

# SQuAD

## The Stanford Question Answering Dataset

### What is SQuAD?

Stanford Question Answering Dataset (SQuAD) is a new reading comprehension dataset, consisting of questions posed by crowdworkers on a set of Wikipedia articles, where the answer to every question is a segment of text, or *span*, from the corresponding reading passage. With 100,000+ question-answer pairs on 500+ articles, SQuAD is significantly larger than previous reading comprehension datasets.

Explore SQuAD and model predictions  
(/SQuAD-explorer/explore/1.1/dev/)

Read the paper (Rajpurkar et al. '16)  
(<http://arxiv.org/abs/1606.05250>)

### Getting Started

We've built a few resources to help you get started with the dataset. Download a copy of the dataset (distributed under the CC BY-SA 4.0 (<http://creativecommons.org/licenses/by-sa/4.0/legalcode>) license):

Training Set v1.1 (30 MB) (/SQuAD-explorer/dataset/train-v1.1.json)

Dev Set v1.1 (5 MB) (/SQuAD-explorer/dataset/dev-v1.1.json)

To evaluate your models, we have also made available the evaluation script we will use for official evaluation, along with a sample prediction file that the script will take as input. To run the evaluation, use `python evaluate-v1.1.py <path_to_dev-v1.1> <path_to_predictions>`.

Evaluation Script v1.1  
(<https://worksheets.codalab.org/rest/bundles/0xc83bf36cf871>)

Sample Prediction File (on Dev v1.1)  
(<https://worksheets.codalab.org/rest/bundles/0xc83bf36cf871>)

### Leaderboard

Since the release of our dataset, the community has made rapid progress! Here are the ExactMatch (EM) and F1 scores of the best models evaluated on the test and development sets of v1.1. Will your model outperform humans on the QA task?

Rank	Model	EM	F1
1 <small>Oct 17, 2017</small>	Interactive AoA Reader+ (ensemble) <i>Joint Laboratory of HIT and iFLYTEK</i>	79.083	86.450
2 <small>Oct 24, 2017</small>	FusionNet (ensemble) <i>Microsoft Business AI Solutions Team</i>	78.978	86.016
3 <small>Nov 03, 2017</small>	BiDAF + Self Attention + ELMo (single model) <i>Allen Institute for Artificial Intelligence</i>	78.580	85.833
3 <small>Oct 12, 2017</small>	r-net (ensemble) <i>Microsoft Research Asia</i> <a href="http://aka.ms/rnet">http://aka.ms/rnet</a> ( <a href="http://aka.ms/rnet">http://aka.ms/rnet</a> )	78.926	85.722
3 <small>Oct 22, 2017</small>	DCN+ (ensemble) <i>Salesforce Research</i>	78.852	85.996
4 <small>Oct 22, 2017</small>	BiDAF + Self Attention + ELMo (single model) <i>Allen Institute for Artificial Intelligence</i>	77.856	85.344
5 <small>Jul 25, 2017</small>	Interactive AoA Reader (ensemble) <i>Joint Laboratory of HIT and iFLYTEK Research</i>	77.845	85.297
6 <small>Aug 21, 2017</small>	Reinforced Mnemonic Reader (ensemble) <i>NUDT and Fudan University</i> <a href="https://arxiv.org/abs/1705.02798">https://arxiv.org/abs/1705.02798</a> ( <a href="https://arxiv.org/abs/1705.02798">https://arxiv.org/abs/1705.02798</a> )	77.678	84.888
7 <small>Nov 01, 2017</small>	Unnamed submission by allenlaobupt	76.828	84.396
8 <small>Oct 13, 2017</small>	r-net (single model) <i>Microsoft Research Asia</i> <a href="http://aka.ms/rnet">http://aka.ms/rnet</a> ( <a href="http://aka.ms/rnet">http://aka.ms/rnet</a> )	76.461	84.265
9 <small>Oct 22, 2017</small>	Conductor-net (ensemble) <i>CMU</i>	76.146	83.991
10 <small>Sep 08, 2017</small>	AIR-FusionNet (single model) <i>Microsoft Business AI Solutions team</i>	75.968	83.900

Once you have built a model that works to your expectations on the dev set, you submit it to get official scores on the dev and a hidden test set. To preserve the integrity of test results, we do not release the test set to the public. Instead, we require you to submit your model so that we can run it on the test set for you. Here's a tutorial walking you through official evaluation of your model:

**Submission Tutorial**

(<https://worksheets.codalab.org/worksheets/Ox8403d867f9a3>)

Because SQuAD is an ongoing effort, we expect the dataset to evolve. To keep up to date with major changes to the dataset, please subscribe:

## Have Questions?

Ask us questions at our google group (<https://groups.google.com/forum/#!forum/squad-stanford-qa>) or at [pranavsr@stanford.edu](mailto:pranavsr@stanford.edu) (<mailto:pranavsr@stanford.edu>).

