



Hosted by: Data Science Milan

Colab Code and slides:

<http://bit.ly/datasciencemilan>

MCQ Generation in edtech using Natural Language Processing

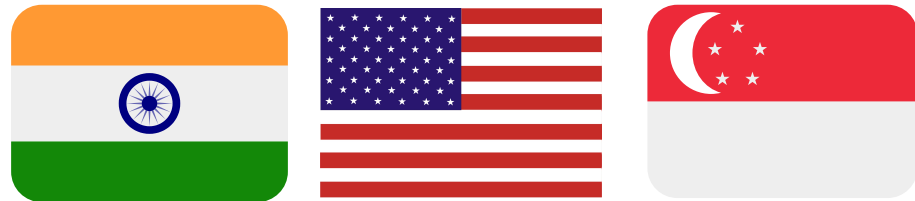
**Thu, May 20th, 2021
10 PM – 11:30 PM (India)
6:30 PM – 8:00 PM (Italy)**



**Talk by
Ramsri Goutham
ramsri.ai**

Who am I ?

- Lead Data Scientist with 8+ years of work experience across Silicon Valley, Singapore, and India.



- Most recently been a Co-founder and CTO of an AI-assisted assessments startup.
- Spent the last 2 years on question generation using Natural Language processing in edtech.



Ramsri Goutham Golla

ramsri.ai

Open Source library for question generation

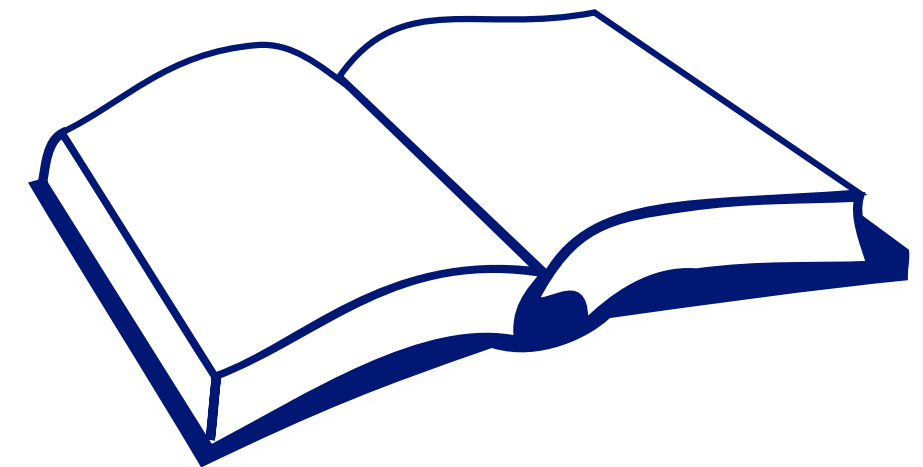
Questgen

<https://github.com/ramsrigouthamg/Questgen.ai>

The problem

Assessment creation (Quizzes and Tests) is manual and not scalable for online transformation due to the pandemic.

A textbook publisher hired you as a data scientist to create a tool that automates the assessment creation process .



The solution

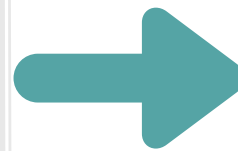
News Article

Elon Musk and Dogecoin

Elon Musk has shown again he can influence the digital currency market with just his tweets.

After saying that his electric vehicle-making company Tesla will not accept payments in Bitcoin because of environmental concerns, he tweeted that he was working with developers of Dogecoin to improve system transaction efficiency.

Following the two distinct statements from him, the world's largest cryptocurrency hit a two-month low, while Dogecoin rallied by about 20 percent.



