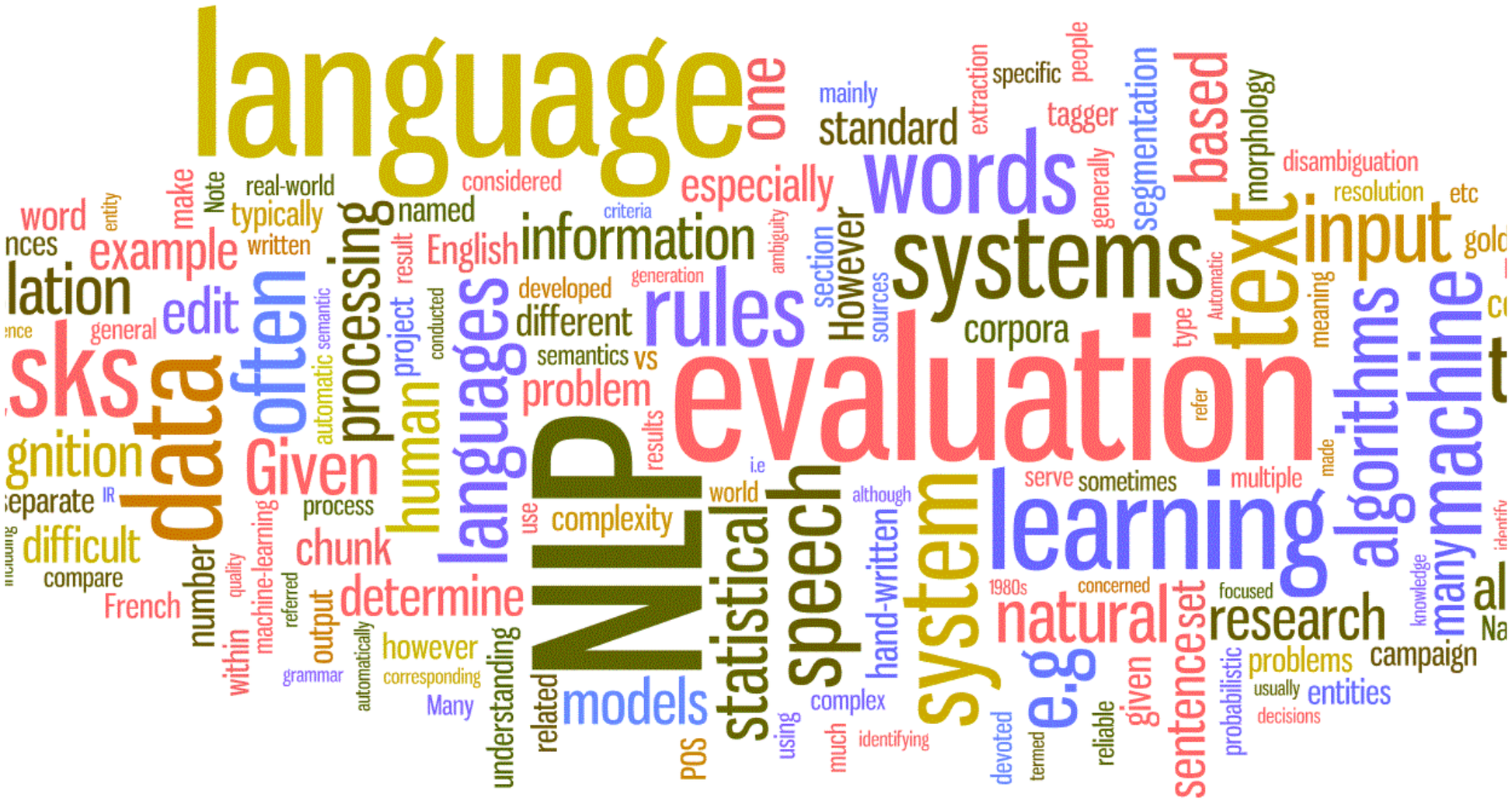


NLP IN DAILY LIFE

by Yuni Susanti

-start-

POPULAR TERMS NOWADAYS

Artificial Intelligence? Machine Learning? Deep learning? Neural Network?

