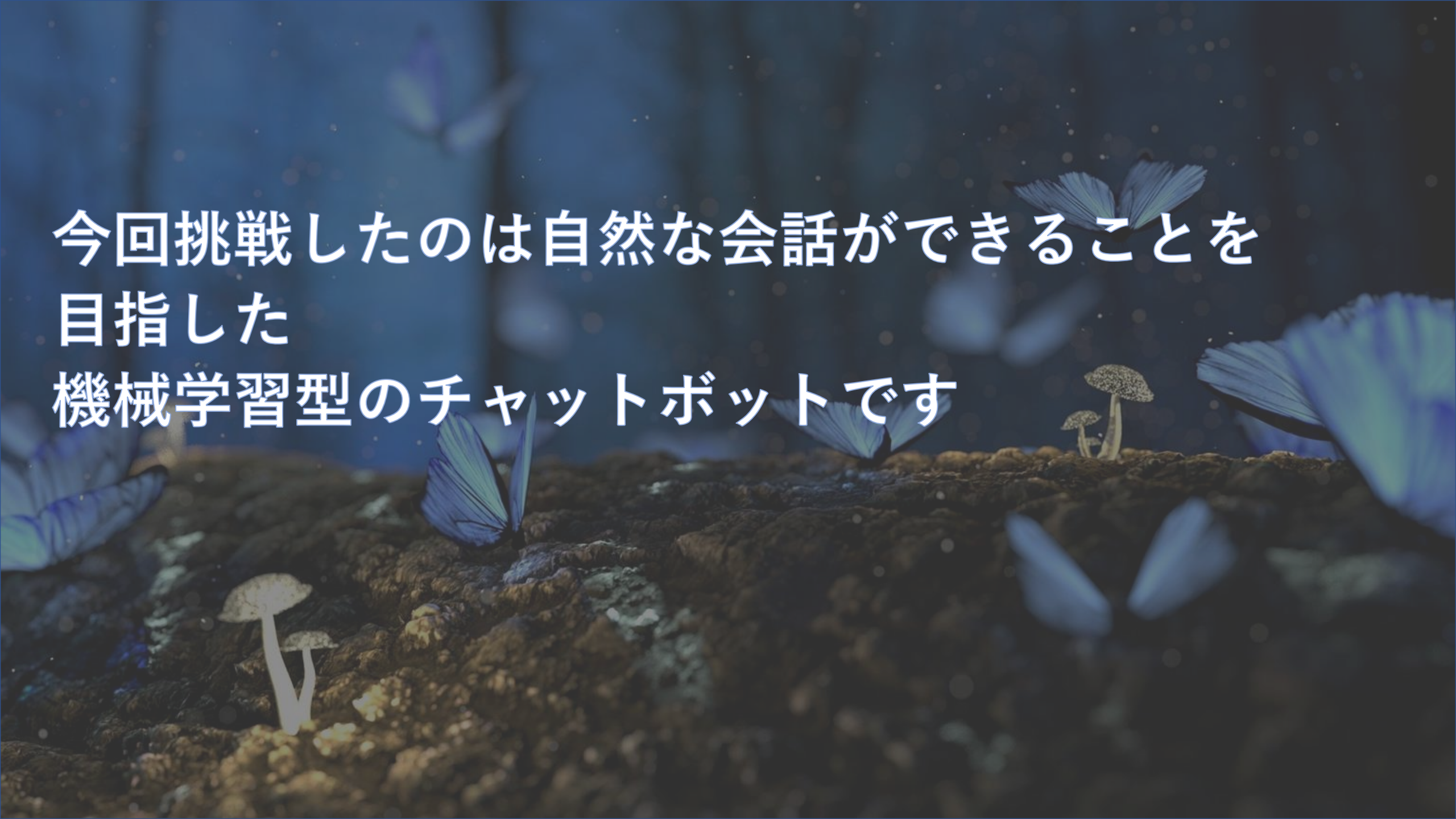


最終プレゼン

IE4A

高田駿介

今回挑戦したのは自然な会話ができることを
目指した
機械学習型のチャットボットです



チャットボットには
ルールベース型
機械学習型
の2種類があります。



ルールベース型


メリット

シナリオ通りに回答するので、誤った回答をする心配がない

デメリット

ルール通りにしか回答できない

機械学習型

The background of the slide is a dark, atmospheric scene of a forest floor. It features several mushrooms of various sizes and colors, including white and brown ones, scattered across a dark, textured ground. Several blue butterflies are in flight, their wings spread, against a dark blue, slightly hazy background. The overall lighting is dim, creating a mysterious and naturalistic setting.

