

AutoQ

Improving reading comprehension through
automatic question generation

W210.5

Julia Buffinton, Saurav Datta, Joanna Huang, Kathryn Plath

1.5 Billion People Study English Worldwide

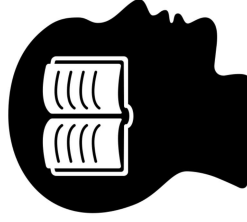
- Over **7.5 million** English language learners (ELLs) take English proficiency tests each year
- Current scores suggest room for improvement:
 - **TOEFL:** average reading scores are **intermediate**
 - **IELTS:** average reading scores are **competent**
 - However, many prestigious schools/companies require higher scores



The Core Issue: Limited Resources

- Current resources require significant financial/time investments or are limited in their scope
- No similar product offers free, comprehensive support for English language learners (ELLs)

AutoQ leverages ML and NLP to democratize reading comprehension questions



AutoQ

Improving reading comprehension through
automatic question generation

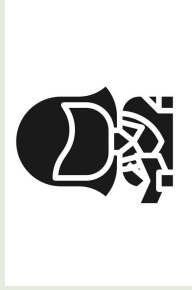
First **free** web app with automatically generated reading comprehension questions targeted to English language learners

Offers over **100,000 practice questions** on over **10,000 Wikipedia articles**, widely used source for English language practice

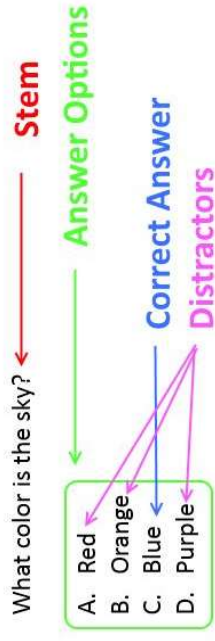
Multiple-choice questions mimic English-language exam format, **scoring system** provides feedback for improvement

The Traditional Question Generation Approach

Experts read source material and write questions and answers.



Experts identify main ideas and common sources of confusion to create wrong answer choices (“distractors”)

