1 Automatic scores for General MT shared task

This document contains automatic scores calculated for the General MT submissions. While human judgement is going to be used as the official ranking of systems and their performance, you may want to use automatic scores in the discussion of your system description paper. Please, find the TEX source for tables in https://github.com/wmt-conference/wmt22-news-systems/tree/main/scores.

We use COMET (Rei et al., 2020) as the primary metric while ChrF (Popović, 2015) as the secondary metric, following recommendation by (Kocmi et al., 2021). The COMET scores are calculated with the default model `wmt20-comet-da`. The ChrF scores are calculated using all available references and SacreBLEU signature (Post, 2018) is `chrF2|nrefs:all|case:mixed|eff:yes|nc:6|nw:0|space:no|version:2.0.0`.

The different suffix represents the name of reference used for calculation (A, B, C, stud), references has been translated by different translators but with the same sponsor. A notable difference is Czech-English, where we are missing reference "A" for it's low quality, which was partly corrected and placed under "C". The second exception is Croatian reference "stud" which was created by students in contrast to "A" prepared by professionals. Lastly, testsets liv-en and ru-sah are reverse testsets to their oposite counterparts (i. e. "en" and "sah" are orifinal sources)

Table 1: Automatic scores for en-cs.
AUTOMATIC METRICS ARE NOT THE OFFICIAL WMT SYSTEM RANKING.

en-cs	Constrained	COMET-B ↓	COMET-C	ChrF-all
Online-W		97.8	79.3	70.4
Online-B		97.5	76.6	71.3
CUNI-Bergamot	✓	96.0	79.0	65.1
JDExploreAcademy	✓	95.3	77.8	67.2
Lan-Bridge		94.7	73.8	70.4
Online-A		92.2	71.1	67.5
CUNI-DocTransformer	✓	91.7	72.2	66.0
CUNI-Transformer	✓	86.6	68.6	64.2
Online-Y		83.7	62.3	64.5
Online-G		82.3	61.5	64.6

Table 2: Automatic scores for en-de.
AUTOMATIC METRICS ARE NOT THE OFFICIAL WMT SYSTEM RANKING.

en-de	Constrained	COMET-A ↓	COMET-B	ChrF-all
Online-W		65.5	64.4	68.7
JDExploreAcademy	✓	63.2	62.5	69.2
Online-B		62.3	61.9	69.7
Online-Y		61.1	60.9	68.7
Online-A		60.6	60.0	68.8
Online-G		60.2	59.3	68.4
Lan-Bridge		58.8	58.3	69.1
OpenNMT		57.2	57.0	66.7
PROMT		55.8	55.3	67.5

Table 3: Automatic scores for en-hr.
AUTOMATIC METRICS ARE NOT THE OFFICIAL WMT SYSTEM RANKING.

en-hr	Constrained	COMET-A ↓	COMET-stud	ChrF-all
Online-B		80.4	77.6	63.8
Lan-Bridge		79.6	76.7	63.7
GTCOM		77.4	74.7	63.3
Online-A		69.5	67.1	61.7
SRPOL	✓	69.4	67.6	61.4
HuaweiTSC	✓	67.6	66.3	61.8
NiuTrans	✓	65.5	63.4	61.5
Online-G		64.2	63.0	57.8
Online-Y		56.7	55.1	59.2

Table 4: Automatic scores for en-ja.
AUTOMATIC METRICS ARE NOT THE OFFICIAL WMT SYSTEM RANKING.

en-ja	Constrained	COMET-A ↓	ChrF-all
JDExploreAcademy	✓	65.1	36.1
NT5	✓	64.1	36.8
LanguageX		62.1	36.1
Online-B		60.8	35.5
DLUT	✓	60.5	36.1
Online-W		59.8	35.2
Online-Y		56.8	34.4
Lan-Bridge		56.5	34.1
Online-A		53.6	34.1
NAIST-NICT-TIT	✓	53.3	33.8
AISP-SJTU	✓	52.4	33.9
KYB	✓	31.8	28.6
Online-G		24.9	28.0

Table 5: Automatic scores for en-liv.
AUTOMATIC METRICS ARE NOT THE OFFICIAL WMT SYSTEM RANKING.

en-liv	Constrained	COMET-A ↓	ChrF-all
TAL-SJTU		-29.5	43.8
TartuNLP		-36.8	39.2
HuaweiTSC		-38.9	37.7
Liv4ever		-39.4	39.6
NiuTrans	✓	-81.9	30.5

Table 6: Automatic scores for en-ru.
AUTOMATIC METRICS ARE NOT THE OFFICIAL WMT SYSTEM RANKING.

en-ru	Constrained	COMET-A ↓	ChrF-all
Online-W		75.1	58.3
Online-G		73.1	59.5
Online-B		72.9	59.7
Online-Y		69.8	58.3
JDExploreAcademy	✓	69.6	58.4
Lan-Bridge		67.3	59.0
Online-A		67.3	58.1
PROMT		60.3	56.1
SRPOL	✓	59.7	56.4
HuaweiTSC	✓	59.2	56.1
eTranslation	✓	57.9	55.8