- Jika digeneralkan, semua hal ini bisa merujuk pada hal yang sama

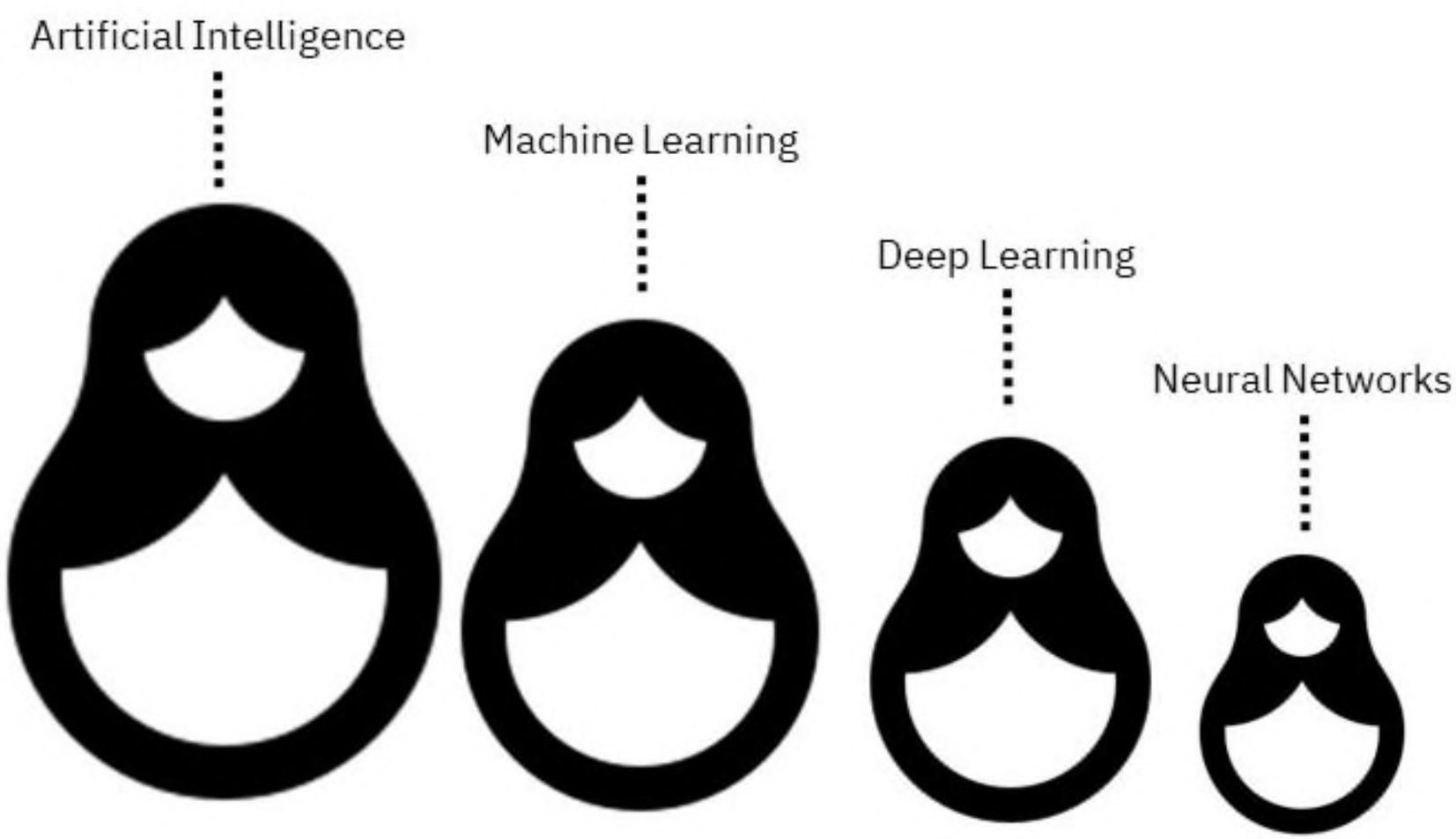
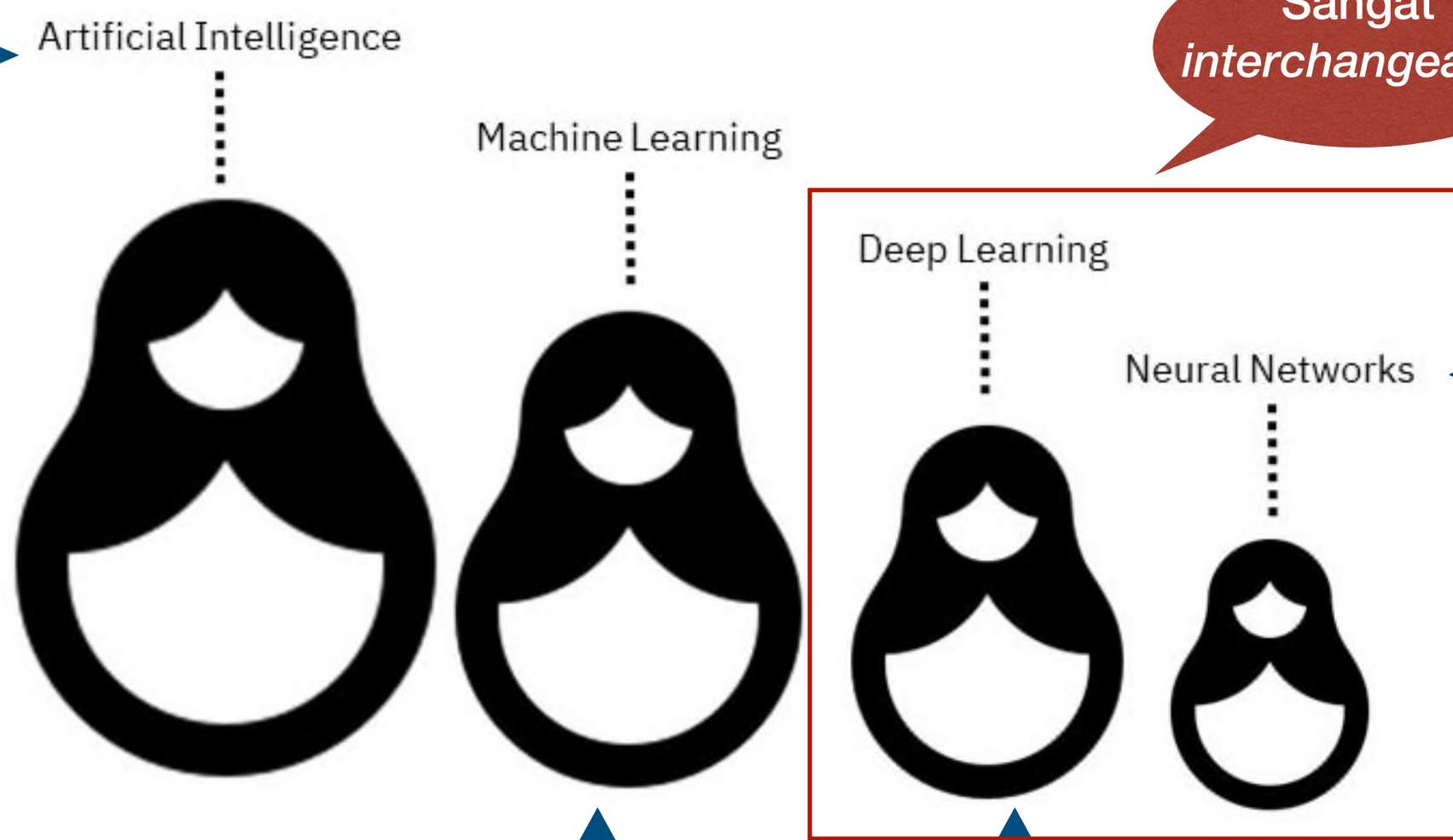


*image taken from https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks

Artificial Intelligence? Machine Learning? Deep learning? Neural Network?

