SQuAD2.0

The Stanford Question Answering Dataset

What is SQuAD?

Stanford **Qu**estion **A**nswering **D**ataset (SQuAD) is a 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, or the question might be unanswerable.

New **SQuAD2.0** combines the 100,000 questions in SQuAD1.1 with over 50,000 new, unanswerable questions written adversarially by crowdworkers to look similar to answerable ones. To do well on SQuAD2.0, systems must not only answer questions when possible, but also determine when no answer is supported by the paragraph and abstain from answering. SQuAD2.0 is a challenging natural language understanding task for existing models, and we release SQuAD2.0 to the community as the successor to SQuAD1.1. We are optimistic that this new dataset will encourage the development of reading comprehension systems that know what they don't know.

Explore SQuAD2.0 and model predictions (/SQuAD-explorer/explore/v2.0/dev/)

SQuAD2.0 paper (Rajpurkar & Jia et al. '18) (http://arxiv.org/abs/1806.03822)

SQuAD 1.1, the previous version of the SQuAD dataset, contains 100,000+ question-answer pairs on 500+ articles.

Explore SQuAD1.1 and model predictions (/SQuAD-explorer/explore/1.1/dev/)

SQuAD1.0 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 v2.0 (40 MB) (/SQuAD-explorer/dataset/train-v2.0.json)

Dev Set v2.0 (4 MB) (/SQuAD-explorer/dataset/dev-v2.0.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-v2.0.py <path_to_dev-v2.0> <path_to_predictions>`.

Evaluation Script v2.0
(https://worksheets.codalab.org/rest/bundles/0x6b567e1cf2e041ec80d7098f031c5c9e/contents/blob/)

Sample Prediction File (on Dev v2.0)
(https://worksheets.codalab.org/bundles/0x8731effab84f41b7b874a070e40f61e2/)

Once you have a 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/0x8212d84ca41c4150b555a075b19ccc05/)

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:

email address 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) and robinjia@stanford.edu (mailto:robinjia@stanford.edu).

Tweet ★ Star

Leaderboard

SQuAD2.0 tests the ability of a system to not only answer reading comprehension questions, but also abstain when presented with a question that cannot be answered based on the provided paragraph. How will your system compare to humans on this task?

Rank	Model	EM	F1
	Human Performance *Stanford University* (Rajpurkar & Jia et al. '18) (http://arxiv.org/abs/1606.05250)	86.831	89.452
1 Mar 20, 2019	BERT + DAE + AoA (ensemble) *Joint Laboratory of HIT and iFLYTEK Research*	**87.147**	**89.474**
2 Mar 15, 2019	BERT + ConvLSTM + MTL + Verifier (ensemble) *Layer 6 AI*	86.730	89.286
3 Mar 05, 2019	BERT + N-Gram Masking + Synthetic Self-Training (ensemble) *Google AI Language* https://github.com/google-research/bert (https://github.com/google-research/bert)	86.673	89.147
4 Apr 13, 2019	SemBERT(ensemble) *Shanghai Jiao Tong University*	86.166	88.886
4 May 14, 2019	SG-Net (ensemble) *Anonymous*	86.211	88.848
5 Mar 16, 2019	BERT + DAE + AoA (single model) *Joint Laboratory of HIT and iFLYTEK Research*	85.884	88.621
6 May 14, 2019	SG-Net (single model) *Anonymous*	85.229	87.926

7 Mar 05, 2019	BERT + N-Gram Masking + Synthetic Self-Training (single model) *Google AI Language* https://github.com/google-research/bert (https://github.com/google-research/bert)	85.150	87.715
8 Apr 16, 2019	Insight-baseline-BERT (single model) *PAII Insight Team*	84.834	87.644
8 Jan 15, 2019	BERT + MMFT + ADA (ensemble) *Microsoft Research Asia*	85.082	87.615
8 Mar 13, 2019	BERT + ConvLSTM + MTL + Verifier (single model) *Layer 6 AI*	84.924	88.204
9 Apr 11, 2019	SemBERT (single model) *Shanghai Jiao Tong University*	84.800	87.864
10 Jan 10, 2019	BERT + Synthetic Self-Training (ensemble) *Google AI Language* https://github.com/google-research/bert (https://github.com/google-research/bert)	84.292	86.967
11 Dec 13, 2018	BERT finetune baseline (ensemble) *Anonymous*	83.536	86.096
12 Dec 16, 2018	Lunet + Verifier + BERT (ensemble) *Layer 6 AI NLP Team*	83.469	86.043
12 Dec 21, 2018	PAML+BERT (ensemble model) *PINGAN GammaLab*	83.457	86.122
12 Mar 20, 2019	Bert-raw (ensemble) *None*	83.604	86.036
13 May 14, 2019	ATB (single model) *Anonymous*	82.882	86.002
13 Jan 14, 2019	BERT + MMFT + ADA (single model) *Microsoft Research Asia*	83.040	85.892
14 Jan 10, 2019	BERT + Synthetic Self-Training (single model) *Google AI Language* https://github.com/google-research/bert (https://github.com/google-research/bert)	82.972	85.810
14 Feb 26, 2019	BERT with Something (ensemble) *Anonymous*	83.051	85.737

