5$	$\alpha = 32$ $\beta = 1$	$\alpha = 32$ $\beta = 0.5$	Avg.
non-BERT-based	[2]	Acc ↓	57.63	53.75	52.55	61.63	57.75	74.38	59.61	54.38	**51.37**	51.60*	56.97	52.35	59.50	**54.38**
		F1 ↓	50.65	**34.63**	46.31*	60.89	66.65	73.89	55.50	47.09	47.55	49.47	52.35	49.20	57.70	**50.56**
	[3]	Acc ↓	64.50	**58.75**	62.50	87.63	80.75	94.38	74.75	64.75	61.05*	61.88	67.42	63.83	68.73	**64.61**
		F1 ↓	66.91	**56.46**	64.95	88.03	82.14	94.33	75.47	63.85	61.05*	67.03	66.16	63.82	68.83	**65.12**
BERT-based	[4]	Acc ↓	87.38	74.03	92.25	92.75	86.81	97.38	88.43	70.63	65.43*	**65.06**	65.87	65.56	73.25	**67.63**
		F1 ↓	87.39	74.21	91.99	93.21	86.85	97.35	88.50	70.44	**61.48**	65.28	65.15*	67.53	74.27	**67.36**
	[5]	Acc ↓	90.75	78.75	95.25	95.25	88.38	98.13	91.08	70.38	65.50*	**64.38**	71.12	65.50*	74.25	**68.52**
		F1 ↓	90.74	78.75	95.25	95.25	88.36	98.12	91.08	70.34	65.50	**64.27**	70.99	65.31*	74.25	**68.44**
	[6]	Acc ↓	92.50	75.38	95.88	94.75	88.32	97.62	90.74	70.12	**64.25**	**64.25**	73.75	66.50*	73.88	**68.79**
		F1 ↓	92.19	73.20	95.62	94.53	88.26	97.60	90.23	72.31	**66.19**	66.29*	71.47	70.02	69.04	**69.22**
LLMs-based	[7]	Acc ↓	89.63	78.13	93.13	88.63	76.38	86.75	85.44	63.38	59.63*	61.44	63.25	**59.38**	62.13	**61.53**
		F1 ↓	89.59	78.06	93.12	88.62	76.13	86.75	85.38	63.24	**60.15**	60.59*	63.13	59.15	61.86	**61.35**
	[8]	Acc ↓	97.38	92.38	98.38	98.13	93.88	98.88	96.50	75.38	72.63	71.75*	81.75	**69.38**	83.25	**75.69**
		F1 ↓	96.87	92.97	98.57	97.10	92.97	98.97	96.25	75.27	72.41	71.35*	81.55	**69.32**	83.15	**75.51**

[1] VAE-Stega: Linguistic Steganography Based on Variational Auto-Encoder. *IEEE Trans. Inf. Forensics and Security*, 880–895, 2021.
[2] A Fast and Efficient Text Steganalysis Method. *IEEE Signal Processing Lett.*, 26(4): 627–631, 2019.
[3] A Hybrid R-BiLSTM-C Neural Network Based Text Steganalysis. *IEEE Signal Processing Lett.*, 26(12): 1907–1911, 2019.
[4] High-performance Linguistic Steganalysis, Capacity Estimation and Steganographic Positioning. In *Proceeding of IWDW*, 2021.
[5] Small-Scale Linguistic Steganalysis for Multi-Concealed Scenarios. *IEEE Signal Processing Lett.*, 29: 130–134, 2022.
[6] SeSy: Linguistic Steganalysis Framework Integrating Semantic and Syntactic Features. *IEEE Signal Processing Lett.*, 31–35, 2021.
[7] Towards Next-Generation Steganalysis: LLMs Unleash the Power of Detecting Steganography. *arXiv:2405.09090*, 2024.
[8] Linguistic Steganalysis via LLMs: Two Modes for Efficient Detection of Strongly Concealed Stego. *arXiv:2406.04218*, 2024.

E. Example of cross-modal secret information hiding

Table 13: Examples of longer stegos generated by DAIRstega using cross-modal information. Bitstream is the binary corresponding to secret.

Secret&Bitsream	Instruction	Stegos	PPL	ΔLDA
Secret:  **Bitstream**: 100010010101 000001001110 010001110000 110100001010 000110100000 101000000000 000000000000 000000001101 010010010100 100001000100 010100100000 000000000000 000000000110 110000000000 00...(31520 bit)	Discuss the causes of the Great Depression.	The Great Depression was a severe worldwide economic depression that took place from 1929 to 1939. It was the longest and most severe depression ever experienced by the industrialized world, surpassing in overall duration the Great Depression of 1873. The Depression began in the United States after the stock market crash of October 1929, and peaked between 1930 and 1933. By some measures, the Great Depression was the longest economic depression in the history of the industrialized world. The Great Depression had many causes, including overproduction, lack of consumer demand, high tariffs, banking crisis, stock market crash, deflation, wage reductions, and the 1929 collapse of Worldwide prices for agricultural commodities. The collapse of agricultural prices resulted from overproduction, due to improvements in seed stock, fertilizers, pesticides, transportation, and farm equipment, as well as increases in acreage and livestock numbers.	2.5611	0.021
	Write a general overview of quantum computing.	Quantum computing is a new and revolutionary form of computing that is based on the principles of quantum mechanics. Unlike traditional computing, which uses binary bits to represent information, quantum computing uses quantum bits (qubits) which can exist in a superposition of 0 and 1 simultaneously. This means that quantum computers can process information exponentially faster than traditional computers, and they can also solve certain types of problems that are intractable for traditional computers. Quantum computing has the potential to revolutionize many fields, including machine learning, cryptography, drug discovery, and weather forecasting. For example, quantum computers can be used to train machine learning models more efficiently, which could lead to more accurate and efficient models. They can also be used to break current encryption methods, which could have serious implications for cybersecurity. In the field of drug discovery, quantum computers can be used to simulate molecular interactions, which could lead to the discovery of new and more effective drugs. Finally, in the field of weather forecasting, quantum computers can be used to simulate the complex interactions between different weather systems, which could lead to more accurate and precise forecasts.	1.9153	0.009
	...	...		