

SQuAD 2.0

The Stanford Question Answering Dataset

What is SQuAD?

Stanford **Question Answering Dataset** (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.

(/SQuAD-explorer/explore/v2.0/dev/)

(<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.

(/SQuAD-explorer/explore/1.1/dev/)

(<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 \(40 MB\) \(/SQuAD-explorer/dataset/train-v2.0.json\)](/SQuAD-explorer/dataset/train-v2.0.json)

[Dev Set v2.0 \(4 MB\) \(/SQuAD-explorer/dataset/dev-v2.0.json\)](/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/cont)
(<https://worksheets.codalab.org/rest/bundles/0x6b567e1cf2e041ec80d7098f031c5c9e/cont>)

Sample Prediction File (on Dev v2.0)

(<https://worksheets.codalab.org/bundles/0x8731effab84f41b7b874a070e40f61e2/>)

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/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:

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>).

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 <i>Stanford University</i> (Rajpurkar & Jia et al. '18) (http://arxiv.org/abs/1606.05250)	86.831	89.452
1 Mar 20, 2019	BERT + DAE + AoA (ensemble) <i>Joint Laboratory of HIT and iFLYTEK Research</i>	87.147	89.474
2 Mar 15, 2019	BERT + ConvLSTM + MTL + Verifier (ensemble) <i>Layer 6 AI</i>	86.730	89.286
3 Mar 05, 2019	BERT + N-Gram Masking + Synthetic Self-Training (ensemble) <i>Google AI Language</i> https://github.com/google-research/bert (https://github.com/google-research/bert)	86.673	89.147
4 Mar 16, 2019	BERT + DAE + AoA (single model) <i>Joint Laboratory of HIT and iFLYTEK Research</i>	85.884	88.621
5 Jan 15, 2019	BERT + MMFT + ADA (ensemble) <i>Microsoft Research Asia</i>	85.082	87.615

5 Mar 13, 2019	BERT + ConvLSTM + MTL + Verifier (single model) <i>Layer 6 AI</i>	84.924	88.204
5 Mar 05, 2019	BERT + N-Gram Masking + Synthetic Self-Training (single model) <i>Google AI Language</i> https://github.com/google-research/bert (https://github.com/google-research/bert)	85.150	87.715
6 Jan 10, 2019	BERT + Synthetic Self-Training (ensemble) <i>Google AI Language</i> https://github.com/google-research/bert (https://github.com/google-research/bert)	84.292	86.967
7 Dec 13, 2018	BERT finetune baseline (ensemble) <i>Anonymous</i>	83.536	86.096
7 Mar 20, 2019	Bert-raw (ensemble) <i>None</i>	83.604	86.036
7 Dec 21, 2018	PAML+BERT (ensemble model) <i>PINGAN GammaLab</i>	83.457	86.122
7 Dec 16, 2018	Lunet + Verifier + BERT (ensemble) <i>Layer 6 AI NLP Team</i>	83.469	86.043
8 Mar 04, 2019	SemBERT (ensemble model) <i>Shanghai Jiao Tong University</i>	83.243	85.821
8 Dec 15, 2018	Lunet + Verifier + BERT (single model) <i>Layer 6 AI NLP Team</i>	82.995	86.035
8 Jan 14, 2019	BERT + MMFT + ADA (single model) <i>Microsoft Research Asia</i>	83.040	85.892
9 Feb 15, 2019	BERT + NeurQuRI (ensemble) <i>2SAH</i>	82.803	85.703
9 Feb 16, 2019	Bert-raw (ensemble) <i>None</i>	83.175	85.635
10 Dec 16, 2018	PAML+BERT (single model) <i>PINGAN GammaLab</i>	82.577	85.603
10 Feb 28, 2019	BERT + NeurQuRI (ensemble) <i>2SAH</i>	82.713	85.584
10 Mar 11, 2019	Bert-raw (ensemble) <i>None</i>	83.119	85.510
10 Jan 10, 2019	BERT + Synthetic Self-Training (single model) <i>Google AI Language</i> https://github.com/google-research/bert (https://github.com/google-research/bert)	82.972	85.810
10 Feb 26, 2019	BERT with Something (ensemble) <i>Anonymous</i>	83.051	85.737
11 Mar 02, 2019	SemBERT (single model) <i>Shanghai Jiao Tong University</i>	82.431	85.178