15 Feb 15, 2019	BERT + NeurQuRI (ensemble) *2SAH*	82.803	85.703
15 Feb 16, 2019	Bert-raw (ensemble) *None*	83.175	85.635
15 Dec 15, 2018	Lunet + Verifier + BERT (single model) *Layer 6 AI NLP Team*	82.995	86.035
16 Dec 16, 2018	PAML+BERT (single model) *PINGAN GammaLab*	82.577	85.603
16 Mar 11, 2019	Bert-raw (ensemble) *None*	83.119	85.510
17 May 13, 2019	BERT-Base + QA Pre-training (single model) *Anonymous*	82.724	85.491
17 Feb 27, 2019	BERT + NeurQuRI (ensemble) *2SAH*	82.713	85.584
18 Nov 16, 2018	AoA + DA + BERT (ensemble) *Joint Laboratory of HIT and iFLYTEK Research*	82.374	85.310
18 Mar 02, 2019	Unnamed submission by cooelf	82.431	85.178
19 Dec 12, 2018	BERT finetune baseline (single model) *Anonymous*	82.126	84.820
19 Feb 28, 2019	BERT_s (single model) *Anonymous*	81.979	84.846
19 Dec 10, 2018	Candi-Net+BERT (ensemble) *42Maru NLP Team*	82.126	84.624
20 Feb 28, 2019	BERT-large+UBFT (single model) *anonymous*	81.573	84.535
21 Feb 25, 2019	BERT with Something (single model) *Anonymous*	81.110	84.386
21 Feb 15, 2019	BERT + NeurQuRI (single model) *2SAH*	81.257	84.342
22 Nov 16, 2018	AoA + DA + BERT (single model) *Joint Laboratory of HIT and iFLYTEK Research*	81.178	84.251

23 Mar 07, 2019	BERT + UnAnsQ (single model) *Anonymous*	80.749	83.851
23 Mar 20, 2019	Bert-raw (single) *None*	80.693	83.922
23 Apr 07, 2019	BERT + AL (single model) *Anonymous*	80.715	83.827
24 Dec 19, 2018	Candi-Net+BERT (single model) *42Maru NLP Team*	80.659	83.562
25 Jan 09, 2019	Unnamed submission by null	80.512	83.539
26 Mar 11, 2019	Bert-raw (single) *None*	80.411	83.457
26 Jan 22, 2019	BERT + NeurQuRI (single model) *2SAH*	80.591	83.391
27 Apr 19, 2019	Unnamed submission by null	80.354	83.329
28 Feb 16, 2019	Bert-raw (single model) *None*	80.343	83.243
28 Jan 09, 2019	Unnamed submission by null	80.343	83.221
29 Feb 19, 2019	BERT + UDA (single model) *Anonymous*	80.005	83.208
29 Dec 03, 2018	PwP+BERT (single model) *AITRICS*	80.117	83.189
30 Apr 10, 2019	bert (single model) *vinda msqjmxx*	79.971	83.184
30 Apr 04, 2019	BISAN-CC (single model) *Seoul National University & Hyundai Motors*	80.208	83.149
30 Dec 05, 2018	Candi-Net+BERT (single model) *42Maru NLP Team*	80.388	82.908
30 Nov 08, 2018	BERT (single model) *Google AI Language*	80.005	83.061

31 Feb 12, 2019	BERT + Sparse-Transformer *single model*	79.948	83.023
32 Dec 06, 2018	NEXYS_BASE (single model) *NEXYS, DGIST R7*	79.779	82.912
32 Mar 07, 2019	BERT uncased (single model) *Anonymous*	79.745	83.020
32 Feb 28, 2019	ST_bl *single model*	80.140	82.962
33 Feb 01, 2019	{bert-finetuning} (single model) *ksai*	79.632	82.852
34 Mar 14, 2019	{Anonymous} (single model) *Anonymous*	78.876	82.524
34 Nov 09, 2018	L6Net + BERT (single model) *Layer 6 AI*	79.181	82.259
35 Mar 14, 2019	BISAN (single model) *Seoul National University & Hyundai Motors*	78.481	81.531
35 Apr 24, 2019	BERT + WIAN (ensemble) *Infosys Limited*	78.650	81.497
36 Jan 09, 2019	Unnamed submission by null	78.301	81.350
37 Dec 14, 2018	BERT+AC(single model) *Hithink RoyalFlush*	78.052	81.174
38 Nov 06, 2018	SLQA+BERT (single model) *Alibaba DAMO NLP* http://www.aclweb.org/anthology/P18-1158 (http://www.aclweb.org/anthology/P18-1158)	77.003	80.209
39 Jan 05, 2019	synss (single model) *bert_finetune*	76.055	79.329
40 Dec 18, 2018	ARSG-BERT (single model) *TRINITI RESEARCH LABS, Active.ai* https://active.ai (https://active.ai)	74.746	78.227
40 Nov 05, 2018	MIR-MRC(F-Net) (single model) *Kangwon National University, Natural Language Processing Lab. & ForceWin, KP Lab.*	74.791	77.988