Sistem yang "mimics" (me niru) human intelligence, i.e. learning, problem solving

Sangat *interchangeable*



Cara belajar (learning) yang terinspirasi dari human brain

Sistem yang bisa "belajar sendiri" dari data untuk digunakan prediksi data baru

machine learning yang menggunakan neural network dengan banyak layer (deep)

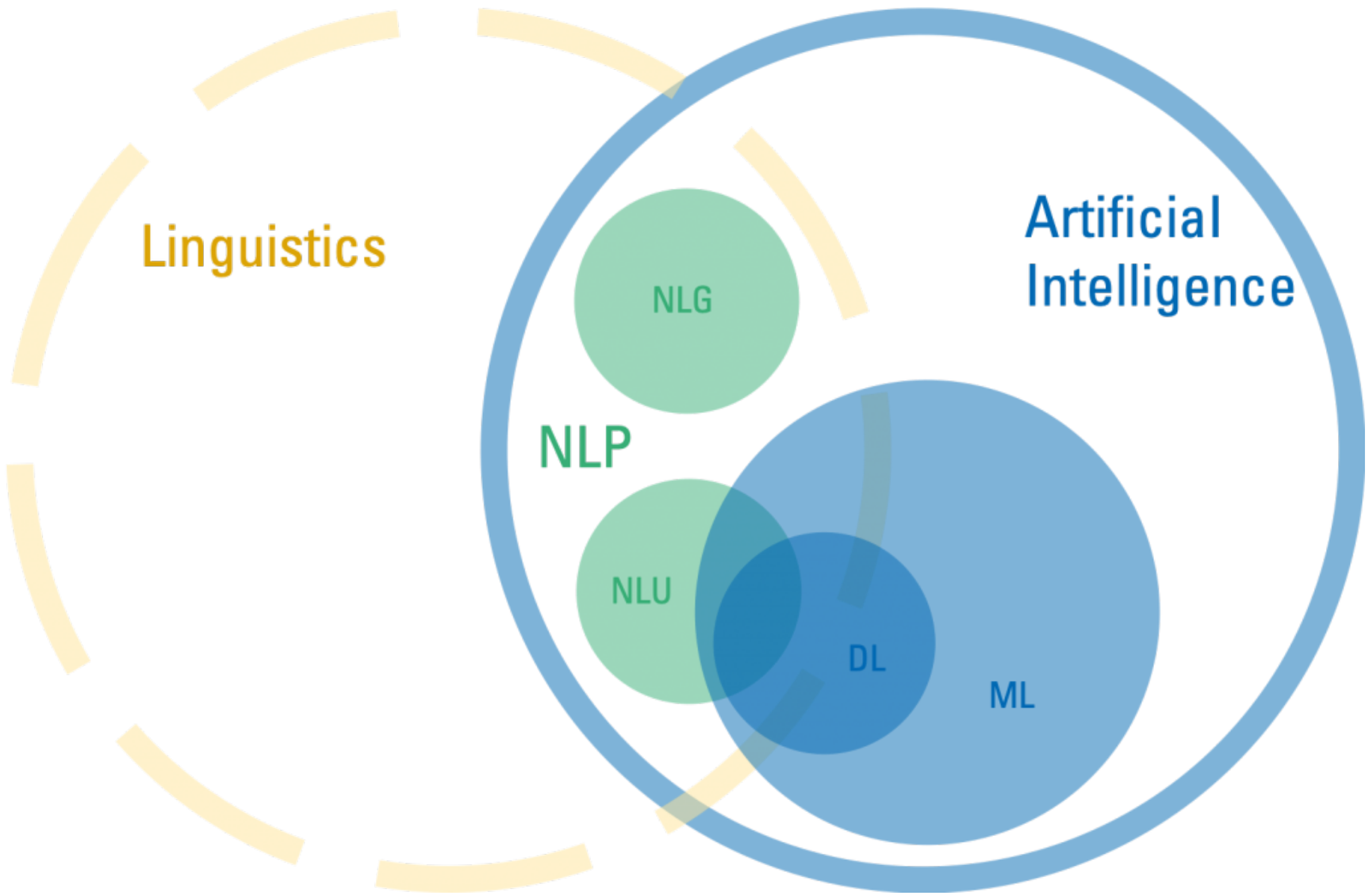
Natural Language Processing

(very loose) definition:

a branch of artificial intelligence that deals with the interaction between computers and humans using the natural language.

e.g. Bahasa Indonesia, Inggris, Jepang.. dll

NLP and its related fields

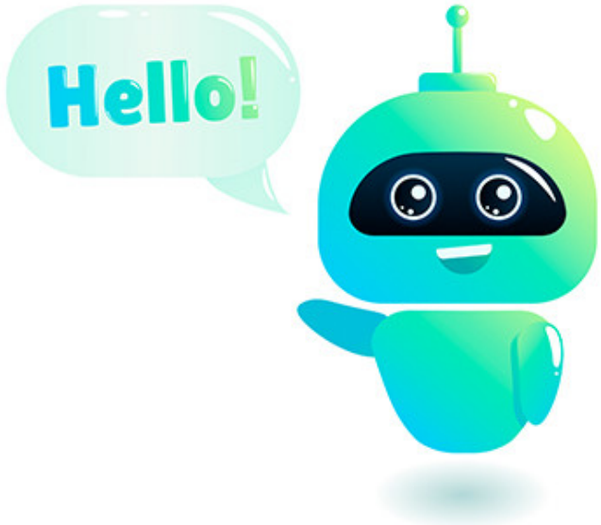


Intersection dari Linguistics (bahasa) dan Artificial Intelligence (Computer Science)

NLP Applications

(as you might be aware)