1

Multiple Choice Questions (MCQs)

2

True or False Questions

3

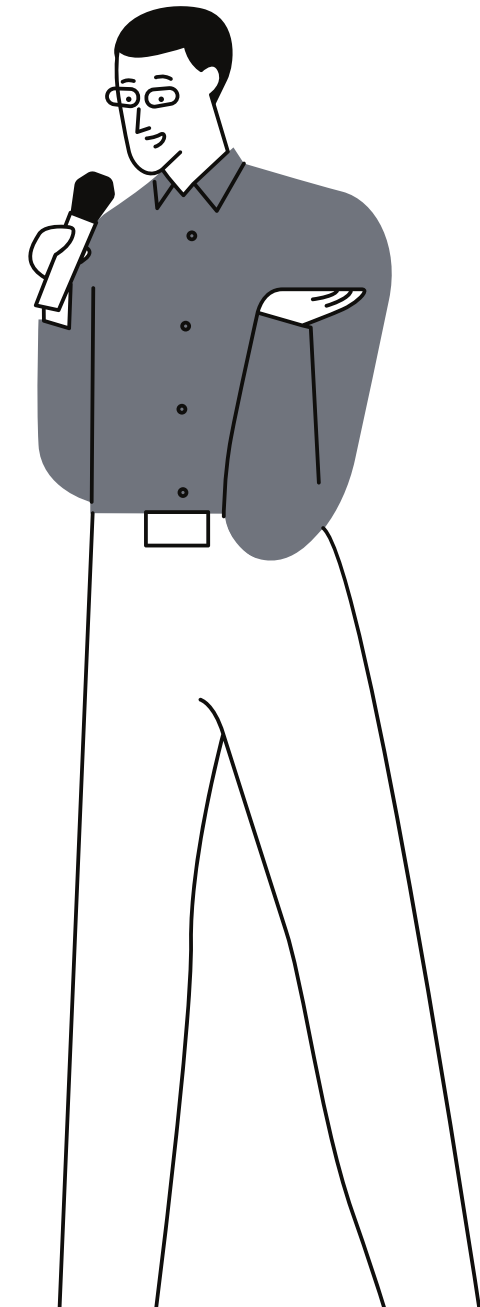
Fill in the Blanks

4

Match the following

Let's get started!

Are you ready?



Talk Outline

We will use Google Colab notebooks as the development environment for our demos

1

Text2MCQ generation using **T5** transformers and sense2vec vectors

2

Use **ONNX** quantization to create a light-weight model of **T5**

Step 1 : Identify key Sentences/Concepts

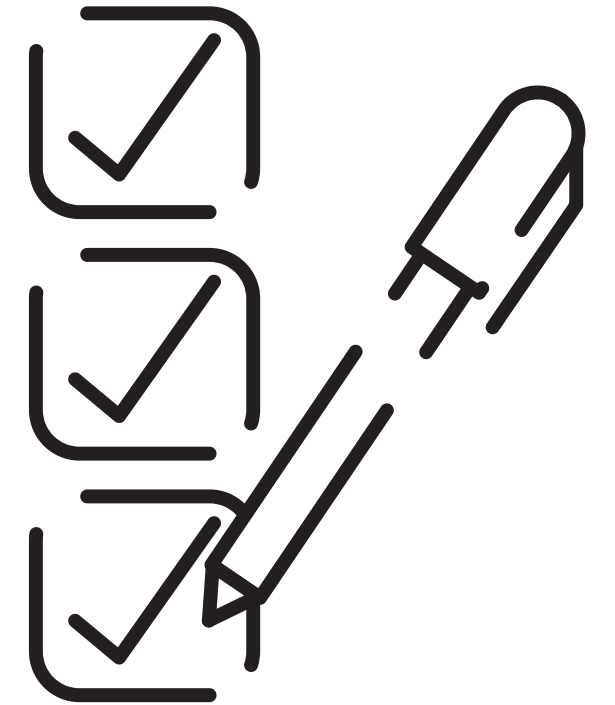
News Article

Elon Musk and Dogecoin

Elon Musk has shown again he can influence the digital currency market with just his tweets. After saying that his electric vehicle-making company **Tesla will not accept payments in Bitcoin** because of environmental concerns, he tweeted that he was working with developers of **Dogecoin to improve system transaction efficiency**.