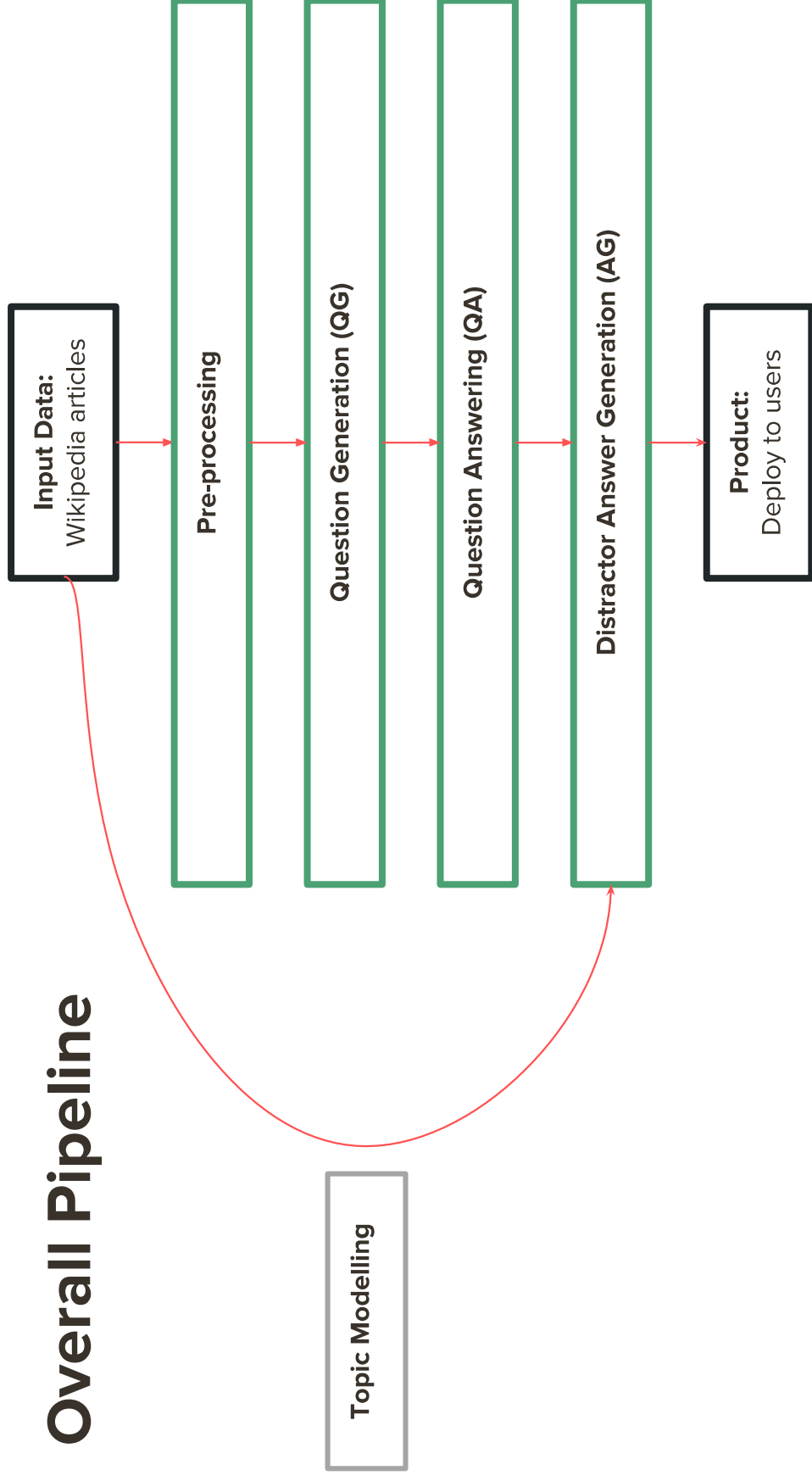


Learner reads the passage and attempts to answer the question.