41 Sep 13, 2018	nlnet (single model) *Microsoft Research Asia*	74.272	77.052
42 Dec 22, 2018	Unnamed submission by null	73.234	76.790
42 Dec 29, 2018	MMIPN *Single*	73.505	76.424
43 Apr 20, 2019	BERT-Base (single model) *Dining Philosophers*	73.099	76.236
44 Oct 12, 2018	YARCS (ensemble) *IBM Research AI*	72.670	75.507
45 Nov 14, 2018	BERT+Answer Verifier (single model) *Pingan Tech Olatop Lab*	71.666	75.457
45 Nov 10, 2018	Unnamed submission by null	72.580	75.075
46 Sep 17, 2018	Unet (ensemble) *Fudan University & Liulishuo Lab* https://arxiv.org/abs/1810.06638 (https://arxiv.org/abs/1810.06638)	71.417	74.869
47 Jan 19, 2019	{BERT-base} (single-model) *Anonymous*	70.763	74.449
47 Aug 28, 2018	SLQA+ (single model) *Alibaba DAMO NLP* http://www.aclweb.org/anthology/P18-1158 (http://www.aclweb.org/anthology/P18-1158)	71.462	74.434
47 Apr 24, 2019	BERT-Base (single) *GreenflyAI* https://greenfly.ai (https://greenfly.ai)	71.699	74.430
47 Aug 15, 2018	Reinforced Mnemonic Reader + Answer Verifier (single model) *NUDT* https://arxiv.org/abs/1808.05759 (https://arxiv.org/abs/1808.05759)	71.767	74.295
48 Sep 14, 2018	SAN (ensemble model) *Microsoft Business Applications AI Research* https://arxiv.org/abs/1712.03556 (https://arxiv.org/abs/1712.03556)	71.316	73.704

49 Nov 10, 2018	Unnamed submission by null	70.718	73.403
50 Sep 14, 2018	Unet (single model) *Fudan University & Liulishuo Lab*	69.262	72.642
50 Aug 21, 2018	FusionNet++ (ensemble) *Microsoft Business Applications Group AI Research* https://arxiv.org/abs/1711.07341 (https://arxiv.org/abs/1711.07341)	70.300	72.484
50 Sep 26, 2018	Multi-Level Attention Fusion(MLAF) (single model) *Chonbuk National University, Cognitive Computing Lab.*	69.476	72.857
51 Dec 20, 2018	DocQA + NeurQuRI (single model) *2SAH*	68.766	71.662
52 Sep 13, 2018	BiDAF++ with pair2vec (single model) *UW and FAIR*	68.021	71.583
52 Aug 21, 2018	SAN (single model) *Microsoft Business Applications AI Research* https://arxiv.org/abs/1712.03556 (https://arxiv.org/abs/1712.03556)	68.653	71.439
52 Nov 10, 2018	Unnamed submission by null	68.653	71.124
53 Jul 13, 2018	VS^3-NET (single model) *Kangwon National University in South Korea*	67.897	70.884
53 Jun 24, 2018	KACTEIL-MRC(GFN-Net) (single model) *Kangwon National University, Natural Language Processing Lab.*	68.213	70.878
54 Jan 01, 2019	EBB-Net (single model) *Enliple AI*	66.610	70.303
55 Jun 25, 2018	KakaoNet2 (single model) *Kakao NLP Team*	65.719	69.381
56 Jul 11, 2018	abcNet (single model) *Fudan University & Liulishuo AI Lab*	65.256	69.206
56 Sep 13, 2018	BiDAF++ (single model) *UW and FAIR*	65.651	68.866

57 Jun 27, 2018	BSAE AddText (single model) *reciTAL.ai*	63.338	67.422
58 Aug 14, 2018	eeAttNet (single model) *BBD NLP Team* https://www.bbdservice.com (https://www.bbdservice.com)	63.327	66.633
58 May 30, 2018	BiDAF + Self Attention + ELMo (single model) *Allen Institute for Artificial Intelligence [modified by Stanford]*	63.372	66.251
59 Nov 27, 2018	Tree-LSTM + BiDAF + ELMo (single model) *Carnegie Mellon University*	57.707	62.341
59 May 30, 2018	BiDAF + Self Attention (single model) *Allen Institute for Artificial Intelligence [modified by Stanford]*	59.332	62.305
60 May 30, 2018	BiDAF-No-Answer (single model) *University of Washington [modified by Stanford]*	59.174	62.093

SQuAD1.1 Leaderboard

Here are the ExactMatch (EM) and F1 scores evaluated on the test set of SQuAD v1.1.