Following the two distinct statements from him, **the world's largest cryptocurrency hit a two-month low, while Dogecoin rallied by about 20 percent**. The SpaceX CEO has in recent months **often tweeted in support of Dogecoin, but rarely for Bitcoin**. In a recent tweet, Musk put out a statement from Tesla that it was “concerned” about the rapidly increasing use of fossil fuels for Bitcoin mining and transaction, and hence was suspending vehicle purchases using the cryptocurrency.

A day later he again tweeted saying, “To be clear, I strongly believe in crypto, but it can't drive a massive increase in fossil fuel use, especially coal”. It triggered a downward spiral for Bitcoin value but the cryptocurrency has stabilized since.



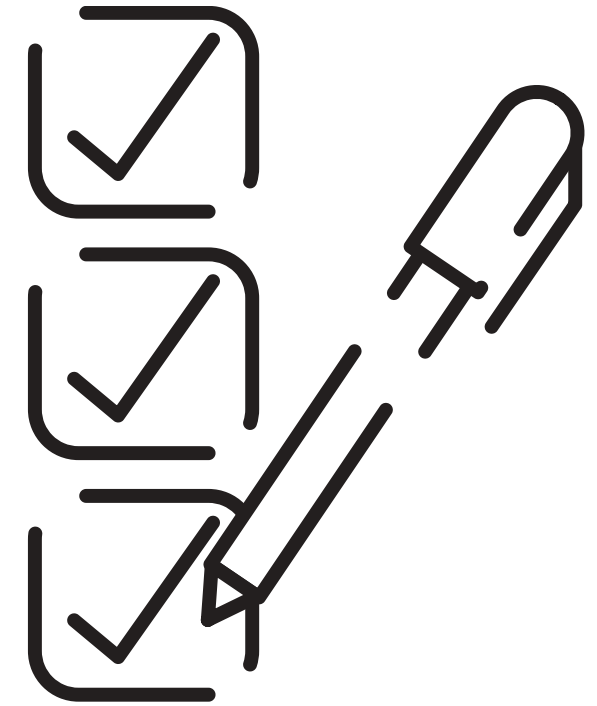
Step 2 : Identify keywords from sentences

News Article

Elon Musk and Dogecoin

Musk tweeted that his electric vehicle-making company **tesla** will not accept payments in **bitcoin** because of environmental concerns. He also said that the company was working with developers of **dogecoin** to improve system transaction efficiency.

The world's largest cryptocurrency hit a two-month low, while **dogecoin** rallied by about 20 percent. Musk has in recent months often tweeted in support of crypto, but rarely for bitcoin.



Extract Keywords

Musk

Tesla

Bitcoin

Dogecoin

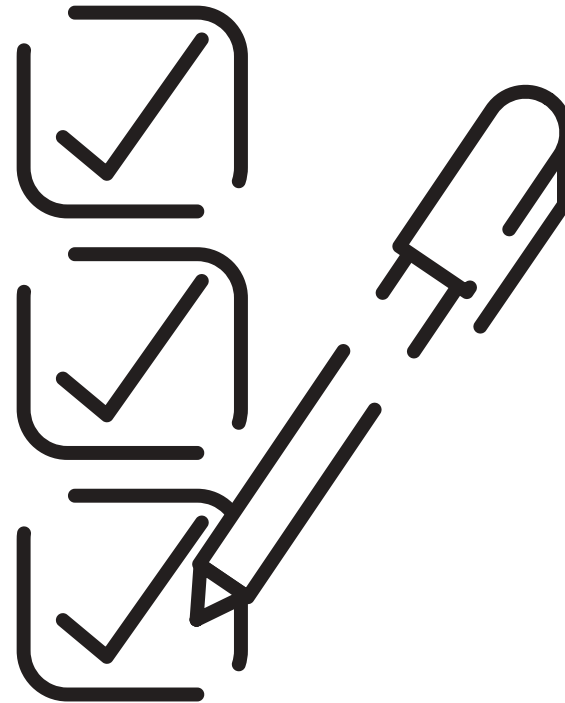


Step 3: Form multiple-choice questions (MCQs)

Elon Musk and Dogecoin

Musk tweeted that his electric vehicle-making company **tesla** will not accept payments in **bitcoin** because of environmental concerns. He also said that the company was working with developers of **dogecoin** to improve system transaction efficiency.