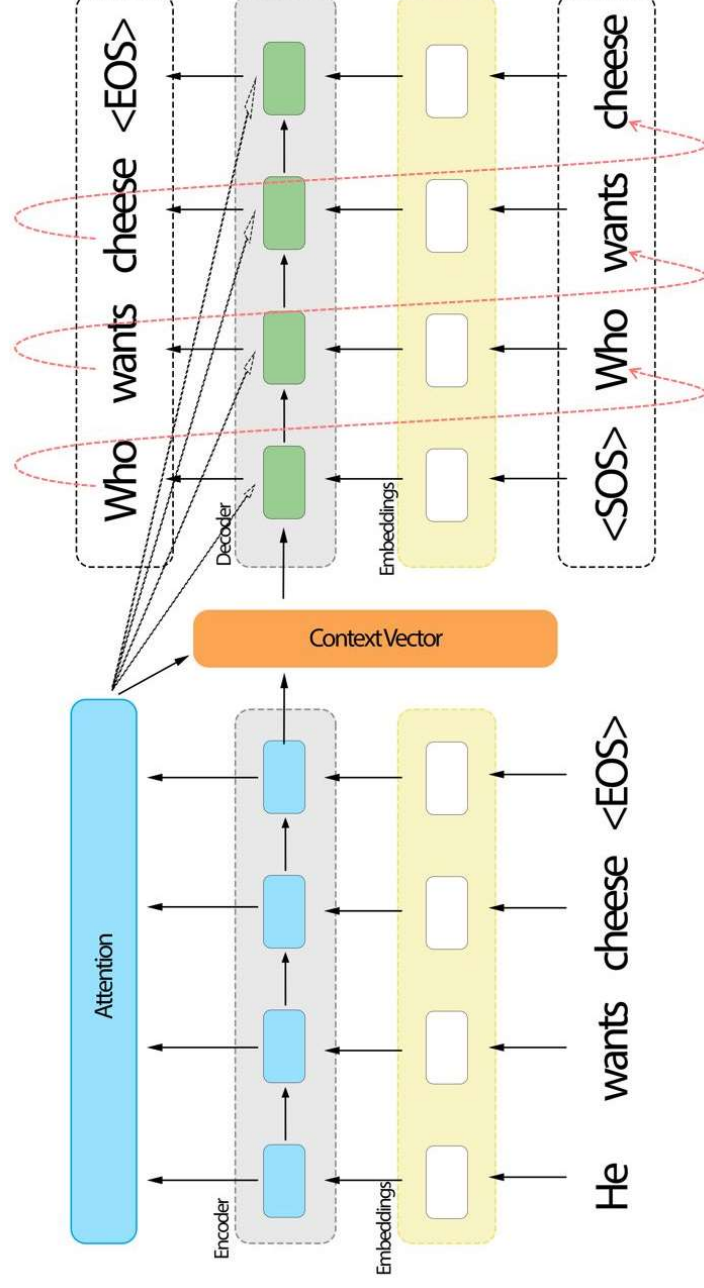


AutoQ alters this traditionally human-to-human function. By automating the question generation steps, we reduce the costs associated with exam prep to allow learners more access to educational opportunities.

Overall Pipeline



How it works: Question Generation (QG) Model



Try it!

Evaluation

Automatic

	QG	QA
Bleu_1	0.301	0.628
Bleu_2	0.185	0.568
Bleu_3	0.127	0.502
Bleu_4	0.089	0.430
ROUGE_L	0.280	0.679

Answer Example:

Reference: english-medium

Generated Question: Chinese

Human

	SQuAD Questions	Generated Questions	SQuAD Answers	Generated Answers
Clarity (1-5)	4.21	3.65	4.35	4.05
Fluency (1-5)	4.15	3.61	4.37	4.03
Grammatical	94%	82%	94%	85%
Answer Present	88%	63%	91%	91%

Question Example:

Reference: If Roman numerals were used, what would Super Bowl 50 have been called?

Generated Question: What is the “golden anniversary?”

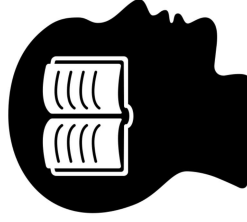
Sample Error Analysis

Error type	Example
Input sentence not grammatical	<p><u>Sentence</u>: a hindu could:</p> <p><u>Question</u>: hindu could: a hindu could: could: a hindu could: could:</p>
Synonyms/named entities replaced when they're not supposed to be	<p><u>Sentence</u>: most hindus are found in asian countries .</p> <p><u>Question</u>: most muslims are found in what asian country ?</p>
Complex time periods incorrectly interpreted	<p><u>Sentence</u>: august is the eighth month of the year in the julian and gregorian calendars , and the fifth of seven months to have a length of 31 days .</p> <p><u>Question</u>: what is the length of the gregorian calendar ?</p>
Formula based sentence selected	<p><u>Sentence</u>: in other words , there are elements `` k " , \xa0`` x " , \xa0`` x " , \xa0 . . . , \xa0`` x " of the field `` f " such that `` p " -lrb- `` x .</p> <p><u>Question</u>: what does `` x " mean ?</p>
Multiple categories of errors - time period and grammar	<p><u>Sentence</u>: the virginids also shower at various dates in april .</p> <p><u>Question</u>: what month and year did the virginids also at various dates ?</p>



Future Work

- Additional/improved filters for articles/questions
- Generating different types of questions
- Real-time multiple choice question generation on user-input text
- Feedback loop for model improvement



AutoQ

Improving reading comprehension through
automatic question generation

First **free** web app with automatically generated reading comprehension questions targeted to English language learners