Tweet (<https://twitter.com/share>)

11 <input type="button" value="Oct 22, 2017"/>	Interactive AoA Reader+ (single model) <i>Joint Laboratory of HIT and iFLYTEK</i>	75.821	83.843
11 <input type="button" value="Jul 14, 2017"/>	smarnet (ensemble) <i>Eigen Technology &amp; Zhejiang University</i>	75.989	83.475
12 <input type="button" value="Aug 18, 2017"/>	Reg-RaSoR (single model) <i>Google NY, Tel-Aviv University</i>	75.789	83.261
13 <input type="button" value="Jul 10, 2017"/>	DCN+ (single model) <i>Salesforce Research</i>	74.866	82.806
13 <input type="button" value="Sep 19, 2017"/>	SLQA (ensemble model) <i>Alibaba iDST NLP</i>	75.212	82.681
13 <input type="button" value="Oct 23, 2017"/>	DCN+ (single model) <i>Salesforce Research</i>	75.087	83.081
13 <input type="button" value="May 21, 2017"/>	MEMEN (ensemble) <i>Eigen Technology &amp; Zhejiang University</i> <a href="https://arxiv.org/abs/1707.09098">https://arxiv.org/abs/1707.09098</a> ( <a href="https://arxiv.org/abs/1707.09098">https://arxiv.org/abs/1707.09098</a> )	75.370	82.658
14 <input type="button" value="Mar 09, 2017"/>	ReasonNet (ensemble) <i>MSR Redmond</i> <a href="https://arxiv.org/abs/1609.05284">https://arxiv.org/abs/1609.05284</a> ( <a href="https://arxiv.org/abs/1609.05284">https://arxiv.org/abs/1609.05284</a> )	75.034	82.552
15 <input type="button" value="Oct 18, 2017"/>	Unnamed submission by eddielin	74.489	82.312
15 <input type="button" value="Jul 14, 2017"/>	Mnemonic Reader (ensemble) <i>NUDT and Fudan University</i> <a href="https://arxiv.org/abs/1705.02798">https://arxiv.org/abs/1705.02798</a> ( <a href="https://arxiv.org/abs/1705.02798">https://arxiv.org/abs/1705.02798</a> )	74.268	82.371
16 <input type="button" value="Jul 29, 2017"/>	SEDT (ensemble model) <i>CMU</i> <a href="https://arxiv.org/abs/1703.00572">https://arxiv.org/abs/1703.00572</a> ( <a href="https://arxiv.org/abs/1703.00572">https://arxiv.org/abs/1703.00572</a> )	74.090	81.761
17 <input type="button" value="Jul 06, 2017"/>	SSAE (ensemble) <i>Tsinghua University</i>	74.080	81.665
17 <input type="button" value="Aug 21, 2017"/>	Reinforced Mnemonic Reader (single model) <i>NUDT and Fudan University</i> <a href="https://arxiv.org/abs/1705.02798">https://arxiv.org/abs/1705.02798</a> ( <a href="https://arxiv.org/abs/1705.02798">https://arxiv.org/abs/1705.02798</a> )	73.188	81.816
17 <input type="button" value="Apr 22, 2017"/>	SEDT+BiDAF (ensemble) <i>CMU</i> <a href="https://arxiv.org/abs/1703.00572">https://arxiv.org/abs/1703.00572</a> ( <a href="https://arxiv.org/abs/1703.00572">https://arxiv.org/abs/1703.00572</a> )	73.723	81.530
17 <input type="button" value="Jul 25, 2017"/>	Interactive AoA Reader (single model) <i>Joint Laboratory of HIT and iFLYTEK Research</i>	73.639	81.931
17 <input type="button" value="Jan 24, 2017"/>	Multi-Perspective Matching (ensemble) <i>IBM Research</i> <a href="https://arxiv.org/abs/1612.04211">https://arxiv.org/abs/1612.04211</a> ( <a href="https://arxiv.org/abs/1612.04211">https://arxiv.org/abs/1612.04211</a> )	73.765	81.257
17 <input type="button" value="Feb 22, 2017"/>	BiDAF (ensemble)	73.744	81.525