11 Nov 16, 2018	AoA + DA + BERT (ensemble) <i>Joint Laboratory of HIT and iFLYTEK Research</i>	82.374	85.310
12 Dec 12, 2018	BERT finetune baseline (single model) <i>Anonymous</i>	82.126	84.820
12 Feb 28, 2019	BERT_s (single model) <i>Anonymous</i>	81.979	84.846
12 Dec 11, 2018	Candi-Net+BERT (ensemble) <i>42Maru NLP Team</i>	82.126	84.624
13 Feb 28, 2019	BERT-large+UBFT (single model) <i>anonymous</i>	81.573	84.535
14 Feb 25, 2019	BERT with Something (single model) <i>Anonymous</i>	81.110	84.386
14 Feb 15, 2019	BERT + NeurQuRI (single model) <i>2SAH</i>	81.257	84.342
15 Nov 16, 2018	AoA + DA + BERT (single model) <i>Joint Laboratory of HIT and iFLYTEK Research</i>	81.178	84.251
16 Mar 20, 2019	Bert-raw (single) <i>None</i>	80.693	83.922
16 Mar 07, 2019	BERT + UnAnsQ (single model) <i>Anonymous</i>	80.749	83.851
17 Dec 19, 2018	Candi-Net+BERT (single model) <i>42Maru NLP Team</i>	80.659	83.562
18 Jan 09, 2019	Unnamed submission by null	80.512	83.539
19 Mar 12, 2019	Bert-raw (single) <i>None</i>	80.411	83.457
19 Jan 22, 2019	BERT + NeurQuRI (single model) <i>2SAH</i>	80.591	83.391
20 Mar 09, 2019	Insight-baseline (single model) <i>PAll Insight Team</i>	80.264	83.376
20 Feb 16, 2019	Bert-raw (single model) <i>None</i>	80.343	83.243
20 Jan 09, 2019	Unnamed submission by null	80.343	83.221
21 Feb 19, 2019	BERT + UDA (single model) <i>Anonymous</i>	80.005	83.208
21 Dec 03, 2018	PwP+BERT (single model) <i>AITRICS</i>	80.117	83.189
22 Nov 09, 2018	BERT (single model) <i>Google AI Language</i>	80.005	83.061
	BERT + Sparse-Transformer	79.948	83.023

23	<i>single model</i>		
Feb 12, 2019			
24	BERT uncased (single model)	79.745	83.020
Mar 07, 2019	<i>Anonymous</i>		
24	ST_bl	80.140	82.962
Feb 28, 2019	<i>single model</i>		
25	NEXYS_BASE (single model)	79.779	82.912
Dec 06, 2018	<i>NEXYS, DGIST R7</i>		
25	Candi-Net+BERT (single model)	80.388	82.908
Dec 05, 2018	<i>42Maru NLP Team</i>		
26	{bert-finetuning} (single model)	79.632	82.852
Feb 02, 2019	<i>ksai</i>		
27	{Anonymous} (single model)	78.876	82.524
Mar 14, 2019	<i>Anonymous</i>		
27	L6Net + BERT (single model)	79.181	82.259
Nov 09, 2018	<i>Layer 6 AI</i>		
28	BISAN (single model)	78.481	81.531
Mar 14, 2019	<i>Seoul National University & Hyundai Motors</i>		
29	Unnamed submission by null	78.301	81.350
Jan 09, 2019			
30	BERT+AC(single model)	78.052	81.174
Dec 14, 2018	<i>Hithink RoyalFlush</i>		
31	SLQA+BERT (single model)	77.003	80.209
Nov 06, 2018	<i>Alibaba DAMO NLP</i> http://www.aclweb.org/anthology/P18-1158 (http://www.aclweb.org/anthology/P18-1158)		
32	synss (single model)	76.055	79.329
Jan 05, 2019	<i>bert_finetune</i>		
33	ARSG-BERT (single model)	74.746	78.227
Dec 19, 2018	<i>TRINITY RESEARCH LABS, Active.ai</i> https://active.ai (https://active.ai)		
33	MIR-MRC(F-Net) (single model)	74.791	77.988
Nov 05, 2018	<i>Kangwon National University, Natural Language Processing Lab. & ForceWin, KP Lab.</i>		
34	nlnet (single model)	74.272	77.052
Sep 13, 2018	<i>Microsoft Research Asia</i>		
35	Unnamed submission by null	73.234	76.790
Dec 22, 2018			
35	MMIPN	73.505	76.424
Dec 29, 2018	<i>Single</i>		
36	YARCS (ensemble)	72.670	75.507
Oct 12, 2018	<i>IBM Research AI</i>		