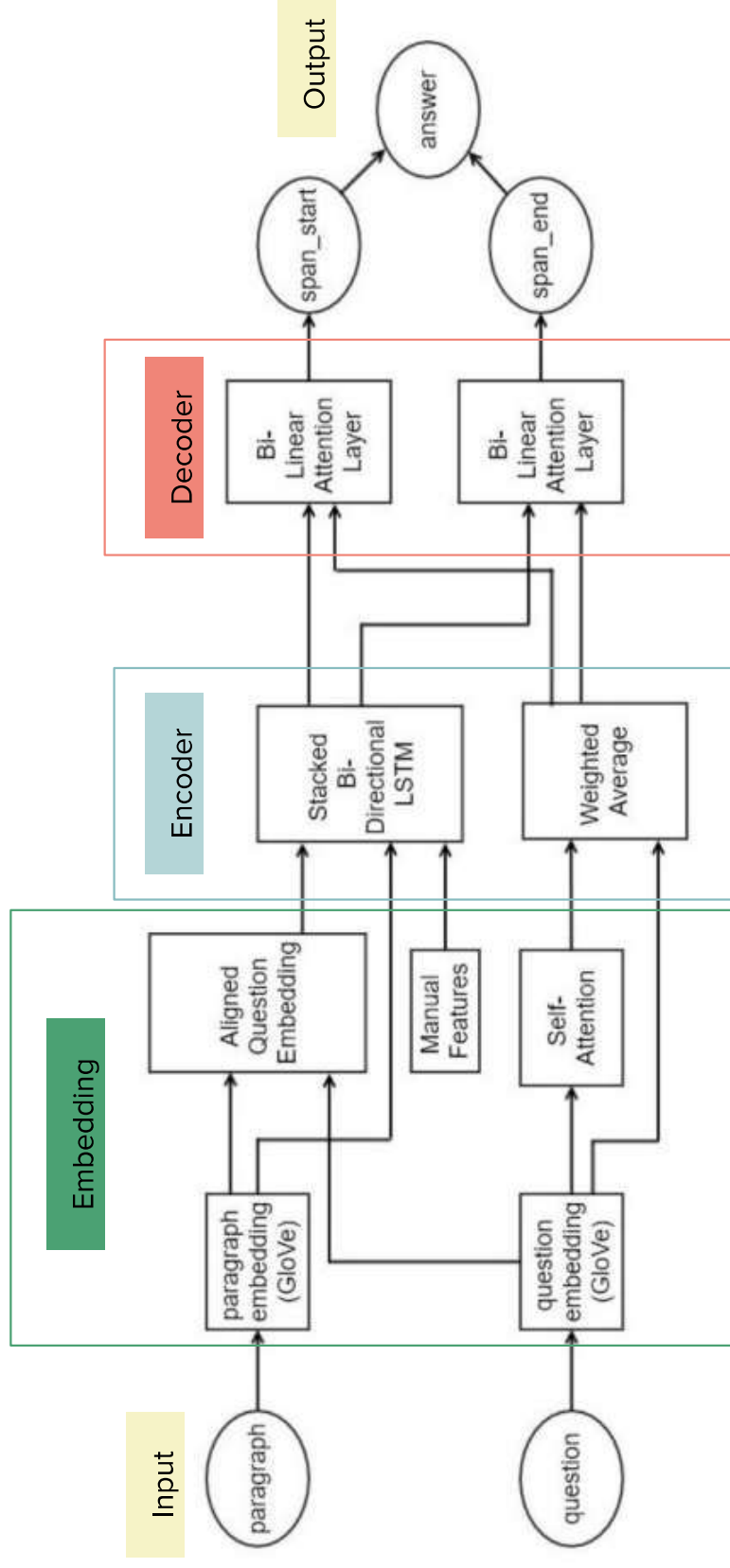
Offers over **100,000 practice questions** on over **10,000 Wikipedia articles**, widely used source for English language practice

Multiple-choice questions mimic English-language exam format, **scoring system** provides feedback for improvement

