

Training and Validation Data

Abbreviation	Model Name	Output Format	factual	counterfactual	empty	random context	disentanglement	File Name Folders: DisentQA Data , /more_baselines
(s) cb	Closed book	baselines			✓			v10-simplified_simplified-nq-train_closed_book_baseline_{}_split.csv.zip train , val
(s) factual	Single answer, Factual	baselines	✓					v10-simplified_simplified-nq-train_contextual_baseline_{}_split.csv.zip train , val
(s) f+cf	+Counterfactual	Single answer	✓	✓				more_baselines/v10-simplified_simplified-nq-train_factual_counterfactual_contextual_baseline_{}_split.csv.zip train , val
(s) f+cf	+Answerability	Single answer	✓		✓	✓		more_baselines/v10-simplified_simplified-nq-train_factual_answerability_baseline_{}_split.csv train , val
(s) f+a	+Counterfactual+ Answerability	Single answer	✓	✓	✓	✓		more_baselines/v10-simplified_simplified-nq-train_factual_counterfactual_answerability_contextual_baseline_{}_split.csv.zip train , val
(m) f+cf	+Counterfactual	Multi answer	✓	✓			✓	more_baselines/v10-simplified_simplified-nq-train_factual_counterfactual_disentangled_baseline_{}_split.csv.zip train , val
(m) f+a	+Answerability	Multi answer	✓		✓	✓	✓	more_baselines/v10-simplified_simplified-nq-train_factual_disentangled_answerability_baseline_{}_split.csv.zip train , val
(m) f+cf+a	+Counterfactual+ Answerability	Multi answer	✓	✓	✓	✓	✓	v10-simplified_simplified-nq-train_full_no_dec_{}_split.csv.zip train , val

Evaluation Set

[Evaluation sets/simplified-nq-dev_any_from_cf_full_subset_no_dec.csv.zip](#) (NQ Dev subset (1365*4))

Includes Factual, Counterfactual, Empty, Random altogether, in a multi answer output format (disentangled), that allows for an evaluation of the relevant part in the model's output (in a single answer output format the answer to evaluate against is the **contextual answer**).

Counts

	Factual	Counterfactual	Empty	Random
Train	85,540	30,653	85,540	85,540
Validation	21,386	7,698	21,386	21,386
Test (Evaluation Set)	1,365	1,365	1,365	1,365

References

Natural Questions (NQ) dataset:

Tom Kwiatkowski, Jennimaria Palomaki, Olivia Redfield, Michael Collins, Ankur Parikh, Chris Alberti, Danielle Epstein, Illia Polosukhin, Jacob Devlin, Kenton Lee, Kristina Toutanova, Llion Jones, Matthew Kelcey, Ming-Wei Chang, Andrew M. Dai, Jakob Uszkoreit, Quoc Le, and Slav Petrov. 2019. [Natural Questions: A Benchmark for Question Answering Research - ACL Anthology](#). Transactions of the Association for Computational Linguistics, 7:452–466.

Counterfactual Data Augmentation, obtained by automatically altering facts in a given QA corpus:

Shayne Longpre, Kartik Perisetla, Anthony Chen, Nikhil Ramesh, Chris DuBois, and Sameer Singh. 2021. [Entity-Based Knowledge Conflicts in Question Answering - ACL Anthology](#). In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pages 7052–7063, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.