The world's largest cryptocurrency hit a two-month low, while **dogecoin** rallied by about 20 percent. Musk has in recent months often tweeted in support of crypto, but rarely for bitcoin.



What cryptocurrency did Musk rarely tweet about?

Bitcoin

Dogecoin

ETH

Litecoin



What company did Musk say would not accept bitcoin payments?

Tesla

Nissan

Ford

Toyota

Posing it as an NLP problem

News Article

Elon Musk and Dogecoin

Elon Musk has shown again he can influence the digital currency market with just his tweets.

After saying that his electric vehicle-making company Tesla will not accept payments in Bitcoin because of environmental concerns, he tweeted that he was working with developers of Dogecoin to improve system transaction efficiency.

Following the two distinct statements from him, the world's largest cryptocurrency hit a two-month low, while Dogecoin rallied by about 20 percent.

1

Abstractive/Extractive summarization (T5, BART, Pegasus Models)

2

Paraphrasing of sentences (T5, GPT-2 etc)

3

Keyword/Keyphrase Extraction (YAKE, TopicRank, KeyBERT etc)

4

Question generation (T5, GPT-2, GPT-3 etc)

5

Wrong choices/distractors generation (Wordnet, Word2vec, Sense2vec etc)

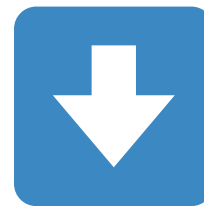
Generate a question using T5 Transformer model

Context:

Musk tweeted that his electric vehicle-making company **Tesla** will not accept payments in bitcoin because of environmental concerns.

Answer:

Tesla



**Text to text transfer
transformer**

Question

What company did Musk say would not accept bitcoin payments?

Squad dataset used for training

<https://rajpurkar.github.io/SQuAD-explorer/>

Nikola_Tesla

The Stanford Question Answering Dataset

Nikola Tesla (Serbian Cyrillic: Никола Тесла; 10 July 1856 – 7 January 1943) was a Serbian American inventor, electrical engineer, mechanical engineer, physicist, and futurist best known for his contributions to the design of the modern alternating current (AC) electricity supply system.

In what year was Nikola Tesla born?

Ground Truth Answers: 1856 1856 1856

Prediction: 1856

What was Nikola Tesla's ethnicity?

Ground Truth Answers: Serbian Serbian Serbian

Prediction: Serbian

In what year did Tesla die?

Ground Truth Answers: 1943 1943 1943

Prediction: 1943

Extract Keywords

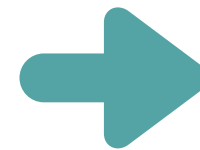
News Article

Elon Musk and Dogecoin

Elon Musk has shown again he can influence the digital currency market with just his tweets.

After saying that his electric vehicle-making company Tesla will not accept payments in Bitcoin because of environmental concerns, he tweeted that he was working with developers of Dogecoin to improve system transaction efficiency.

Following the two distinct statements from him, the world's largest cryptocurrency hit a two-month low, while Dogecoin rallied by about 20 percent.