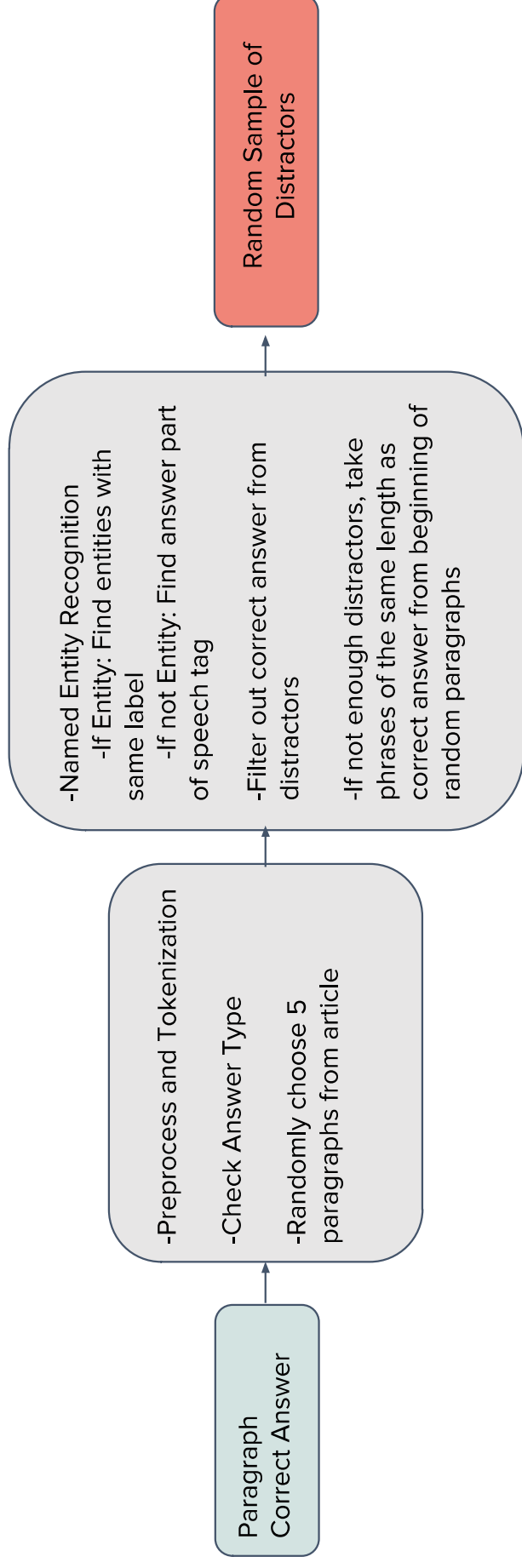


APPENDIX

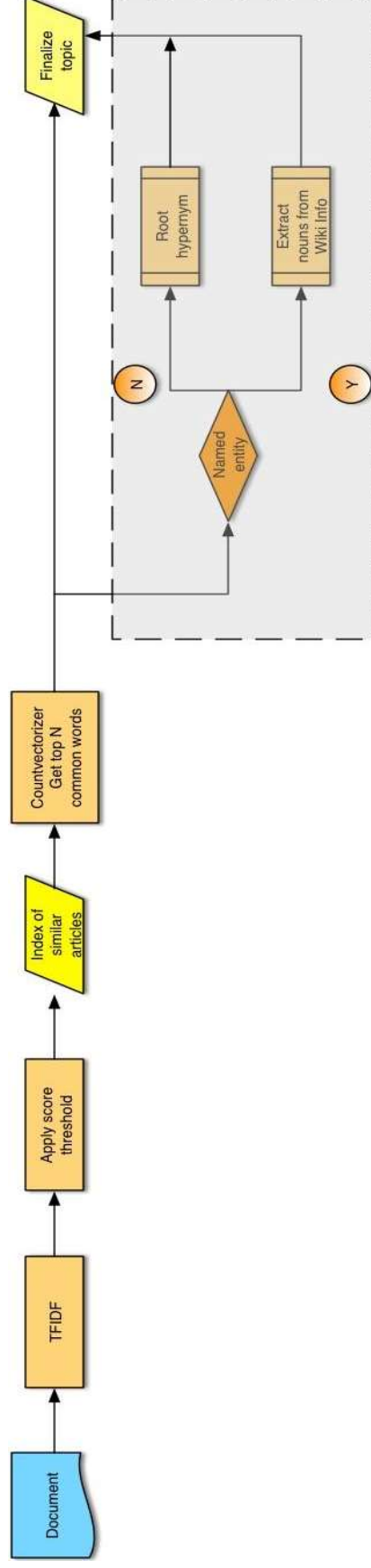
How it works: Question Answer Model



How it works: Distractor Generation Model



How it works: Topic Modelling



Examples of Good Distractors

question: what was one of the first two image formats used on web sites ?,
correct answer: XBM
distractors: apple, png, unisys

question: when did he discover the castle in london ?,
correct answers: late August 1605
distractors: August 28, 1658, June 15, 1669, June 14, 1658