Rank	Model	EM	F1
	Human Performance *Stanford University* (Rajpurkar et al. '16) (http://arxiv.org/abs/1606.05250)	82.304	91.221
1 Oct 05, 2018	BERT (ensemble) *Google AI Language* https://arxiv.org/abs/1810.04805 (https://arxiv.org/abs/1810.04805)	**87.433**	**93.160**
2 May 14, 2019	ATB (single model) *Anonymous*	86.940	92.641
3 Feb 14, 2019	Knowledge-enhanced BERT (single model) *Anonymous*	85.944	92.425

4 Feb 28, 2019	ST_bl *single model*	85.430	91.976
4 Sep 26, 2018	nlnet (ensemble) *Microsoft Research Asia*	85.954	91.677
5 Feb 16, 2019	BERT+Sparse-Transformer *single model*	85.125	91.623
5 Sep 09, 2018	nlnet (ensemble) *Microsoft Research Asia*	85.356	91.202
5 Oct 05, 2018	BERT (single model) *Google AI Language* https://arxiv.org/abs/1810.04805 (https://arxiv.org/abs/1810.04805)	85.083	91.835
5 Mar 14, 2019	BISAN (single model) *Seoul National University & Hyundai Motors*	85.314	91.756
6 Jul 11, 2018	QANet (ensemble) *Google Brain & CMU*	84.454	90.490
6 Apr 21, 2019	Common-sense Governed BERT-123 (single model) *Jerry AGI Ragtag*	83.930	90.613
6 Feb 19, 2019	WD (single model) *Anonymous*	84.402	90.561
7 Jun 20, 2018	MARS (ensemble) *YUANFUDAO research NLP*	83.982	89.796
8 Mar 19, 2018	QANet (ensemble) *Google Brain & CMU*	83.877	89.737
8 Jul 08, 2018	r-net (ensemble) *Microsoft Research Asia*	84.003	90.147
8 Feb 21, 2019	WD1 (single model) *Anonymous*	83.804	90.429
9 Sep 09, 2018	nlnet (single model) *Microsoft Research Asia*	83.468	90.133
10 Sep 01, 2018	MARS (single model) *YUANFUDAO research NLP*	83.185	89.547

11 Jun 21, 2018	MARS (single model) *YUANFUDAO research NLP*	83.122	89.224
12 Mar 06, 2018	QANet (ensemble) *Google Brain & CMU*	82.744	89.045
13 Dec 23, 2018	MMIPN *Single*	81.580	88.948
14 Dec 17, 2018	ARSG-BERT (single model) *TRINITI RESEARCH LABS, Active.ai* https://active.ai (https://active.ai)	81.307	88.909
14 Jan 22, 2018	Hybrid AoA Reader (ensemble) *Joint Laboratory of HIT and iFLYTEK Research*	82.482	89.281
14 Feb 19, 2018	Reinforced Mnemonic Reader + A2D (ensemble model) *Microsoft Research Asia & NUDT*	82.849	88.764
14 May 09, 2018	MARS (single model) *YUANFUDAO research NLP*	82.587	88.880
14 Jun 20, 2018	QANet (single) *Google Brain & CMU*	82.471	89.306
14 Jan 03, 2018	r-net+ (ensemble) *Microsoft Research Asia*	82.650	88.493
14 Jan 05, 2018	SLQA+ (ensemble) *Alibaba iDST NLP*	82.440	88.607
15 Feb 02, 2018	Reinforced Mnemonic Reader (ensemble model) *NUDT and Fudan University* https://arxiv.org/abs/1705.02798 (https://arxiv.org/abs/1705.02798)	82.283	88.533
15 Feb 27, 2018	QANet (single model) *Google Brain & CMU*	82.209	88.608
16 Dec 22, 2017	AttentionReader+ (ensemble) *Tencent DPDAC NLP*	81.790	88.163
17 May 09, 2018	Reinforced Mnemonic Reader + A2D (single model) *Microsoft Research Asia & NUDT*	81.538	88.130

