

Benchmarks



Benchmarks



- Support research on open-domain text generation models.
- Evaluate the ***diversity***, the ***quality*** and the ***consistency*** of the generated texts on various datasets/domains
- Facilitate the ***sharing*** of fine-tuned open-source implementations among researchers

Benchmarks



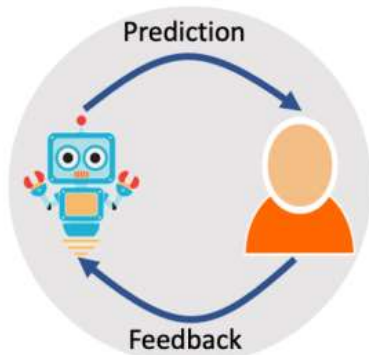
- Text generation benchmarks:
 - On specific text generation tasks
 - Machine Translation, Dialog Modeling, Summarization, etc.
 - On generic text evaluation tasks

General Text Evaluation Platforms

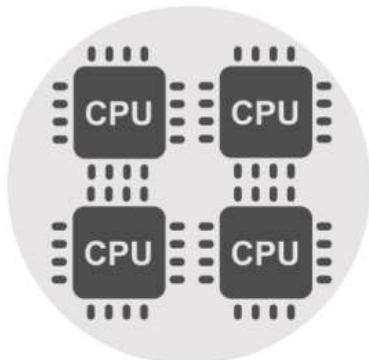
Comparison of Existing AI Evaluation Platforms

Features	OpenML	Topcoder	Kaggle	CrowdAI	ParlAI	CodaLab	EvalAI
AI Challenge Hosting	✗	✓	✓	✓	✗	✓	✓
Custom metrics	✗	✗	✗	✓	✓	✓	✓
Multiple phases/splits	✗	✗	✗	✓	✗	✓	✓
Open Source	✓	✗	✗	✓	✓	✓	✓
Remote Evaluation	✗	✗	✗	✗	✓	✓	✓
Human Evaluation	✗	✗	✗	✗	✓	✗	✓
Environments	✗	✗	✗	✓	✗	✗	✓





Human-in-the-loop
Evaluation



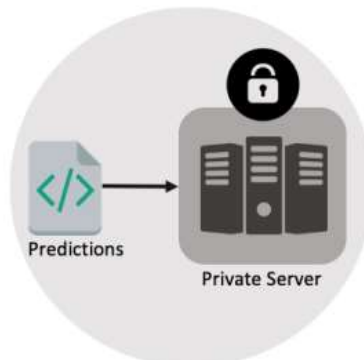
Faster Evaluation



Evaluation inside
RL Environments



EvalAI



Remote Evaluation



Custom Protocols



Portable

Task Specific Text Evaluation Platforms



DialoGLUE

Dialogue Language Understanding Evaluation

- Banking
- HWU
- Clerc
- Restaurant8k
- DSTC8 SGD
- TOP
- MultiWOZ 2.1



- SeqGAN - [SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient](#)
- MaliGAN - [Maximum-Likelihood Augmented Discrete Generative Adversarial Networks](#)
- RankGAN - [Adversarial ranking for language generation](#)
- LeakGAN - [Long Text Generation via Adversarial Training with Leaked Information](#)
- TextGAN - [Adversarial Feature Matching for Text Generation](#)
- GSGAN - [GANS for Sequences of Discrete Elements with the Gumbel-softmax Distribution](#)

<https://github.com/geek-ai/Taxygen>

Sitemap

- [SMT Book](#)
- [Research Survey Wiki](#)
- [Moses MT System](#)
- [Europarl Corpus](#)
- [News Commentary Corpus](#)
- [Online Evaluation](#)
- [Online Moses Demo](#)
- [Translation Tool](#)
- [WMT Workshop 2014](#)
- [WMT Workshop 2013](#)
- [WMT Workshop 2012](#)
- [WMT Workshop 2011](#)
- [WMT Workshop 2010](#)
- [WMT Workshop 2009](#)
- [WMT Workshop 2008](#)
- [WMT Workshop 2007](#)
- [WMT Workshop 2006](#)
- [WPT Workshop 2005](#)
- [ACL SIG MT](#)
- [Edinburgh SMT Group](#)
- [SE Times Corpus](#)

Statistical Machine Translation

This website is dedicated to research in statistical machine translation, i.e. the translation of text from one human language to another by a computer that learned how to translate from vast amounts of translated text.

Introduction to Statistical MT Research

- [The Mathematics of Statistical Machine Translation](#) by Brown, Della Petra, Della Pietra, and Mercer
- [Statistical MT Handbook](#) by Kevin Knight
- [SMT Tutorial \(2003\)](#) by Kevin Knight and Philipp Koehn
- ESSLI Summer Course on SMT (2005), [day1](#), [2](#), [3](#), [4](#), [5](#) by Chris Callison-Burch and Philipp Koehn.
- [MT Archive](#) by John Hutchins, electronic repository and bibliography of articles, books and papers on topics in machine translation and computer-based translation tools

Conferences and Workshops

See [comprehensive list of NLP meetings](#).

Software

- [Giza++](#) a training tool for IBM Model 1-5 ([version for gcc-4](#))
- [Moses](#), a complete SMT system
- [UCAM-SMT](#), the Cambridge Statistical Machine Translation system
- [Phrasal](#), a toolkit for phrase-based SMT
- [cdec](#), a decoder for syntax-based SMT
- [Joshua](#), a decoder for syntax-based SMT
- [Jane](#), decoder for syntax-based SMT
- [Pharaoh](#) a decoder for phrase-based SMT
- [Rewrite](#) a decoder for IBM Model 4
- [BLEU scoring tool](#) for machine translation evaluation

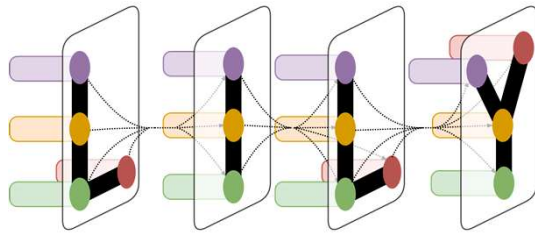
Parallel Corpora

- [LDC](#) Linguistic Data Consortium
- [Canadian Hansards](#)

<http://www.statmt.org/wmt19/>

WMT:

<http://www.statmt.org/wmt19/>



SummEval

- Provides data and evaluation platform for summarization tasks
- Enables benchmarks for more than 10 different trained and un-trained evaluation metrics

References and Additional Reading

- [1] Evaluation of Text Generation, Asli Celikyilmaz, Elizabeth Clark, Jianfeng Gao