Google Translate



Hey Siri

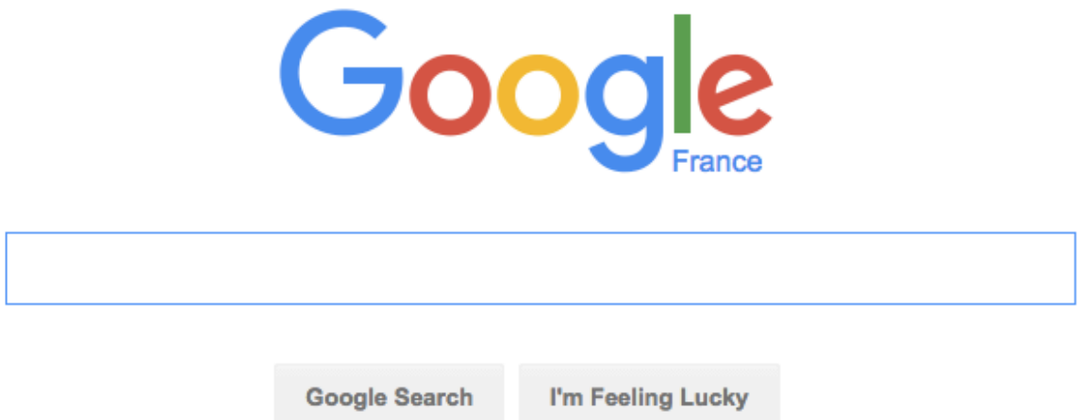


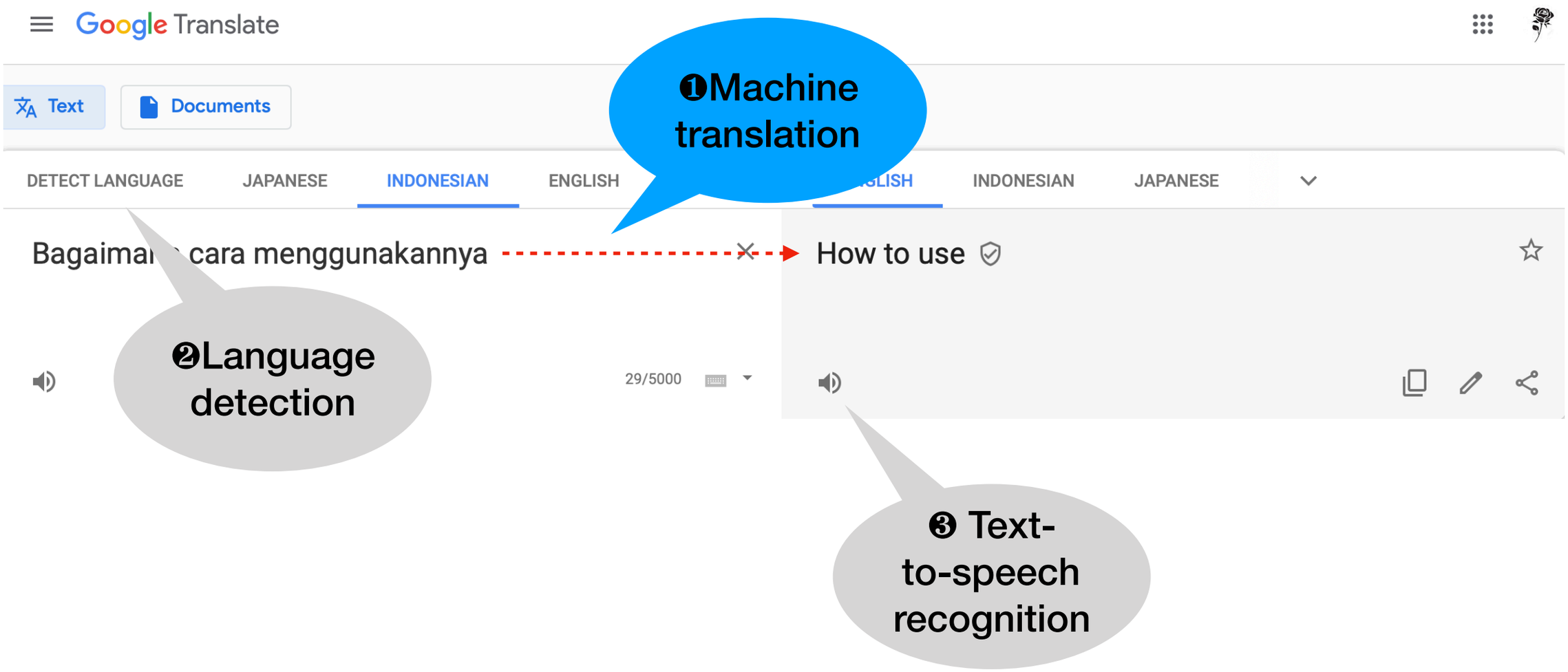
*image taken from the respective brand's website

Let's start with Machine Translation

“Bedah” Google Translate (1)

…adalah sebuah “Machine Translation” system

(tapi sebenarnya lebih kompleks dari pada itu)



Machine Translation?

Machine Translation (MT): *subtask NLP khusus mempelajari sistem penerjemah otomatis teks (atau ucapan) dari satu bahasa ke bahasa yang lain*

MT system

How (simplest) MT works?

Word-to-word translation: cara paling <u>sederhana</u>

Indonesian: *Saya suka apel.* ***English:*** *I love apple.*

Indonesian-English database

Saya ----------→ I
suka ----------→ love
apel ----------→ apple

Cukupkah hanya dengan ini?

How: Menggunakan sebuah database pasangan kata Indonesia-Inggris

Komponen dari bahasa

1) **Kata (*word, token*)**: bagian penyusun paling sederhana dari bahasa

token token token

Saya suka apel

2) **Konteks**: bagian pengatur bahasa, e.g. urutan kata dalam kalimat, grammar, syntax, semantic (meaning)…

<u>Perempuan</u> cantik itu bekerja di <u>Apple</u>

Noun *Adjective*

Grammar: adjective (kata sifat) harus mengikuti noun (kata benda)

Buah?