17 Dec 17, 2017	r-net (ensemble) *Microsoft Research Asia* http://aka.ms/rnet (http://aka.ms/rnet)	82.136	88.126
18 Feb 27, 2018	QANet (single model) *Google Brain & CMU*	80.929	87.773
18 Apr 23, 2018	r-net (single model) *Microsoft Research Asia*	81.391	88.170
18 Apr 03, 2018	KACTEIL-MRC(GF-Net+) (ensemble) *Kangwon National University, Natural Language Processing Lab.*	81.496	87.557
18 May 09, 2018	Reinforced Mnemonic Reader + A2D + DA (single model) *Microsoft Research Asia & NUDT*	81.401	88.122
19 Nov 17, 2017	BiDAF + Self Attention + ELMo (ensemble) *Allen Institute for Artificial Intelligence*	81.003	87.432
19 Feb 19, 2018	Reinforced Mnemonic Reader + A2D (single model) *Microsoft Research Asia & NUDT*	80.919	87.492
20 Apr 12, 2018	AVIQA+ (ensemble) *aviqa team*	80.615	87.311
20 Feb 12, 2018	Reinforced Mnemonic Reader + A2D (single model) *Microsoft Research Asia & NUDT*	80.489	87.454
21 Jan 13, 2018	SLQA+ *single model*	80.436	87.021
21 Jan 22, 2018	Hybrid AoA Reader (single model) *Joint Laboratory of HIT and iFLYTEK Research*	80.027	87.288
21 Jan 12, 2018	EAZI+ (ensemble) *Yiwise NLP Group*	80.426	86.912
21 Jan 04, 2018	{EAZI} (ensemble) *Yiwise NLP Group*	80.436	86.912
22 Feb 23, 2018	MAMCN+ (single model) *Samsung Research*	79.692	86.727
22 Feb 12, 2018	BiDAF + Self Attention + ELMo + A2D (single model) *Microsoft Research Asia & NUDT*	79.996	86.711

22 Mar 20, 2018	DNET (ensemble) *QA geeks*	80.164	86.721
23 Apr 10, 2018	Unnamed submission by null	80.027	86.612
24 Jan 03, 2018	r-net+ (single model) *Microsoft Research Asia*	79.901	86.536
24 Dec 28, 2017	SLQA+ (single model) *Alibaba iDST NLP*	79.199	86.590
24 Dec 05, 2017	SAN (ensemble model) *Microsoft Business AI Solutions Team* https://arxiv.org/abs/1712.03556 (https://arxiv.org/abs/1712.03556)	79.608	86.496
24 Jan 29, 2018	Reinforced Mnemonic Reader (single model) *NUDT and Fudan University* https://arxiv.org/abs/1705.02798 (https://arxiv.org/abs/1705.02798)	79.545	86.654
25 Oct 17, 2017	Interactive AoA Reader+ (ensemble) *Joint Laboratory of HIT and iFLYTEK*	79.083	86.450
25 Nov 05, 2018	MIR-MRC(F-Net) (single model) *ForceWin, KP Lab.*	79.083	86.288
26 Feb 01, 2018	Unnamed submission by null	78.999	86.151
27 Oct 24, 2017	FusionNet (ensemble) *Microsoft Business AI Solutions Team* https://arxiv.org/abs/1711.07341 (https://arxiv.org/abs/1711.07341)	78.978	86.016
27 Jun 01, 2018	MDReader *single model*	79.031	86.006
28 Oct 22, 2017	DCN+ (ensemble) *Salesforce Research* https://arxiv.org/abs/1711.00106 (https://arxiv.org/abs/1711.00106)	78.852	85.996
29 Nov 03, 2017	BiDAF + Self Attention + ELMo (single model) *Allen Institute for Artificial Intelligence*	78.580	85.833

29 Mar 29, 2018	KACTEIL-MRC(GF-Net+) (single model) *Kangwon National University, Natural Language Processing Lab.*	78.664	85.780
30 Nov 30, 2017	SLQA(ensemble) *Alibaba iDST NLP*	78.328	85.682
31 Jun 01, 2018	MDReader0 *single model*	78.171	85.543
31 Sep 18, 2018	BiDAF++ with pair2vec (single model) *UW and FAIR*	78.223	85.535
31 Jan 02, 2018	Conductor-net (ensemble) *CMU* https://arxiv.org/abs/1710.10504 (https://arxiv.org/abs/1710.10504)	78.433	85.517
31 Mar 19, 2018	aviqa (ensemble) *aviqa team*	78.496	85.469
31 May 09, 2018	KakaoNet (single model) *Kakao NLP Team*	78.401	85.724
32 Jan 29, 2018	test *single*	78.087	85.348
32 Jan 03, 2018	MEMEN (single model) *Zhejiang University* https://arxiv.org/abs/1707.09098 (https://arxiv.org/abs/1707.09098)	78.234	85.344
33 Jul 25, 2017	Interactive AoA Reader (ensemble) *Joint Laboratory of HIT and iFLYTEK Research*	77.845	85.297
34 Jan 10, 2018	Unnamed submission by null	77.436	85.130
34 Mar 20, 2018	DNET (single model) *QA geeks*	77.646	84.905
35 Sep 18, 2018	BiDAF++ (single model) *UW and FAIR*	77.573	84.858
36 Jan 23, 2018	MARS (single model) *YUANFUDAO research NLP*	76.859	84.739