question: what company did ge buy electronics for ?
correct answer: Lufkin Industries
distractors: healthcare financial services, ev solar carport, ge lighting

question: who is now professor emeritus of english ?
correct answer: Snyder
distractors: carole lynn koda, joanne kyger, glyn maxwell

question: how many giant pandas were sent to london ?
correct answer: five
distractors: at least two, one, three

question: what is the giraffe used for ?",
correct answer: scientific experiments and discoveries
distractors: okapi genes, various sounds, common targets

question: what is the traditional meals in cyprus ?
correct answer: breakfast dish
distractors: parts boiling water, bread crumbs, grits belt

Examples of Bad Distractors

question: what city did virgil live in ?

correct answer: San Francisco

distractors: haida, japan, oregon

question: where was `munich ' born ?

correct answer: 12 Arnold Grove in Wavertree, Liverpool

distractors: 2, 100, 2001

question: when was the invention of smokeless coating introduced ?

correct answer: late 19th century

distractors: small calibers, large artillery, hexagonal grained powders

question: how many albums did clark release as a solo performer ?

correct answer: several

distractors: new, jangly, irish catholic

question: what planet is a giant planet ?

correct answer: Neptune

distractors: Extrasolar, The, k

question: what is the name of the city that refers chaucer ?

correct answer: City Hustings Roll 110, 5, Ric II, dated June 1380, Chaucer

distractors: troilus, testament, tower

question: what did cukor eventually become?

correct answer: woman's director

distractors: probate court, vice charges, heart attack

Question Generation

- Attention-based sequence learning model
- Modeled after a paper by Du et al. from 2017, "Learning to Ask: Neural Question Generation for Reading Comprehension."

Question Answering

- Multi-layer recurrent neural network machine comprehension model
- Modeled after a paper by Chen et al. from 2017, "Reading Wikipedia to Answer Open-Domain Questions"

Distractor Generation

- Using spaCy and nltk, we find words/phrases with similar POS tags as the correct answer on 3 levels:
 - Same sentence
 - Same article
 - Different topic

Model Details

Text Generation with simus classmate



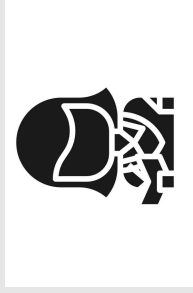
In-Depth Error Analysis

Types of errors:

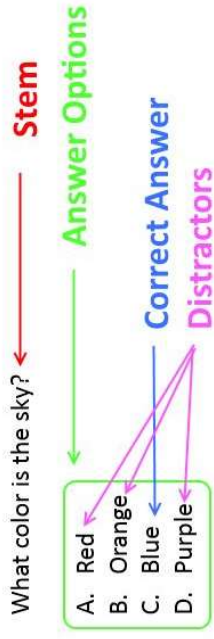
1. Input sentence not grammatical.
2. Synonyms/named entities replaced when they're not supposed to be.
3. We generate bad questions.
4. Sentence selection could have been better
- 5.

The Traditional Question Generation Approach

Experts read source material and write questions and answers.