37 Nov 14, 2018	BERT+Answer Verifier (single model) <i>Pingan Tech Olatop Lab</i>	71.666	75.457
37 Nov 10, 2018	Unnamed submission by null	72.580	75.075
38 Sep 17, 2018	Unet (ensemble) <i>Fudan University & Liulishuo Lab</i> https://arxiv.org/abs/1810.06638 (https://arxiv.org/abs/1810.06638)	71.417	74.869
38 Aug 28, 2018	SLQA+ (single model) <i>Alibaba DAMO NLP</i> http://www.aclweb.org/anthology/P18-1158 (http://www.aclweb.org/anthology/P18-1158)	71.462	74.434
38 Aug 15, 2018	Reinforced Mnemonic Reader + Answer Verifier (single model) <i>NUDT</i> https://arxiv.org/abs/1808.05759 (https://arxiv.org/abs/1808.05759)	71.767	74.295
38 Jan 19, 2019	{BERT-base} (single-model) <i>Anonymous</i>	70.763	74.449
38 Sep 14, 2018	SAN (ensemble model) <i>Microsoft Business Applications AI Research</i> https://arxiv.org/abs/1712.03556 (https://arxiv.org/abs/1712.03556)	71.316	73.704
39 Nov 10, 2018	Unnamed submission by null	70.718	73.403
40 Sep 14, 2018	Unet (single model) <i>Fudan University & Liulishuo Lab</i>	69.262	72.642
40 Aug 21, 2018	FusionNet++ (ensemble) <i>Microsoft Business Applications Group AI Research</i> https://arxiv.org/abs/1711.07341 (https://arxiv.org/abs/1711.07341)	70.300	72.484
40 Sep 26, 2018	Multi-Level Attention Fusion(MLAF) (single model) <i>Chonbuk National University, Cognitive Computing Lab.</i>	69.476	72.857
41 Dec 20, 2018	DocQA + NeurQuRI (single model) <i>2SAH</i>	68.766	71.662
42 Sep 13, 2018	BiDAF++ with pair2vec (single model) <i>UW and FAIR</i>	68.021	71.583
42 Aug 21, 2018	SAN (single model) <i>Microsoft Business Applications AI Research</i> https://arxiv.org/abs/1712.03556 (https://arxiv.org/abs/1712.03556)	68.653	71.439
42 Nov 10, 2018	Unnamed submission by null	68.653	71.124
43 Jul 13, 2018	VS^3-NET (single model) <i>Kangwon National University in South Korea</i>	67.897	70.884

43 Jun 25, 2018	KACTEIL-MRC(GFN-Net) (single model) <i>Kangwon National University, Natural Language Processing Lab.</i>	68.213	70.878
44 Jan 02, 2019	EBB-Net (single model) <i>Enliple AI</i>	66.610	70.303
45 Jun 25, 2018	KakaoNet2 (single model) <i>Kakao NLP Team</i>	65.719	69.381
46 Jul 11, 2018	abcNet (single model) <i>Fudan University & Liulishuo AI Lab</i>	65.256	69.206
46 Sep 13, 2018	BiDAF++ (single model) <i>UW and FAIR</i>	65.651	68.866
47 Jun 27, 2018	BSAE AddText (single model) <i>reciTAL.ai</i>	63.338	67.422
48 Aug 14, 2018	eeAttNet (single model) <i>BBD NLP Team</i> https://www.bbdservice.com (https://www.bbdservice.com)	63.327	66.633
48 May 30, 2018	BiDAF + Self Attention + ELMo (single model) <i>Allen Institute for Artificial Intelligence [modified by Stanford]</i>	63.372	66.251
49 Nov 27, 2018	Tree-LSTM + BiDAF + ELMo (single model) <i>Carnegie Mellon University</i>	57.707	62.341
49 May 30, 2018	BiDAF + Self Attention (single model) <i>Allen Institute for Artificial Intelligence [modified by Stanford]</i>	59.332	62.305
50 May 30, 2018	BiDAF-No-Answer (single model) <i>University of Washington [modified by Stanford]</i>	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 <i>Stanford University</i> (Rajpurkar et al. '16) (http://arxiv.org/abs/1606.05250)	82.304	91.221
1 Oct 05, 2018	BERT (ensemble) <i>Google AI Language</i> https://arxiv.org/abs/1810.04805 (https://arxiv.org/abs/1810.04805)	87.433	93.160
2 Feb 14, 2019	Knowledge-enhanced BERT (single model) <i>Anonymous</i>	85.944	92.425
2 Sep 27, 2018	nlnet (ensemble) <i>Microsoft Research Asia</i>	85.954	91.677