Extract Keyphrases using
Python Keyword Extraction Library

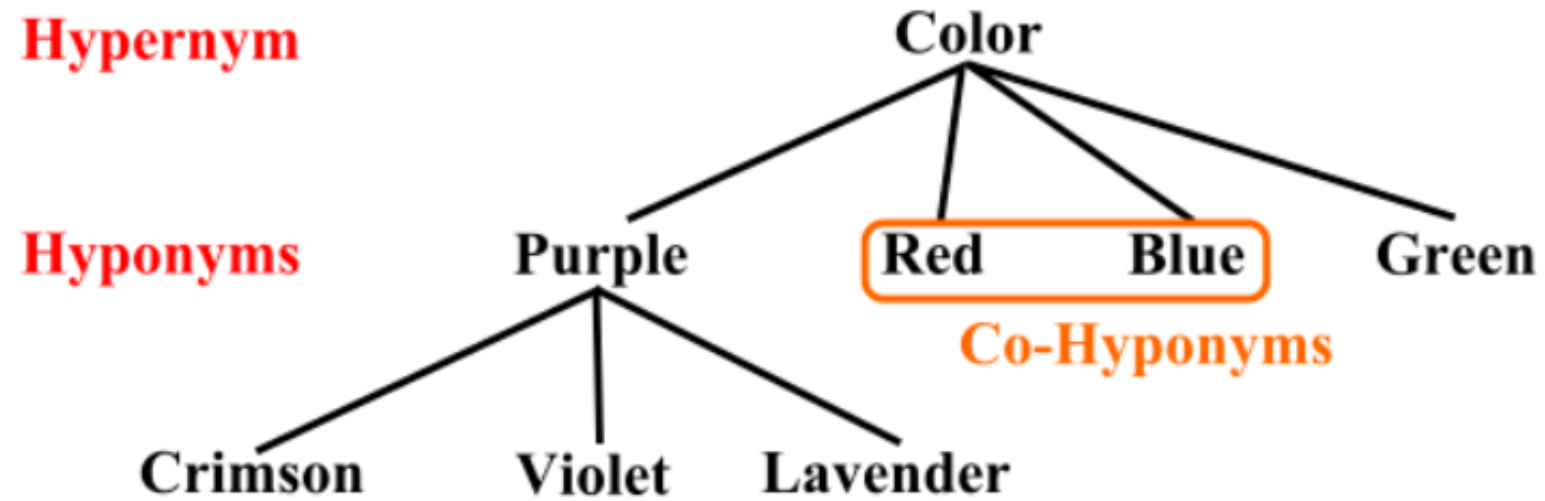
<https://github.com/boudinfl/pke>


Bitcoin, Dogecoin, Tesla, Cryptocurrency

MultipartiteRank

Generate distractors using Wordnet

<https://wordnet.princeton.edu/>



An example of the relationship between hyponyms and hypernym 

Source: https://en.wikipedia.org/wiki/Hyponymy_and_hypernymy

- WordNet® captures relations.
- A hyponym is in a type-of relationship with its hypernym.
- A hypernym is an umbrella term and blanket term. Eg: Color is the hypernym for Purple.
- Red, Blue, Green are the hyponyms of "Color". Since Red is a type of color.
- Our goal is to extract Co-Hyponyms as distractors.