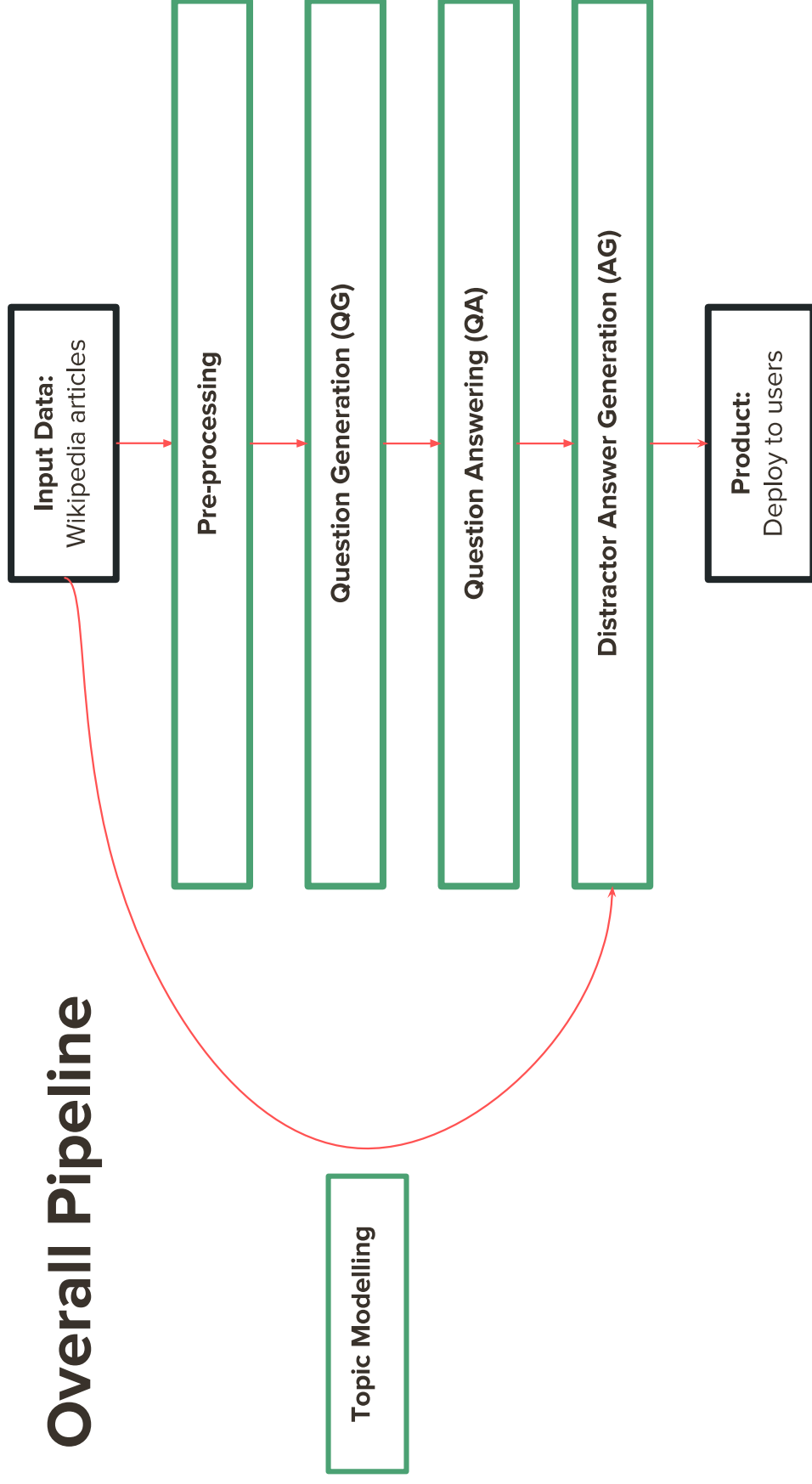
Experts identify main ideas and common sources of confusion to create wrong answer choices (“distractors”)



x 10

By automating the question generation steps with AutoQ, we reduce the costs associated with exam prep to allow learners more access to educational opportunities.

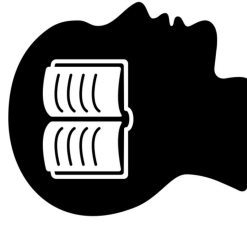
Overall Pipeline





Key Challenges

- Evaluation
 - Automatic evaluation scores are low, without acceptable reference text
 - Human evaluations can be more useful but more resource intensive
- Challenges in wrong answer generation
- Complexity of our pipeline



AutoQ

Improving reading comprehension through
automatic question generation

First web platform with
automatically generated reading
comprehensions targeted towards
English learners

Offers practice questions on —
new Wikipedia articles

Future Work

-

“Insert quote here?”

Why Reading?

Reading is a receptive skill that helps you improve all parts of the English language – vocabulary, spelling, grammar, and writing. It helps to develop language intuition in the corrected form.

Why Wikipedia articles?

- They are easily accessible
- Cited as good sources of practice for academic reading
- Applicable to real life given the range of topics

AutoQ seeks to democratize reading comprehension questions using machine learning and natural language processing