3 Sep 09, 2018	nlnet (ensemble) <i>Microsoft Research Asia</i>	85.356	91.202
3 Oct 05, 2018	BERT (single model) <i>Google AI Language</i> https://arxiv.org/abs/1810.04805 (https://arxiv.org/abs/1810.04805)	85.083	91.835
4 Feb 19, 2019	WD (single model) <i>Anonymous</i>	84.402	90.561
4 Jul 11, 2018	QANet (ensemble) <i>Google Brain & CMU</i>	84.454	90.490
5 Feb 21, 2019	WD1 (single model) <i>Anonymous</i>	83.804	90.429
5 Jul 08, 2018	r-net (ensemble) <i>Microsoft Research Asia</i>	84.003	90.147
6 Jun 20, 2018	MARS (ensemble) <i>YUANFUDAO research NLP</i>	83.982	89.796
6 Sep 09, 2018	nlnet (single model) <i>Microsoft Research Asia</i>	83.468	90.133
6 Mar 19, 2018	QANet (ensemble) <i>Google Brain & CMU</i>	83.877	89.737
7 Sep 01, 2018	MARS (single model) <i>YUANFUDAO research NLP</i>	83.185	89.547
8 Jun 21, 2018	MARS (single model) <i>YUANFUDAO research NLP</i>	83.122	89.224
8 Jan 22, 2018	Hybrid AoA Reader (ensemble) <i>Joint Laboratory of HIT and iFLYTEK Research</i>	82.482	89.281
8 May 09, 2018	MARS (single model) <i>YUANFUDAO research NLP</i>	82.587	88.880
8 Feb 19, 2018	Reinforced Mnemonic Reader + A2D (ensemble model) <i>Microsoft Research Asia & NUDT</i>	82.849	88.764
8 Jun 20, 2018	QANet (single) <i>Google Brain & CMU</i>	82.471	89.306
8 Mar 06, 2018	QANet (ensemble) <i>Google Brain & CMU</i>	82.744	89.045
9 Jan 05, 2018	SLQA+ (ensemble) <i>Alibaba iDST NLP</i>	82.440	88.607
9 Dec 17, 2018	ARSG-BERT (single model) <i>TRINITY RESEARCH LABS, Active.ai</i> https://active.ai (https://active.ai)	81.307	88.909
9 Feb 02, 2018	Reinforced Mnemonic Reader (ensemble model) <i>NUDT and Fudan University</i> https://arxiv.org/abs/1705.02798	82.283	88.533

9 Jan 03, 2018	r-net+ (ensemble) <i>Microsoft Research Asia</i>	82.650	88.493
9 Dec 24, 2018	MMIPN <i>Single</i>	81.580	88.948
9 Feb 28, 2018	QANet (single model) <i>Google Brain & CMU</i>	82.209	88.608
10 May 09, 2018	Reinforced Mnemonic Reader + A2D (single model) <i>Microsoft Research Asia & NUDT</i>	81.538	88.130
10 Apr 23, 2018	r-net (single model) <i>Microsoft Research Asia</i>	81.391	88.170
10 May 09, 2018	Reinforced Mnemonic Reader + A2D + DA (single model) <i>Microsoft Research Asia & NUDT</i>	81.401	88.122
10 Dec 22, 2017	AttentionReader+ (ensemble) <i>Tencent DPDAC NLP</i>	81.790	88.163
11 Apr 03, 2018	KACTEIL-MRC(GF-Net+) (ensemble) <i>Kangwon National University, Natural Language Processing Lab.</i>	81.496	87.557
11 Dec 17, 2017	r-net (ensemble) <i>Microsoft Research Asia</i> http://aka.ms/rnet (http://aka.ms/rnet)	82.136	88.126
12 Feb 27, 2018	QANet (single model) <i>Google Brain & CMU</i>	80.929	87.773
13 Feb 19, 2018	Reinforced Mnemonic Reader + A2D (single model) <i>Microsoft Research Asia & NUDT</i>	80.919	87.492
14 Feb 12, 2018	Reinforced Mnemonic Reader + A2D (single model) <i>Microsoft Research Asia & NUDT</i>	80.489	87.454
14 Nov 17, 2017	BiDAF + Self Attention + ELMo (ensemble) <i>Allen Institute for Artificial Intelligence</i>	81.003	87.432
15 Apr 12, 2018	AVIQA+ (ensemble) <i>aviqa team</i>	80.615	87.311
16 Jan 13, 2018	SLQA+ <i>single model</i>	80.436	87.021
17 Jan 12, 2018	EAZI+ (ensemble) <i>Yiwise NLP Group</i>	80.426	86.912
17 Jan 04, 2018	{EAZI} (ensemble) <i>Yiwise NLP Group</i>	80.436	86.912
18 Feb 23, 2018	MAMCN+ (single model) <i>Samsung Research</i>	79.692	86.727
18 Jan 22, 2018	Hybrid AoA Reader (single model) <i>Joint Laboratory of HIT and iFLYTEK Research</i>	80.027	87.288