36 Apr 10, 2018	Unnamed submission by null	77.489	84.735
36 Dec 06, 2017	AttentionReader+ (single) *Tencent DPDAC NLP*	77.342	84.925
37 Nov 06, 2017	Conductor-net (ensemble) *CMU* https://arxiv.org/abs/1710.10504 (https://arxiv.org/abs/1710.10504)	76.996	84.630
38 Dec 19, 2017	FRC (single model) *in review*	76.240	84.599
38 Dec 21, 2017	Jenga (ensemble) *Facebook AI Research*	77.237	84.466
39 Nov 01, 2017	SAN (single model) *Microsoft Business AI Solutions Team* https://arxiv.org/abs/1712.03556 (https://arxiv.org/abs/1712.03556)	76.828	84.396
40 Oct 13, 2017	r-net (single model) *Microsoft Research Asia* http://aka.ms/rnet (http://aka.ms/rnet)	76.461	84.265
40 Dec 13, 2017	RaSoR + TR + LM (single model) *Tel-Aviv University* https://arxiv.org/abs/1712.03609 (https://arxiv.org/abs/1712.03609)	77.583	84.163
41 Sep 26, 2018	{gqa} (single model) *FAIR*	77.090	83.931
41 May 14, 2018	VS^3-NET (single model) *Kangwon National University in South Korea*	76.775	84.491
42 Oct 22, 2017	Conductor-net (ensemble) *CMU*	76.146	83.991
43 Sep 08, 2017	FusionNet (single model) *Microsoft Business AI Solutions team* https://arxiv.org/abs/1711.07341 (https://arxiv.org/abs/1711.07341)	75.968	83.900
43 Jul 14, 2017	smarnet (ensemble) *Eigen Technology & Zhejiang University*	75.989	83.475

43 Oct 18, 2018	KAR (single model) *York University* https://arxiv.org/abs/1809.03449 (https://arxiv.org/abs/1809.03449)	76.125	83.538
43 Oct 22, 2017	Interactive AoA Reader+ (single model) *Joint Laboratory of HIT and iFLYTEK*	75.821	83.843
43 Mar 15, 2018	AVIQA-v2 (single model) *aviqa team*	75.926	83.305
44 Aug 18, 2017	RaSoR + TR (single model) *Tel-Aviv University* https://arxiv.org/abs/1712.03609 (https://arxiv.org/abs/1712.03609)	75.789	83.261
45 Oct 23, 2017	DCN+ (single model) *Salesforce Research* https://arxiv.org/abs/1711.00106 (https://arxiv.org/abs/1711.00106)	75.087	83.081
45 Oct 05, 2018	Unnamed submission by null	74.950	83.294
45 May 21, 2017	MEMEN (ensemble) *Eigen Technology & Zhejiang University* https://arxiv.org/abs/1707.09098 (https://arxiv.org/abs/1707.09098)	75.370	82.658
45 Nov 01, 2017	Mixed model (ensemble) *Sean*	75.265	82.769
46 Nov 17, 2017	two-attention-self-attention (ensemble) *guotong1988*	75.223	82.716
46 Jul 10, 2017	DCN+ (single model) *Salesforce Research* https://arxiv.org/abs/1711.00106 (https://arxiv.org/abs/1711.00106)	74.866	82.806
47 Jan 02, 2018	Conductor-net (single model) *CMU* https://arxiv.org/abs/1710.10504 (https://arxiv.org/abs/1710.10504)	74.405	82.742
47 Feb 06, 2018	Jenga (single model) *Facebook AI Research*	74.373	82.845

47 Aug 14, 2018	eeAttNet (single model) *BBD NLP Team* https://www.bbdservice.com (https://www.bbdservice.com)	74.604	82.501
48 Feb 13, 2018	SSR-BiDAF *ensemble model*	74.541	82.477
48 Oct 31, 2017	SLQA (single model) *Alibaba iDST NLP*	74.489	82.815
48 Mar 09, 2017	ReasoNet (ensemble) *MSR Redmond* https://arxiv.org/abs/1609.05284 (https://arxiv.org/abs/1609.05284)	75.034	82.552
49 Jul 14, 2017	Mnemonic Reader (ensemble) *NUDT and Fudan University* https://arxiv.org/abs/1705.02798 (https://arxiv.org/abs/1705.02798)	74.268	82.371
50 Dec 23, 2017	S^3-Net (ensemble) *Kangwon National University in South Korea*	74.121	82.342
50 Oct 27, 2017	Unnamed submission by null	74.489	82.312
51 Jul 25, 2017	Interactive AoA Reader (single model) *Joint Laboratory of HIT and iFLYTEK Research*	73.639	81.931
51 Jul 29, 2017	SEDT (ensemble model) *CMU* https://arxiv.org/abs/1703.00572 (https://arxiv.org/abs/1703.00572)	74.090	81.761
51 Nov 06, 2017	Conductor-net (single) *CMU* https://arxiv.org/abs/1710.10504 (https://arxiv.org/abs/1710.10504)	73.240	81.933
51 Jul 06, 2017	SSAE (ensemble) *Tsinghua University*	74.080	81.665
51 Dec 14, 2017	Jenga (single model) *Facebook AI Research*	73.303	81.754