Table 7: Automatic scores for en-uk.
AUTOMATIC METRICS ARE NOT THE OFFICIAL WMT SYSTEM RANKING.

en-uk	Constrained	COMET-A ↓	ChrF-all
Online-B		73.2	59.3
GTCOM		72.0	59.0
Online-G		69.9	57.2
Lan-Bridge		65.7	58.8
Online-A		60.9	56.0
eTranslation	✓	54.5	54.8
HuaweiTSC	✓	54.4	54.8
Online-Y		51.9	54.9
ARC-NKUA		49.2	54.0

Table 8: Automatic scores for en-zh.
AUTOMATIC METRICS ARE NOT THE OFFICIAL WMT SYSTEM RANKING.

en-zh	Constrained	COMET-A ↓	COMET-B	ChrF-all
GTCOM		64.7	69.4	51.9
LanguageX		63.8	71.5	59.5
Online-B		61.8	80.4	70.3
JDExploreAcademy	✓	61.7	70.6	55.8
Lan-Bridge		61.4	69.4	53.7
Online-W		61.0	69.5	51.4
Manifold	✓	60.1	71.2	57.7
Online-Y		59.7	71.7	57.0
HuaweiTSC		59.5	73.1	61.0
Online-A		57.3	70.1	58.1
AISP-SJTU	✓	56.5	66.6	55.4
DLUT	✓	52.1	63.0	53.1
Online-G		51.2	62.5	52.4

Table 9: Automatic scores for cs-en.
AUTOMATIC METRICS ARE NOT THE OFFICIAL WMT SYSTEM RANKING.

cs-en	Constrained	COMET-B ↓	COMET-C	ChrF-all
Online-W		77.5	45.6	79.7
JDExploreAcademy	✓	74.7	49.0	75.0
Lan-Bridge		71.8	47.2	74.6
Online-B		71.8	47.4	74.5
CUNI-DocTransformer	✓	70.6	45.3	73.0
Online-A		69.8	44.3	74.1
CUNI-Transformer	✓	69.2	43.2	72.4
Online-G		63.0	38.8	71.1
SHOPLINE-PL	✓	61.1	39.6	70.1
Online-Y		58.6	35.2	68.8
ALMAnaCH-Inria	✓	19.3	4.9	58.3

Table 10: Automatic scores for de-en.
AUTOMATIC METRICS ARE NOT THE OFFICIAL WMT SYSTEM RANKING.

de-en	Constrained	COMET-A ↓	COMET-B	ChrF-all
JDExploreAcademy	✓	58.0	63.5	65.8
Online-B		56.9	63.6	65.7
Lan-Bridge		56.5	63.6	66.2
Online-G		55.2	61.7	66.3
Online-Y		54.6	61.4	65.7
Online-A		54.5	62.2	66.4
Online-W		54.3	61.7	65.4
PROMT		51.8	59.4	65.6
LT22	✓	25.6	33.3	58.4

Table 11: Automatic scores for ja-en.
AUTOMATIC METRICS ARE NOT THE OFFICIAL WMT SYSTEM RANKING.

ja-en	Constrained	COMET-A ↓	ChrF-all
NT5	✓	42.0	51.3
Online-W		41.2	51.7
JDExploreAcademy	✓	40.6	50.1
Online-B		39.6	49.9
DLUT	✓	37.2	49.8
NAIST-NICT-TIT	✓	33.4	48.3
Online-A		32.9	48.4
LanguageX		32.9	49.1
Online-Y		32.3	48.2
Lan-Bridge		31.9	48.7
AISP-SJTU	✓	30.1	48.0
Online-G		22.3	45.7
KYB	✓	17.3	43.4
AIST		-152.7	11.4

Table 12: Automatic scores for liv-en.
AUTOMATIC METRICS ARE NOT THE OFFICIAL WMT SYSTEM RANKING.

liv-en	Constrained	COMET-A ↓	ChrF-all
TartuNLP		-5.8	53.5
TAL-SJTU		-8.4	53.2
HuaweiTSC		-27.3	48.4
Liv4ever		-44.0	46.7
NiuTrans	✓	-88.3	35.6

Table 13: Automatic scores for ru-en.
AUTOMATIC METRICS ARE NOT THE OFFICIAL WMT SYSTEM RANKING.

ru-en	Constrained	COMET-A ↓	ChrF-all
Online-G		65.1	70.0
JDExploreAcademy	✓	64.9	68.9
Online-Y		64.1	68.2
Lan-Bridge		63.1	68.5
Online-B		63.1	68.3
Online-A		62.2	68.3
Online-W		61.6	66.3
HuaweiTSC	✓	60.9	68.5
SRPOL	✓	59.5	67.2
ALMAnaCH-Inria	✓	26.8	57.9

Table 14: Automatic scores for uk-en.
AUTOMATIC METRICS ARE NOT THE OFFICIAL WMT SYSTEM RANKING.