18 Mar 20, 2018	DNET (ensemble) <i>QA geeks</i>	80.164	86.721
19 Feb 13, 2018	BiDAF + Self Attention + ELMo + A2D (single model) <i>Microsoft Research Asia & NUDT</i>	79.996	86.711
20 Jan 29, 2018	Reinforced Mnemonic Reader (single model) <i>NUDT and Fudan University</i> https://arxiv.org/abs/1705.02798 (https://arxiv.org/abs/1705.02798)	79.545	86.654
20 Apr 10, 2018	Unnamed submission by null	80.027	86.612
21 Jan 03, 2018	r-net+ (single model) <i>Microsoft Research Asia</i>	79.901	86.536
22 Dec 05, 2017	SAN (ensemble model) <i>Microsoft Business AI Solutions Team</i> https://arxiv.org/abs/1712.03556 (https://arxiv.org/abs/1712.03556)	79.608	86.496
22 Dec 28, 2017	SLQA+ (single model) <i>Alibaba iDST NLP</i>	79.199	86.590
23 Oct 18, 2017	Interactive AoA Reader+ (ensemble) <i>Joint Laboratory of HIT and iFLYTEK</i>	79.083	86.450
23 Nov 05, 2018	MIR-MRC(F-Net) (single model) <i>ForceWin, KP Lab.</i>	79.083	86.288
24 Jun 02, 2018	MDReader <i>single model</i>	79.031	86.006
24 Feb 01, 2018	Unnamed submission by null	78.999	86.151
25 Oct 24, 2017	FusionNet (ensemble) <i>Microsoft Business AI Solutions Team</i> https://arxiv.org/abs/1711.07341 (https://arxiv.org/abs/1711.07341)	78.978	86.016
26 Oct 22, 2017	DCN+ (ensemble) <i>Salesforce Research</i> https://arxiv.org/abs/1711.00106 (https://arxiv.org/abs/1711.00106)	78.852	85.996
27 Nov 03, 2017	BiDAF + Self Attention + ELMo (single model) <i>Allen Institute for Artificial Intelligence</i>	78.580	85.833
27 Mar 30, 2018	KACTEIL-MRC(GF-Net+) (single model) <i>Kangwon National University, Natural Language Processing Lab.</i>	78.664	85.780
28 May 10, 2018	KakaoNet (single model) <i>Kakao NLP Team</i>	78.401	85.724
28 Mar 19, 2018	aviqa (ensemble) <i>aviqa team</i>	78.496	85.469
28	Conductor-net (ensemble)	78.433	85.517

Jan 02, 2018	CMU https://arxiv.org/abs/1710.10504 (https://arxiv.org/abs/1710.10504)		
28 Nov 30, 2017	SLQA(ensemble) <i>Alibaba iDST NLP</i>	78.328	85.682
29 Sep 18, 2018	BiDAF++ with pair2vec (single model) <i>UW and FAIR</i>	78.223	85.535
29 Jun 01, 2018	MDReader0 <i>single model</i>	78.171	85.543
29 Jan 03, 2018	MEMEN (single model) <i>Zhejiang University</i> https://arxiv.org/abs/1707.09098 (https://arxiv.org/abs/1707.09098)	78.234	85.344
29 Jan 29, 2018	test <i>single</i>	78.087	85.348
30 Jul 26, 2017	Interactive AoA Reader (ensemble) <i>Joint Laboratory of HIT and iFLYTEK Research</i>	77.845	85.297
31 Jan 11, 2018	Unnamed submission by null	77.436	85.130
31 Mar 20, 2018	DNET (single model) <i>QA geeks</i>	77.646	84.905
32 Sep 18, 2018	BiDAF++ (single model) <i>UW and FAIR</i>	77.573	84.858
32 Dec 14, 2017	RaSoR + TR + LM (single model) <i>Tel-Aviv University</i> https://arxiv.org/abs/1712.03609 (https://arxiv.org/abs/1712.03609)	77.583	84.163
32 Apr 10, 2018	Unnamed submission by null	77.489	84.735
32 Dec 06, 2017	AttentionReader+ (single) <i>Tencent DPDAC NLP</i>	77.342	84.925
33 Nov 06, 2017	Conductor-net (ensemble) <i>CMU</i> https://arxiv.org/abs/1710.10504 (https://arxiv.org/abs/1710.10504)	76.996	84.630
34 Dec 19, 2017	FRC (single model) <i>in review</i>	76.240	84.599
34 Dec 21, 2017	Jenga (ensemble) <i>Facebook AI Research</i>	77.237	84.466
34 Jan 23, 2018	MARS (single model) <i>YUANFUDAO research NLP</i>	76.859	84.739
35 Nov 01, 2017	SAN (single model) <i>Microsoft Business AI Solutions Team</i>	76.828	84.396