18 <span>May 01, 2017</span>	jNet (ensemble) USTC & National Research Council Canada & York University <a href="https://arxiv.org/abs/1703.04617">https://arxiv.org/abs/1703.04617</a> ( <a href="https://arxiv.org/abs/1703.04617">https://arxiv.org/abs/1703.04617</a> )	73.010	81.517
19 <span>Oct 22, 2017</span>	Conductor-net (single) CMU	72.590	81.415
20 <span>Sep 20, 2017</span>	BiDAF + Self Attention (single model) Allen Institute for Artificial Intelligence	72.139	81.048
20 <span>Apr 12, 2017</span>	T-gating (ensemble) Peking University	72.758	81.001
21 <span>Sep 08, 2017</span>	SLQA (single model) Alibaba iDST NLP	72.023	80.648
22 <span>Nov 01, 2016</span>	Dynamic Coattention Networks (ensemble) Salesforce Research <a href="https://arxiv.org/abs/1611.01604">https://arxiv.org/abs/1611.01604</a> ( <a href="https://arxiv.org/abs/1611.01604">https://arxiv.org/abs/1611.01604</a> )	71.625	80.383
23 <span>Jul 14, 2017</span>	smarnet (single model) Eigen Technology & Zhejiang University	71.415	80.160
23 <span>Apr 13, 2017</span>	QFASE NUS	71.898	79.989
24 <span>Oct 27, 2017</span>	M-NET (single) UFL	71.016	79.835
24 <span>Jul 14, 2017</span>	Mnemonic Reader (single model) NUDT and Fudan University <a href="https://arxiv.org/abs/1705.02798">https://arxiv.org/abs/1705.02798</a> ( <a href="https://arxiv.org/abs/1705.02798">https://arxiv.org/abs/1705.02798</a> )	70.995	80.146
25 <span>Apr 02, 2017</span>	Ruminating Reader (single model) New York University <a href="https://arxiv.org/abs/1704.07415">https://arxiv.org/abs/1704.07415</a> ( <a href="https://arxiv.org/abs/1704.07415">https://arxiv.org/abs/1704.07415</a> )	70.639	79.456
26 <span>Mar 08, 2017</span>	ReasoNet (single model) MSR Redmond <a href="https://arxiv.org/abs/1609.05284">https://arxiv.org/abs/1609.05284</a> ( <a href="https://arxiv.org/abs/1609.05284">https://arxiv.org/abs/1609.05284</a> )	70.555	79.364
26 <span>Mar 14, 2017</span>	Document Reader (single model) Facebook AI Research <a href="https://arxiv.org/abs/1704.00051">https://arxiv.org/abs/1704.00051</a> ( <a href="https://arxiv.org/abs/1704.00051">https://arxiv.org/abs/1704.00051</a> )	70.733	79.353
26 <span>Dec 28, 2016</span>	FastQAExt German Research Center for Artificial Intelligence <a href="https://arxiv.org/abs/1703.04816">https://arxiv.org/abs/1703.04816</a> ( <a href="https://arxiv.org/abs/1703.04816">https://arxiv.org/abs/1703.04816</a> )	70.849	78.857
26 <span>Mar 24, 2017</span>	jNet (single model) USTC & National Research Council Canada & York University <a href="https://arxiv.org/abs/1703.04617">https://arxiv.org/abs/1703.04617</a> ( <a href="https://arxiv.org/abs/1703.04617">https://arxiv.org/abs/1703.04617</a> )	70.607	79.821