51 Feb 22, 2017	BiDAF (ensemble) *Allen Institute for AI & University of Washington* https://arxiv.org/abs/1611.01603 (https://arxiv.org/abs/1611.01603)	73.744	81.525
51 Apr 22, 2017	SEDT+BiDAF (ensemble) *CMU* https://arxiv.org/abs/1703.00572 (https://arxiv.org/abs/1703.00572)	73.723	81.530
51 Jan 24, 2017	Multi-Perspective Matching (ensemble) *IBM Research* https://arxiv.org/abs/1612.04211 (https://arxiv.org/abs/1612.04211)	73.765	81.257
51 May 01, 2017	jNet (ensemble) *USTC & National Research Council Canada & York University* https://arxiv.org/abs/1703.04617 (https://arxiv.org/abs/1703.04617)	73.010	81.517
52 Apr 17, 2018	Unnamed submission by null	72.831	80.622
52 Apr 17, 2018	Unnamed submission by null	72.831	80.622
52 Nov 16, 2017	two-attention-self-attention (single model) *guotong1988*	72.600	81.011
52 Oct 22, 2017	Conductor-net (single) *CMU*	72.590	81.415
53 Sep 20, 2017	BiDAF + Self Attention (single model) *Allen Institute for Artificial Intelligence* https://arxiv.org/abs/1710.10723 (https://arxiv.org/abs/1710.10723)	72.139	81.048
54 Dec 15, 2017	S^3-Net (single model) *Kangwon National University in South Korea*	71.908	81.023
54 Apr 12, 2017	T-gating (ensemble) *Peking University*	72.758	81.001
55 Mar 03, 2018	AVIQA (single model) *aviqa team*	72.485	80.550

56 Nov 06, 2017	attention+self-attention (single model) *guotong1988*	71.698	80.462
57 Nov 01, 2016	Dynamic Coattention Networks (ensemble) *Salesforce Research* https://arxiv.org/abs/1611.01604 (https://arxiv.org/abs/1611.01604)	71.625	80.383
58 Jul 14, 2017	smarnet (single model) *Eigen Technology & Zhejiang University* https://arxiv.org/abs/1710.02772 (https://arxiv.org/abs/1710.02772)	71.415	80.160
58 Apr 13, 2017	QFASE *NUS*	71.898	79.989
59 Oct 27, 2017	M-NET (single) *UFL*	71.016	79.835
59 Apr 22, 2018	MAMCN (single model) *Samsung Research*	70.985	79.939
59 Jul 14, 2017	Mnemonic Reader (single model) *NUDT and Fudan University* https://arxiv.org/abs/1705.02798 (https://arxiv.org/abs/1705.02798)	70.995	80.146
59 May 23, 2018	AttReader (single) *College of Computer & Information Science, SouthWest University, Chongqing, China*	71.373	79.725
59 Mar 24, 2017	jNet (single model) *USTC & National Research Council Canada & York University* https://arxiv.org/abs/1703.04617 (https://arxiv.org/abs/1703.04617)	70.607	79.821
59 Apr 02, 2017	Ruminating Reader (single model) *New York University* https://arxiv.org/abs/1704.07415 (https://arxiv.org/abs/1704.07415)	70.639	79.456
59 Mar 14, 2017	Document Reader (single model) *Facebook AI Research* https://arxiv.org/abs/1704.00051 (https://arxiv.org/abs/1704.00051)	70.733	79.353