uk-en	Constrained	COMET-A ↓	ChrF-all
Online-B		62.5	67.2
Lan-Bridge		62.4	67.3
GTCOM		61.9	67.1
Online-G		57.4	66.0
Online-A		52.1	65.2
HuaweiTSC	✓	50.1	63.9
Online-Y		49.8	64.6
PROMT		49.6	64.7
ARC-NKUA		49.6	64.6
ALMAnaCH-Inria	✓	21.8	55.6

Table 15: Automatic scores for zh-en.
AUTOMATIC METRICS ARE NOT THE OFFICIAL WMT SYSTEM RANKING.

zh-en	Constrained	COMET-A ↓	COMET-B	ChrF-all
Online-G		45.6	36.2	60.9
JDExploreAcademy	✓	45.1	35.2	62.4
LanguageX		44.9	35.3	61.7
Lan-Bridge		43.0	34.0	59.1
HuaweiTSC		42.8	33.5	59.6
Online-B		42.1	32.8	59.4
AISP-SJTU	✓	41.6	32.8	60.5
Online-Y		40.8	31.0	58.6
Online-A		35.2	26.0	58.4
Online-W		31.6	23.1	55.6
NiuTrans	✓	31.3	22.3	57.2
DLUT	✓	30.6	22.0	56.3

Table 16: Automatic scores for cs-uk.
AUTOMATIC METRICS ARE NOT THE OFFICIAL WMT SYSTEM RANKING.

cs-uk	Constrained	COMET-A ↓	ChrF-all
AMU	✓	99.4	61.5
Online-B		94.3	64.0
GTCOM		93.4	63.9
Lan-Bridge		91.8	64.0
CharlesTranslator		90.8	61.5
HuaweiTSC	✓	90.7	62.6
CUNI-JL-JH	✓	90.0	61.6
Online-G		88.3	60.8
Online-A		87.8	62.2
CUNI-Transformer	✓	87.3	61.6
Online-Y		78.4	59.6
ALMAnaCH-Inria	✓	61.3	54.5

Table 17: Automatic scores for de-fr.
AUTOMATIC METRICS ARE NOT THE OFFICIAL WMT SYSTEM RANKING.

de-fr	Constrained	COMET-A ↓	ChrF-all
Online-B		70.5	74.6
Online-W		63.6	65.5
Online-Y		57.8	66.8
Online-A		52.2	64.5
Online-G		44.8	62.7
LT22	✓	10.4	54.4

Table 18: Automatic scores for fr-de.
AUTOMATIC METRICS ARE NOT THE OFFICIAL WMT SYSTEM RANKING.

fr-de	Constrained	COMET-A ↓	ChrF-all
Online-W		77.9	81.2
Online-B		63.7	68.7
Online-Y		61.6	67.5
Online-A		59.2	67.2
eTranslation	✓	55.4	68.4
Lan-Bridge		51.1	65.0
Online-G		48.2	66.0

Table 19: Automatic scores for ru-sah.
AUTOMATIC METRICS ARE NOT THE OFFICIAL WMT SYSTEM RANKING.

ru-sah	Constrained	COMET-A ↓	ChrF-all
Online-G		-17.1	47.0
Lan-Bridge	✓	-124.3	11.3

Table 20: Automatic scores for sah-ru.
AUTOMATIC METRICS ARE NOT THE OFFICIAL WMT SYSTEM RANKING.

sah-ru	Constrained	COMET-A ↓	ChrF-all
Online-G		31.1	55.5
Lan-Bridge	✓	-75.9	28.3

Table 21: Automatic scores for uk-cs.
AUTOMATIC METRICS ARE NOT THE OFFICIAL WMT SYSTEM RANKING.

uk-cs	Constrained	COMET-A ↓	ChrF-all
AMU	✓	104.8	60.7
Online-B		96.5	60.3
Lan-Bridge		94.5	60.4
HuaweiTSC	✓	91.4	59.6
CharlesTranslator		90.2	59.0
CUNI-JL-JH	✓	89.0	58.7
CUNI-Transformer	✓	88.5	59.0
Online-A		85.4	57.5
Online-G		84.2	56.3
GTCOM		80.2	55.8
Online-Y		78.6	55.3
ALMAnaCH-Inria	✓	62.4	50.7

References

Tom Kocmi, Christian Federmann, Roman Grundkiewicz, Marcin Junczys-Dowmunt, Hitokazu Matsushita, and Arul Menezes. 2021. To ship or not to ship: An extensive evaluation of automatic metrics for machine translation. In *Proceedings of the Sixth Conference on Machine Translation*, pages 478–494, Online. Association for Computational Linguistics.

Maja Popović. 2015. chrF: character n-gram F-score for automatic MT evaluation. In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pages 392–395, Lisbon, Portugal. Association for Computational Linguistics.

Matt Post. 2018. A call for clarity in reporting BLEU scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191, Brussels, Belgium. Association for Computational Linguistics.

Ricardo Rei, Craig Stewart, Ana C Farinha, and Alon Lavie. 2020. COMET: A neural framework for MT evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2685–2702, Online. Association for Computational Linguistics.