26 May 13, 2017	RaSoR (single model) <i>Google NY, Tel-Aviv University</i> <a href="https://arxiv.org/abs/1611.01436">https://arxiv.org/abs/1611.01436</a> ( <a href="https://arxiv.org/abs/1611.01436">https://arxiv.org/abs/1611.01436</a> )	70.849	78.741
26 Apr 14, 2017	Multi-Perspective Matching (single model) <i>IBM Research</i> <a href="https://arxiv.org/abs/1612.04211">https://arxiv.org/abs/1612.04211</a> ( <a href="https://arxiv.org/abs/1612.04211">https://arxiv.org/abs/1612.04211</a> )	70.387	78.784
27 Aug 30, 2017	SimpleBaseline (single model) <i>Technical University of Vienna</i>	69.600	78.236
28 Apr 12, 2017	SEDT+BiDAF (single model) <i>CMU</i> <a href="https://arxiv.org/abs/1703.00572">https://arxiv.org/abs/1703.00572</a> ( <a href="https://arxiv.org/abs/1703.00572">https://arxiv.org/abs/1703.00572</a> )	68.478	77.971
29 Jun 25, 2017	PQMN (single model) <i>KAIST &amp; AIBrain &amp; Crosscert</i>	68.331	77.783
30 Apr 12, 2017	T-gating (single model) <i>Peking University</i>	68.132	77.569
30 Jul 29, 2017	SEDT (single model) <i>CMU</i> <a href="https://arxiv.org/abs/1703.00572">https://arxiv.org/abs/1703.00572</a> ( <a href="https://arxiv.org/abs/1703.00572">https://arxiv.org/abs/1703.00572</a> )	68.163	77.527
30 Dec 28, 2016	FastQA <i>German Research Center for Artificial Intelligence</i> <a href="https://arxiv.org/abs/1703.04816">https://arxiv.org/abs/1703.04816</a> ( <a href="https://arxiv.org/abs/1703.04816">https://arxiv.org/abs/1703.04816</a> )	68.436	77.070
30 Nov 28, 2016	BiDAF (single model) <i>Allen Institute for AI &amp; University of Washington</i> <a href="https://arxiv.org/abs/1611.01603">https://arxiv.org/abs/1611.01603</a> ( <a href="https://arxiv.org/abs/1611.01603">https://arxiv.org/abs/1611.01603</a> )	67.974	77.323
31 Sep 19, 2017	AllenNLP BiDAF (single model) <i>Allen Institute for AI</i> <a href="http://allennlp.org/">http://allennlp.org/</a> ( <a href="http://allennlp.org/">http://allennlp.org/</a> )	67.618	77.151
31 Oct 26, 2016	Match-LSTM with Ans-Ptr (Boundary) (ensemble) <i>Singapore Management University</i> <a href="https://arxiv.org/abs/1608.07905">https://arxiv.org/abs/1608.07905</a> ( <a href="https://arxiv.org/abs/1608.07905">https://arxiv.org/abs/1608.07905</a> )	67.901	77.022
32 Feb 05, 2017	Iterative Co-attention Network <i>Fudan University</i>	67.502	76.786
33 Nov 01, 2016	Dynamic Coattention Networks (single model) <i>Salesforce Research</i> <a href="https://arxiv.org/abs/1611.01604">https://arxiv.org/abs/1611.01604</a> ( <a href="https://arxiv.org/abs/1611.01604">https://arxiv.org/abs/1611.01604</a> )	66.233	75.896
34 Oct 26, 2016	Match-LSTM with Bi-Ans-Ptr (Boundary) <i>Singapore Management University</i> <a href="https://arxiv.org/abs/1608.07905">https://arxiv.org/abs/1608.07905</a> ( <a href="https://arxiv.org/abs/1608.07905">https://arxiv.org/abs/1608.07905</a> )	64.744	73.743
35 Feb 19, 2017	Attentive CNN context with LSTM <i>NLPR, CASIA</i>	63.306	73.463

36 Nov 02, 2016	Fine-Grained Gating <i>Carnegie Mellon University</i> <a href="https://arxiv.org/abs/1611.01724">https://arxiv.org/abs/1611.01724</a> ( <a href="https://arxiv.org/abs/1611.01724">https://arxiv.org/abs/1611.01724</a> )	62.446	73.327
36 Sep 21, 2017	OTF dict+spelling (single) <i>University of Montreal</i> <a href="https://arxiv.org/abs/1706.00286">https://arxiv.org/abs/1706.00286</a> ( <a href="https://arxiv.org/abs/1706.00286">https://arxiv.org/abs/1706.00286</a> )	64.083	73.056
37 Sep 21, 2017	OTF spelling (single) <i>University of Montreal</i> <a href="https://arxiv.org/abs/1706.00286">https://arxiv.org/abs/1706.00286</a> ( <a href="https://arxiv.org/abs/1706.00286">https://arxiv.org/abs/1706.00286</a> )	62.897	72.016
38 Sep 21, 2017	OTF spelling+lemma (single) <i>University of Montreal</i> <a href="https://arxiv.org/abs/1706.00286">https://arxiv.org/abs/1706.00286</a> ( <a href="https://arxiv.org/abs/1706.00286">https://arxiv.org/abs/1706.00286</a> )	62.604	71.968
39 Sep 28, 2016	Dynamic Chunk Reader <i>IBM</i> <a href="https://arxiv.org/abs/1610.09996">https://arxiv.org/abs/1610.09996</a> ( <a href="https://arxiv.org/abs/1610.09996">https://arxiv.org/abs/1610.09996</a> )	62.499	70.956
40 Aug 27, 2016	Match-LSTM with Ans-Ptr (Boundary) <i>Singapore Management University</i> <a href="https://arxiv.org/abs/1608.07905">https://arxiv.org/abs/1608.07905</a> ( <a href="https://arxiv.org/abs/1608.07905">https://arxiv.org/abs/1608.07905</a> )	60.474	70.695
41 Aug 27, 2016	Match-LSTM with Ans-Ptr (Sentence) <i>Singapore Management University</i> <a href="https://arxiv.org/abs/1608.07905">https://arxiv.org/abs/1608.07905</a> ( <a href="https://arxiv.org/abs/1608.07905">https://arxiv.org/abs/1608.07905</a> )	54.505	67.748
	Human Performance <i>Stanford University</i> (Rajpurkar et al. '16) ( <a href="http://arxiv.org/abs/1606.05250">http://arxiv.org/abs/1606.05250</a> )	82.304	91.221