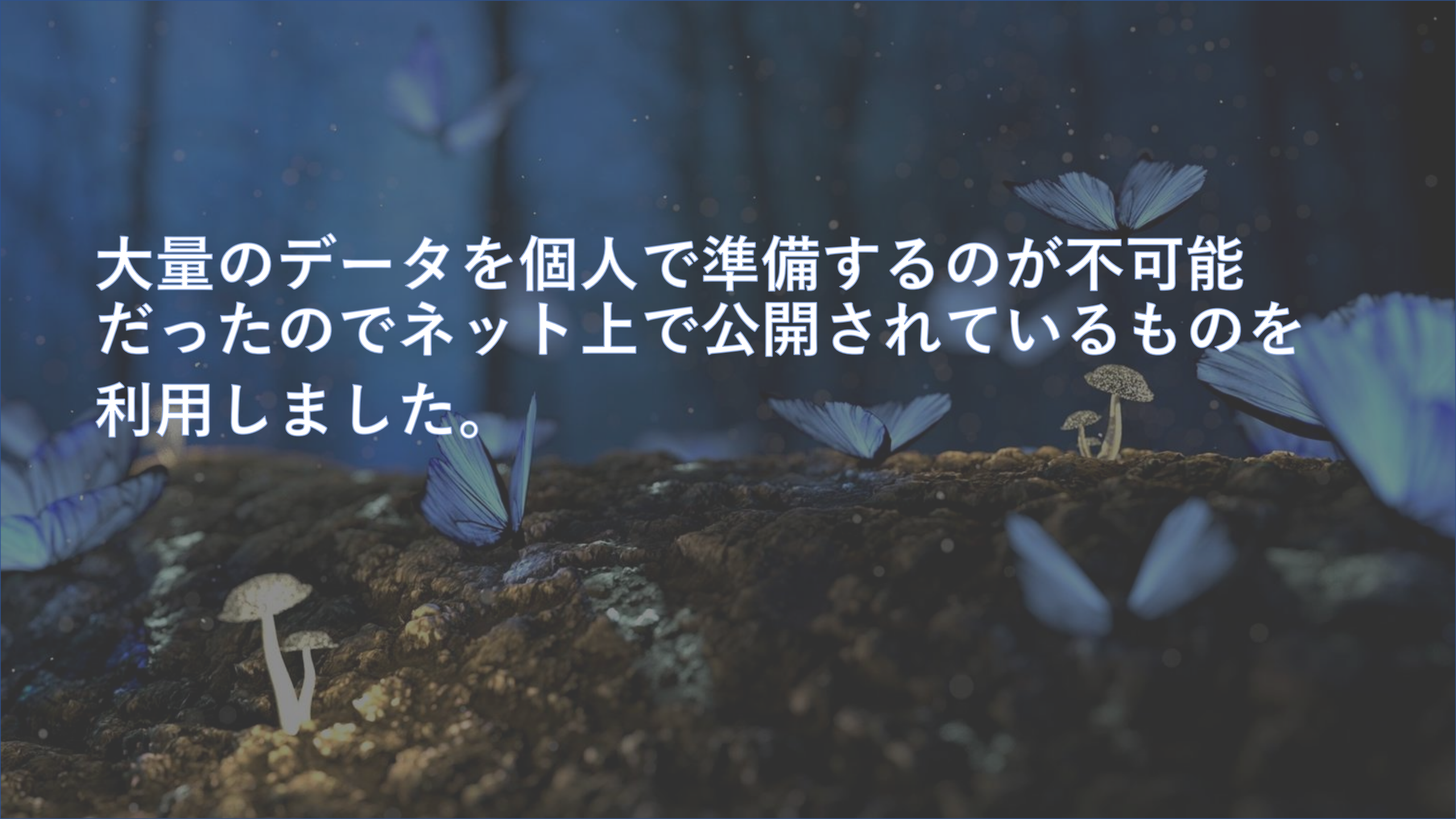
メリット

ルールベース型に比べて、複雑な回答ができる

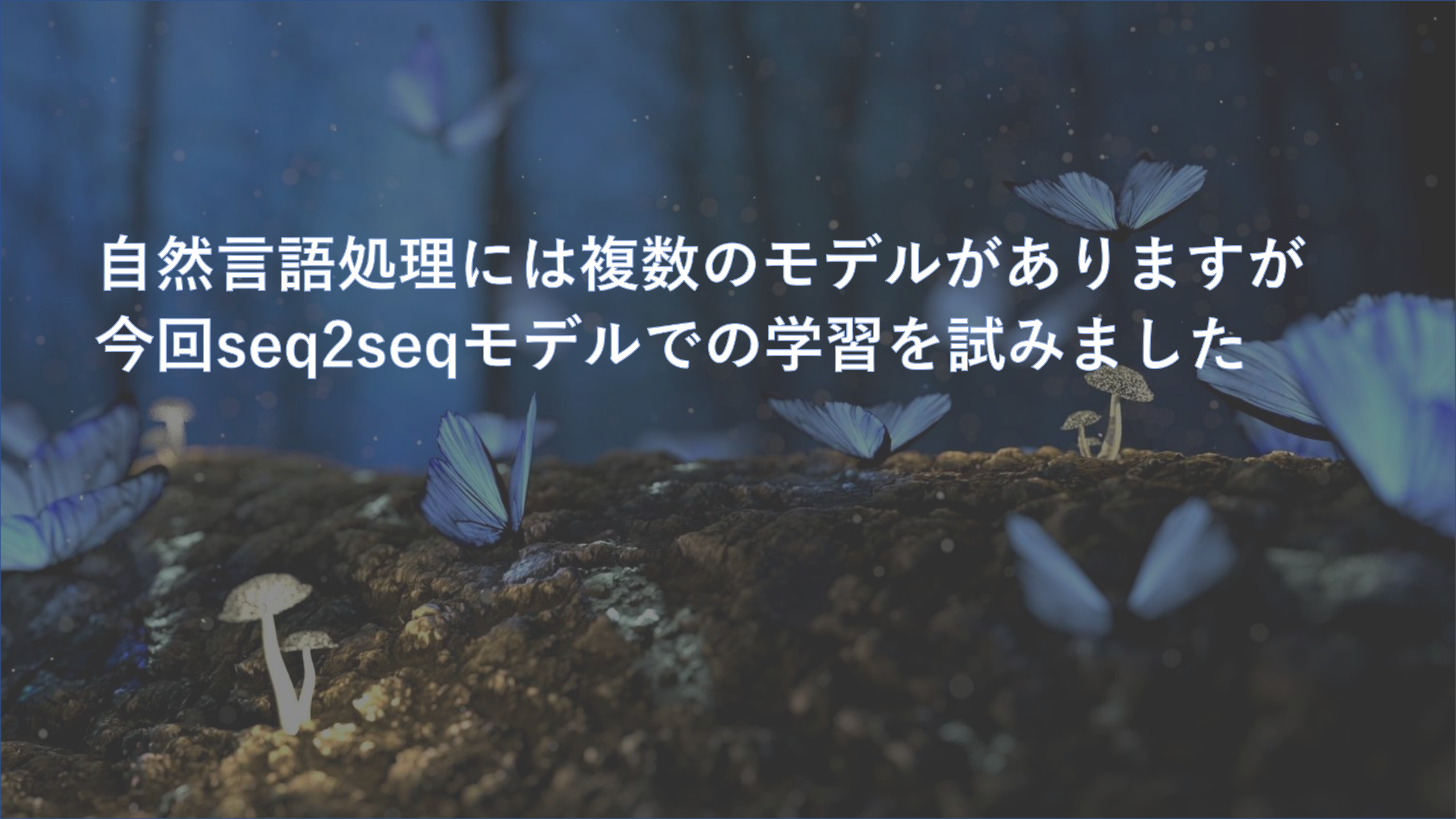
デメリット

会話の精度をあげるために、大量の正しいデータが必要

大量のデータを個人で準備するのが不可能
だったのでネット上で公開されているものを
利用しました。



自然言語処理には複数のモデルがありますが
今回seq2seqモデルでの学習を試みました



seq2seq

EncoderとDecoderを備えたEncoder-Decoderモデルを使って、
系列データを別の系列データに変換するモデル

Word2Vecの発展系

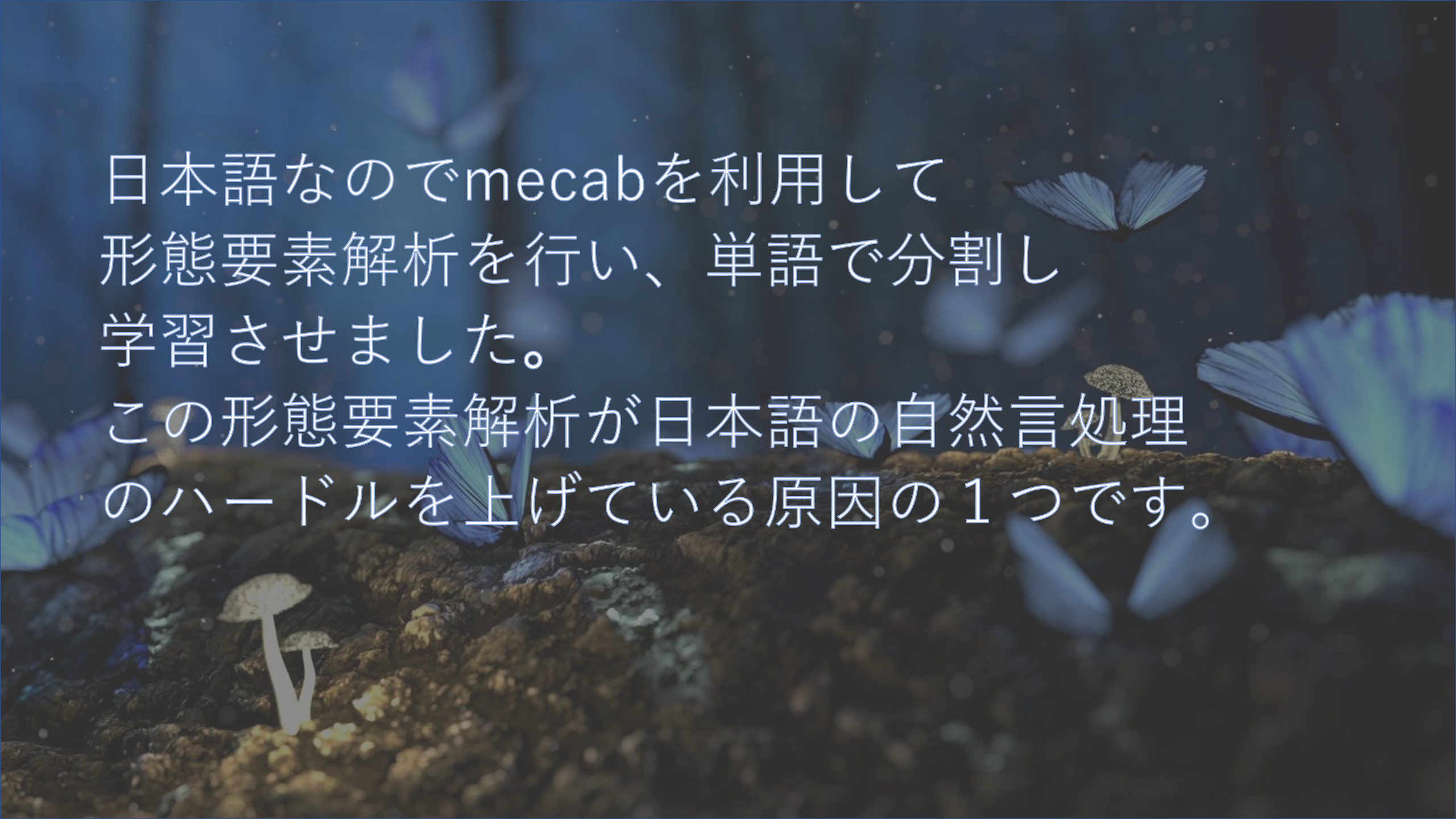
Googleの翻訳などで使われている技術です

日本語 → 英語

発話 → 応答

の形で実装します





日本語なのでmecabを利用して
形態要素解析を行い、単語で分割し
学習させました。

この形態要素解析が日本語の自然言処理
のハードルを上げている原因の1つです。

Perplexityが1.0になり値が変動しなくなっ
たので学習を
終了させました




```
> ^CTraceback (most recent call last):
  File "/Users/shunsuke/Documents/zemi/decode.py", line 180, in <module>
    tf.app.run()
  File "/Users/shunsuke/opt/anaconda3/envs/sub/lib/python3.6/site-packages/tensorflow/python/platform/app.py", line 43, in run