59 Dec 28, 2016	FastQAExt *German Research Center for Artificial Intelligence* https://arxiv.org/abs/1703.04816 (https://arxiv.org/abs/1703.04816)	70.849	78.857
59 May 13, 2017	RaSoR (single model) *Google NY, Tel-Aviv University* https://arxiv.org/abs/1611.01436 (https://arxiv.org/abs/1611.01436)	70.849	78.741
59 Mar 08, 2017	ReasoNet (single model) *MSR Redmond* https://arxiv.org/abs/1609.05284 (https://arxiv.org/abs/1609.05284)	70.555	79.364
60 Apr 14, 2017	Multi-Perspective Matching (single model) *IBM Research* https://arxiv.org/abs/1612.04211 (https://arxiv.org/abs/1612.04211)	70.387	78.784
61 Aug 30, 2017	SimpleBaseline (single model) *Technical University of Vienna*	69.600	78.236
61 Feb 05, 2018	SSR-BiDAF *single model*	69.443	78.358
62 Apr 12, 2017	SEDT+BiDAF (single model) *CMU* https://arxiv.org/abs/1703.00572 (https://arxiv.org/abs/1703.00572)	68.478	77.971
63 Jun 25, 2017	PQMN (single model) *KAIST & AIBrain & Crosscert*	68.331	77.783
64 Apr 12, 2017	T-gating (single model) *Peking University*	68.132	77.569
65 Nov 28, 2016	BiDAF (single model) *Allen Institute for AI & University of Washington* https://arxiv.org/abs/1611.01603 (https://arxiv.org/abs/1611.01603)	67.974	77.323
65 Feb 22, 2018	Unnamed submission by null	68.478	77.220
66 Feb 22, 2018	Unnamed submission by null	68.425	77.077

66 Dec 28, 2016	FastQA *German Research Center for Artificial Intelligence* https://arxiv.org/abs/1703.04816 (https://arxiv.org/abs/1703.04816)	68.436	77.070
66 Jul 29, 2017	SEDT (single model) *CMU* https://arxiv.org/abs/1703.00572 (https://arxiv.org/abs/1703.00572)	68.163	77.527
67 Oct 26, 2016	Match-LSTM with Ans-Ptr (Boundary) (ensemble) *Singapore Management University* https://arxiv.org/abs/1608.07905 (https://arxiv.org/abs/1608.07905)	67.901	77.022
67 Jan 22, 2018	FABIR *Single Model* https://arxiv.org/abs/1810.09580 (https://arxiv.org/abs/1810.09580)	67.744	77.605
68 Sep 19, 2017	AllenNLP BiDAF (single model) *Allen Institute for AI* http://allennlp.org/ (http://allennlp.org/)	67.618	77.151
69 Feb 05, 2017	Iterative Co-attention Network *Fudan University*	67.502	76.786
70 Jan 03, 2018	newtest *single model*	66.527	75.787
70 Nov 01, 2016	Dynamic Coattention Networks (single model) *Salesforce Research* https://arxiv.org/abs/1611.01604 (https://arxiv.org/abs/1611.01604)	66.233	75.896
71 Feb 24, 2018	Unnamed submission by null	65.992	75.469
72 Jan 10, 2018	Unnamed submission by null	64.796	74.272
73 Dec 09, 2017	Unnamed submission by ravioncodalab	64.439	73.921
73 Oct 26, 2016	Match-LSTM with Bi-Ans-Ptr (Boundary) *Singapore Management University* https://arxiv.org/abs/1608.07905 (https://arxiv.org/abs/1608.07905)	64.744	73.743

74 Feb 19, 2017	Attentive CNN context with LSTM *NLPR, CASIA*	63.306	73.463
75 Nov 02, 2016	Fine-Grained Gating *Carnegie Mellon University* https://arxiv.org/abs/1611.01724 (https://arxiv.org/abs/1611.01724)	62.446	73.327
75 Sep 21, 2017	OTF dict+spelling (single) *University of Montreal* https://arxiv.org/abs/1706.00286 (https://arxiv.org/abs/1706.00286)	64.083	73.056
76 Sep 21, 2017	OTF spelling (single) *University of Montreal* https://arxiv.org/abs/1706.00286 (https://arxiv.org/abs/1706.00286)	62.897	72.016
77 Sep 21, 2017	OTF spelling+lemma (single) *University of Montreal* https://arxiv.org/abs/1706.00286 (https://arxiv.org/abs/1706.00286)	62.604	71.968
78 Sep 28, 2016	Dynamic Chunk Reader *IBM* https://arxiv.org/abs/1610.09996 (https://arxiv.org/abs/1610.09996)	62.499	70.956
79 Aug 27, 2016	Match-LSTM with Ans-Ptr (Boundary) *Singapore Management University* https://arxiv.org/abs/1608.07905 (https://arxiv.org/abs/1608.07905)	60.474	70.695
80 Sep 18, 2018	Unnamed submission by null	59.058	69.436
81 Jan 10, 2018	Unnamed submission by null	58.764	69.276
82 Aug 27, 2016	Match-LSTM with Ans-Ptr (Sentence) *Singapore Management University* https://arxiv.org/abs/1608.07905 (https://arxiv.org/abs/1608.07905)	54.505	67.748
83 Nov 14, 2018	Unnamed submission by jinhyuklee	52.544	62.780

84 Oct 26, 2018	Unnamed submission by minjoon	52.533	62.757