<https://arxiv.org/abs/1712.03556>
(<https://arxiv.org/abs/1712.03556>)

36 Oct 13, 2017	r-net (single model) <i>Microsoft Research Asia</i> http://aka.ms/rnet (http://aka.ms/rnet)	76.461	84.265
37 Oct 22, 2017	Conductor-net (ensemble) <i>CMU</i>	76.146	83.991
37 Sep 26, 2018	{gqa} (single model) <i>FAIR</i>	77.090	83.931
37 May 14, 2018	VS ³ -NET (single model) <i>Kangwon National University in South Korea</i>	76.775	84.491
38 Sep 08, 2017	FusionNet (single model) <i>Microsoft Business AI Solutions team</i> https://arxiv.org/abs/1711.07341 (https://arxiv.org/abs/1711.07341)	75.968	83.900
39 Oct 22, 2017	Interactive AoA Reader+ (single model) <i>Joint Laboratory of HIT and iFLYTEK</i>	75.821	83.843
39 Oct 18, 2018	KAR (single model) <i>York University</i> https://arxiv.org/abs/1809.03449 (https://arxiv.org/abs/1809.03449)	76.125	83.538
40 Jul 14, 2017	smarnet (ensemble) <i>Eigen Technology & Zhejiang University</i>	75.989	83.475
41 Mar 15, 2018	AVIQA-v2 (single model) <i>aviqa team</i>	75.926	83.305
42 Aug 18, 2017	RaSoR + TR (single model) <i>Tel-Aviv University</i> https://arxiv.org/abs/1712.03609 (https://arxiv.org/abs/1712.03609)	75.789	83.261
43 Oct 23, 2017	DCN+ (single model) <i>Salesforce Research</i> https://arxiv.org/abs/1711.00106 (https://arxiv.org/abs/1711.00106)	75.087	83.081
43 Oct 05, 2018	Unnamed submission by null	74.950	83.294
43 May 21, 2017	MEMEN (ensemble) <i>Eigen Technology & Zhejiang University</i> https://arxiv.org/abs/1707.09098 (https://arxiv.org/abs/1707.09098)	75.370	82.658
43 Nov 01, 2017	Mixed model (ensemble) <i>Sean</i>	75.265	82.769
44 Nov 17, 2017	two-attention-self-attention (ensemble) <i>guotong1988</i>	75.223	82.716
44	DCN+ (single model)	74.866	82.806

Jul 10, 2017	<i>Salesforce Research</i> https://arxiv.org/abs/1711.00106 (https://arxiv.org/abs/1711.00106)		
45 Jan 02, 2018	Conductor-net (single model) <i>CMU</i> https://arxiv.org/abs/1710.10504 (https://arxiv.org/abs/1710.10504)	74.405	82.742
45 Feb 06, 2018	Jenga (single model) <i>Facebook AI Research</i>	74.373	82.845
45 Aug 14, 2018	eeAttNet (single model) <i>BBD NLP Team</i> https://www.bbdservice.com (https://www.bbdservice.com)	74.604	82.501
46 Feb 13, 2018	SSR-BiDAF <i>ensemble model</i>	74.541	82.477
46 Oct 31, 2017	SLQA (single model) <i>Alibaba iDST NLP</i>	74.489	82.815
46 Mar 09, 2017	ReasonNet (ensemble) <i>MSR Redmond</i> https://arxiv.org/abs/1609.05284 (https://arxiv.org/abs/1609.05284)	75.034	82.552
47 Jul 14, 2017	Mnemonic Reader (ensemble) <i>NUDT and Fudan University</i> https://arxiv.org/abs/1705.02798 (https://arxiv.org/abs/1705.02798)	74.268	82.371
48 Dec 23, 2017	S ³ -Net (ensemble) <i>Kangwon National University in South Korea</i>	74.121	82.342
48 Oct 27, 2017	Unnamed submission by null	74.489	82.312
49 Jul 25, 2017	Interactive AoA Reader (single model) <i>Joint Laboratory of HIT and iFLYTEK Research</i>	73.639	81.931
49 Jul 29, 2017	SEDT (ensemble model) <i>CMU</i> https://arxiv.org/abs/1703.00572 (https://arxiv.org/abs/1703.00572)	74.090	81.761
49 Nov 06, 2017	Conductor-net (single) <i>CMU</i> https://arxiv.org/abs/1710.10504 (https://arxiv.org/abs/1710.10504)	73.240	81.933
49 Jul 06, 2017	SSAE (ensemble) <i>Tsinghua University</i>	74.080	81.665
49 Dec 14, 2017	Jenga (single model) <i>Facebook AI Research</i>	73.303	81.754
49 Feb 22, 2017	BiDAF (ensemble) <i>Allen Institute for AI & University of Washington</i>	73.744	81.525