Generate distractors using Sense2vec

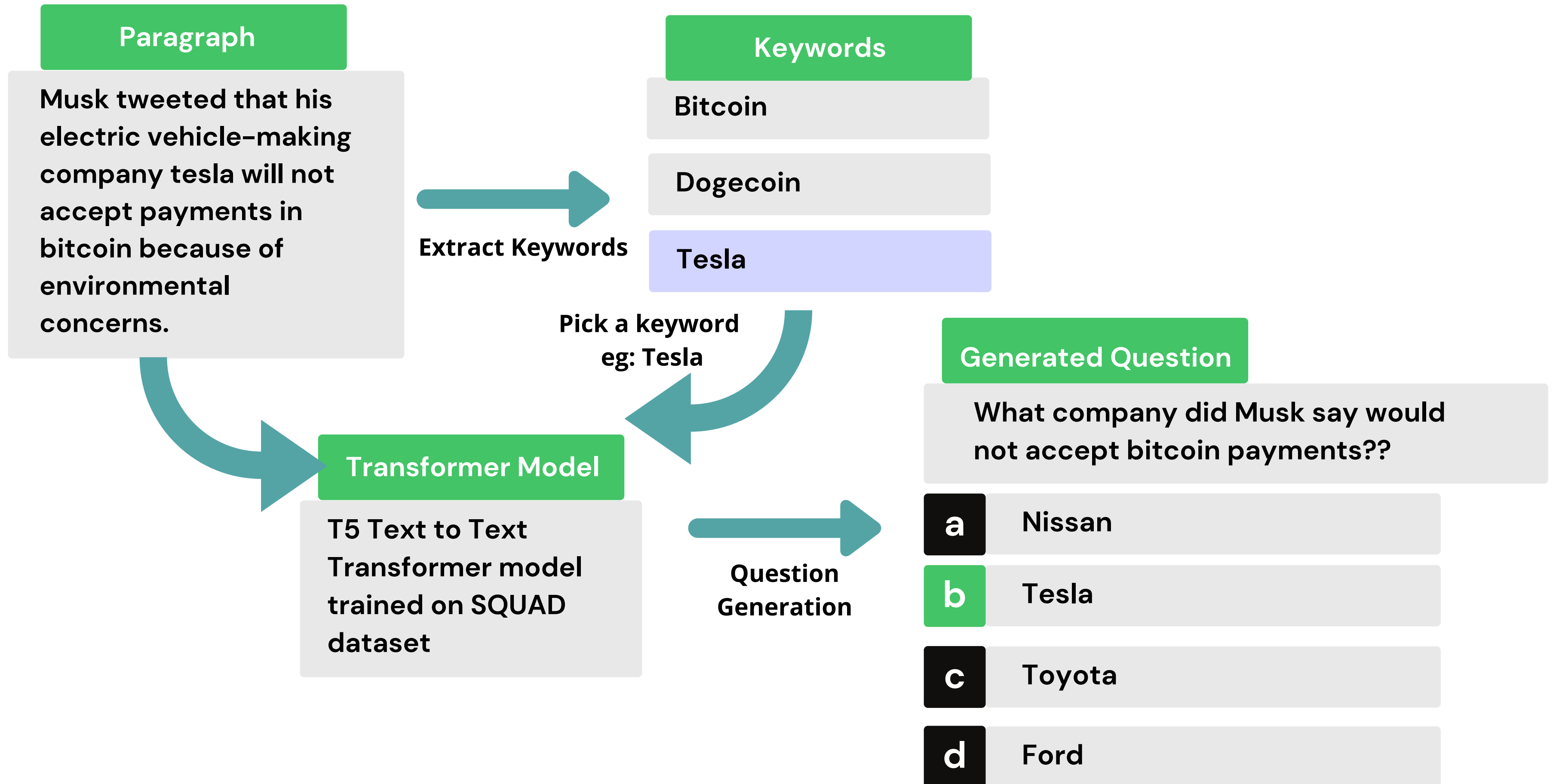
<https://explosion.ai/blog/sense2vec-reloaded>

Term	Sense ?	Year ?
natural language processing	auto	2015
machine learning >		90%
computer vision >		86%
deep learning >		86%
data analysis >		84%
neural nets >		83%
relational databases >		82%

Source: <https://explosion.ai/demos/sense2vec>

- Contextual information is captured. Trained on Reddit comments.
- The words with the same senses are differentiated with parts of speech eg: duck|VERB and duck|NOUN
- Noun phrases and named entities are annotated during training so multiword phrases like "natural language processing" also have an entry.

Overview: Generate MCQs from any content.



Text2MCQ

Google Colab Coding Session

Colab Code and slides:

<http://bit.ly/datasciencemilan>

Udemy Course: Question generation using
Natural Language Processing (bit.ly/questiongeneration)

Coupon Code: **DSMILAN**



Connect with me
on Social Media:

ramsri.ai

Email:

ramsri@questgen.ai