    sys.exit(main(sys.argv[:1] + flags_passthrough))
  File "/Users/shunsuke/Documents/zemi/decode.py", line 177, in main
    decode()
  File "/Users/shunsuke/Documents/zemi/decode.py", line 140, in decode
    sentence = sys.stdin.readline()
KeyboardInterrupt
(sub) takadashunsukenoMacBook-Pro:zemi shunsuke$ /Users/shunsuke/opt/anaconda3/envs/sub/bin/python /Users/shunsuke/Documents/zemi/decode.py
/Users/shunsuke/opt/anaconda3/envs/sub/lib/python3.6/site-packages/tensorflow/python/framework/dtypes.py:463: FutureWarning: Passing (type,
a future version of numpy, it will be understood as (type, (1,)) / '(1,)type'.
  _np_qint8 = np.dtype(["qint8", np.int8, 1])
/Users/shunsuke/opt/anaconda3/envs/sub/lib/python3.6/site-packages/tensorflow/python/framework/dtypes.py:464: FutureWarning: Passing (type,
a future version of numpy, it will be understood as (type, (1,)) / '(1,)type'.
  _np_quint8 = np.dtype(["quint8", np.uint8, 1])
/Users/shunsuke/opt/anaconda3/envs/sub/lib/python3.6/site-packages/tensorflow/python/framework/dtypes.py:465: FutureWarning: Passing (type,
a future version of numpy, it will be understood as (type, (1,)) / '(1,)type'.
  _np_qint16 = np.dtype(["qint16", np.int16, 1])
/Users/shunsuke/opt/anaconda3/envs/sub/lib/python3.6/site-packages/tensorflow/python/framework/dtypes.py:466: FutureWarning: Passing (type,
a future version of numpy, it will be understood as (type, (1,)) / '(1,)type'.
  _np_quint16 = np.dtype(["quint16", np.uint16, 1])
/Users/shunsuke/opt/anaconda3/envs/sub/lib/python3.6/site-packages/tensorflow/python/framework/dtypes.py:467: FutureWarning: Passing (type,
a future version of numpy, it will be understood as (type, (1,)) / '(1,)type'.
  _np_qint32 = np.dtype(["qint32", np.int32, 1])
WARNING:tensorflow:From /Users/shunsuke/opt/anaconda3/envs/sub/lib/python3.6/site-packages/tensorflow/models/rnn/translate/seq2seq_model.py
hon.ops.variables) is deprecated and will be removed after 2017-03-02.
Instructions for updating:
Please use tf.global_variables instead.
Reading model parameters from /Users/shunsuke/Documents/zemi/datas/translate.ckpt-55700
Hello!!
> █
```


結果

条件を変えて学習させましたが最終的には
時々まともな返答をするものになりました



考察

おそらくデータの精度がダメだった
けれどコーパスを個人で集めるのは難易度が高く、
英語ならばデータがあるが日本語は公開されているデータが
少なすぎる

考察

Q&Aなど利用用途を限定して
そのデータのみを学習させるなどであれば
精度の高いものができると思うのですが
自然な会話は現在では不可能だと思いました。