<https://arxiv.org/abs/1611.01603>
(<https://arxiv.org/abs/1611.01603>)

49 Apr 22, 2017	SEDT+BiDAF (ensemble) <i>CMU</i> https://arxiv.org/abs/1703.00572 (https://arxiv.org/abs/1703.00572)	73.723	81.530
49 Jan 24, 2017	Multi-Perspective Matching (ensemble) <i>IBM Research</i> https://arxiv.org/abs/1612.04211 (https://arxiv.org/abs/1612.04211)	73.765	81.257
49 May 01, 2017	jNet (ensemble) <i>USTC & National Research Council Canada & York University</i> https://arxiv.org/abs/1703.04617 (https://arxiv.org/abs/1703.04617)	73.010	81.517
50 Apr 18, 2018	Unnamed submission by null	72.831	80.622
50 Apr 18, 2018	Unnamed submission by null	72.831	80.622
50 Nov 16, 2017	two-attention-self-attention (single model) <i>guotong1988</i>	72.600	81.011
50 Oct 22, 2017	Conductor-net (single) <i>CMU</i>	72.590	81.415
51 Sep 20, 2017	BiDAF + Self Attention (single model) <i>Allen Institute for Artificial Intelligence</i> https://arxiv.org/abs/1710.10723 (https://arxiv.org/abs/1710.10723)	72.139	81.048
52 Dec 15, 2017	S ³ -Net (single model) <i>Kangwon National University in South Korea</i>	71.908	81.023
52 Apr 12, 2017	T-gating (ensemble) <i>Peking University</i>	72.758	81.001
53 Mar 03, 2018	AVIQA (single model) <i>aviqa team</i>	72.485	80.550
54 Nov 06, 2017	attention+self-attention (single model) <i>guotong1988</i>	71.698	80.462
55 Nov 02, 2016	Dynamic Coattention Networks (ensemble) <i>Salesforce Research</i> https://arxiv.org/abs/1611.01604 (https://arxiv.org/abs/1611.01604)	71.625	80.383
56 Jul 14, 2017	smarnet (single model) <i>Eigen Technology & Zhejiang University</i> https://arxiv.org/abs/1710.02772 (https://arxiv.org/abs/1710.02772)	71.415	80.160
56 Apr 13, 2017	QFASE <i>NUS</i>	71.898	79.989

57 Oct 27, 2017	M-NET (single) UFL	71.016	79.835
57 Apr 22, 2018	MAMCN (single model) Samsung Research	70.985	79.939
57 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
57 May 23, 2018	AttReader (single) College of Computer & Information Science, SouthWest University, Chongqing, China	71.373	79.725
57 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
57 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
57 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
57 Dec 29, 2016	FastQAExt German Research Center for Artificial Intelligence https://arxiv.org/abs/1703.04816 (https://arxiv.org/abs/1703.04816)	70.849	78.857
57 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
57 Mar 08, 2017	ReasoNet (single model) MSR Redmond https://arxiv.org/abs/1609.05284 (https://arxiv.org/abs/1609.05284)	70.555	79.364
58 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
59 Aug 30, 2017	SimpleBaseline (single model) Technical University of Vienna	69.600	78.236
59 Feb 06, 2018	SSR-BiDAF single model	69.443	78.358
60 Apr 12, 2017	SEDT+BiDAF (single model) CMU https://arxiv.org/abs/1703.00572	68.478	77.971

61 Jun 25, 2017	PQMN (single model) <i>KAIST & AIBrain & Crosscert</i>	68.331	77.783
62 Apr 12, 2017	T-gating (single model) <i>Peking University</i>	68.132	77.569
63 Nov 28, 2016	BiDAF (single model) <i>Allen Institute for AI & University of Washington</i> https://arxiv.org/abs/1611.01603 (https://arxiv.org/abs/1611.01603)	67.974	77.323
63 Feb 22, 2018	Unnamed submission by null	68.478	77.220
64 Feb 22, 2018	Unnamed submission by null	68.425	77.077
64 Dec 29, 2016	FastQA <i>German Research Center for Artificial Intelligence</i> https://arxiv.org/abs/1703.04816 (https://arxiv.org/abs/1703.04816)	68.436	77.070
64 Jul 29, 2017	SEDT (single model) <i>CMU</i> https://arxiv.org/abs/1703.00572 (https://arxiv.org/abs/1703.00572)	68.163	77.527
65 Oct 26, 2016	Match-LSTM with Ans-Ptr (Boundary) (ensemble) <i>Singapore Management University</i> https://arxiv.org/abs/1608.07905 (https://arxiv.org/abs/1608.07905)	67.901	77.022
65 Jan 22, 2018	FABIR <i>Single Model</i> https://arxiv.org/abs/1810.09580 (https://arxiv.org/abs/1810.09580)	67.744	77.605
66 Sep 19, 2017	AllenNLP BiDAF (single model) <i>Allen Institute for AI</i> http://allennlp.org/ (http://allennlp.org/)	67.618	77.151
67 Feb 05, 2017	Iterative Co-attention Network <i>Fudan University</i>	67.502	76.786
68 Jan 03, 2018	newtest <i>single model</i>	66.527	75.787
68 Nov 02, 2016	Dynamic Coattention Networks (single model) <i>Salesforce Research</i> https://arxiv.org/abs/1611.01604 (https://arxiv.org/abs/1611.01604)	66.233	75.896
69 Feb 24, 2018	Unnamed submission by null	65.992	75.469
70 Jan 11, 2018	Unnamed submission by null	64.796	74.272

71 Dec 09, 2017	Unnamed submission by ravioncodalab	64.439	73.921
71 Oct 26, 2016	Match-LSTM with Bi-Ans-Ptr (Boundary) <i>Singapore Management University</i> https://arxiv.org/abs/1608.07905 (https://arxiv.org/abs/1608.07905)	64.744	73.743
72 Feb 19, 2017	Attentive CNN context with LSTM <i>NLPR, CASIA</i>	63.306	73.463
73 Nov 02, 2016	Fine-Grained Gating <i>Carnegie Mellon University</i> https://arxiv.org/abs/1611.01724 (https://arxiv.org/abs/1611.01724)	62.446	73.327
73 Sep 21, 2017	OTF dict+spelling (single) <i>University of Montreal</i> https://arxiv.org/abs/1706.00286 (https://arxiv.org/abs/1706.00286)	64.083	73.056
74 Sep 21, 2017	OTF spelling (single) <i>University of Montreal</i> https://arxiv.org/abs/1706.00286 (https://arxiv.org/abs/1706.00286)	62.897	72.016
75 Sep 21, 2017	OTF spelling+lemma (single) <i>University of Montreal</i> https://arxiv.org/abs/1706.00286 (https://arxiv.org/abs/1706.00286)	62.604	71.968
76 Sep 28, 2016	Dynamic Chunk Reader <i>IBM</i> https://arxiv.org/abs/1610.09996 (https://arxiv.org/abs/1610.09996)	62.499	70.956
77 Aug 27, 2016	Match-LSTM with Ans-Ptr (Boundary) <i>Singapore Management University</i> https://arxiv.org/abs/1608.07905 (https://arxiv.org/abs/1608.07905)	60.474	70.695
78 Sep 11, 2018	Unnamed submission by Will_Wu	59.058	69.436
79 Jan 11, 2018	Unnamed submission by null	58.764	69.276
80 Aug 27, 2016	Match-LSTM with Ans-Ptr (Sentence) <i>Singapore Management University</i> https://arxiv.org/abs/1608.07905 (https://arxiv.org/abs/1608.07905)	54.505	67.748
81 Nov 14, 2018	Unnamed submission by jinhyuklee	52.544	62.780
82 Oct 26, 2018	Unnamed submission by minjoon	52.533	62.757