Device?

Company?

Karena adanya KONTEKS, word-to-word translation jadi tidak akurat

How current MT works?

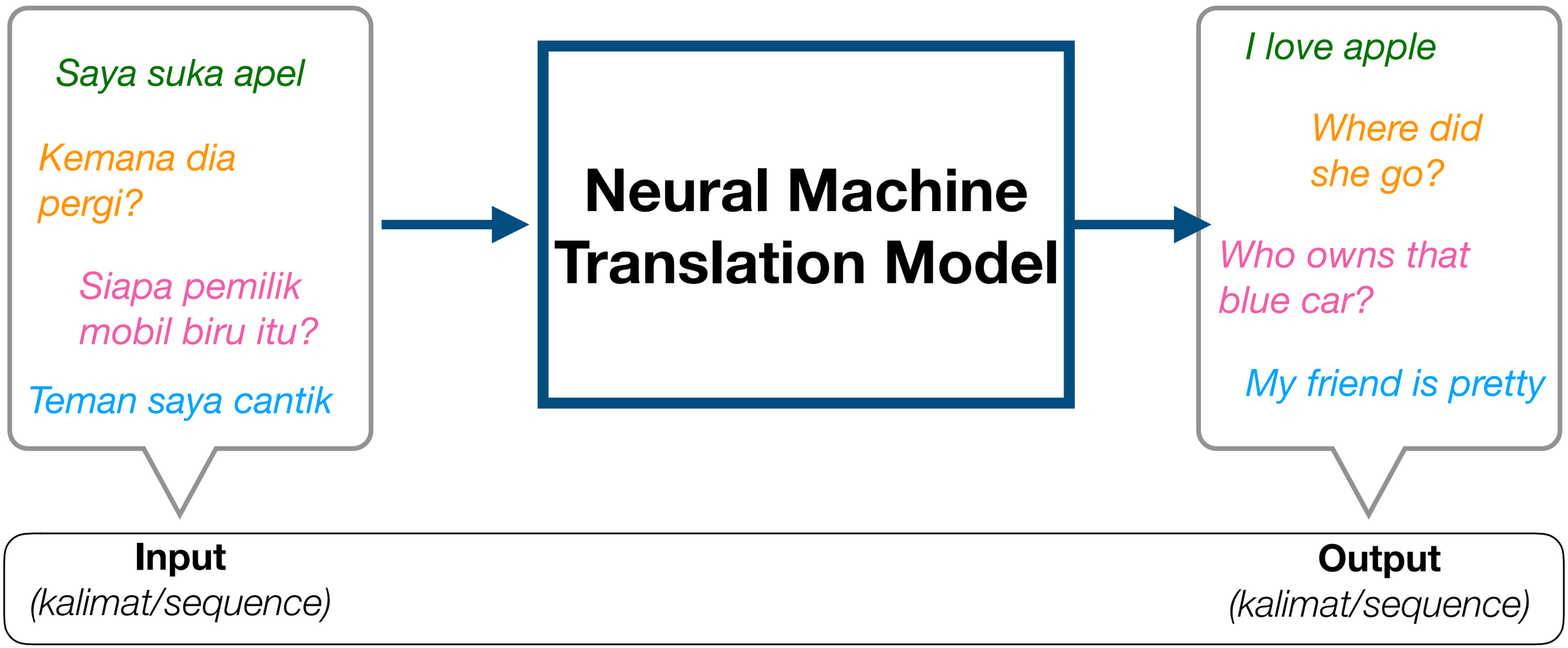
Saya
suka
apel

Deep (machine) Learning
(Neural Machine Translation)

I
love
apple

Machine Learning bisa “belajar sendiri” tentang struktur bahasa dan kata (syntax, context, etc) dalam suatu bahasa dengan mengamati pattern dalam data

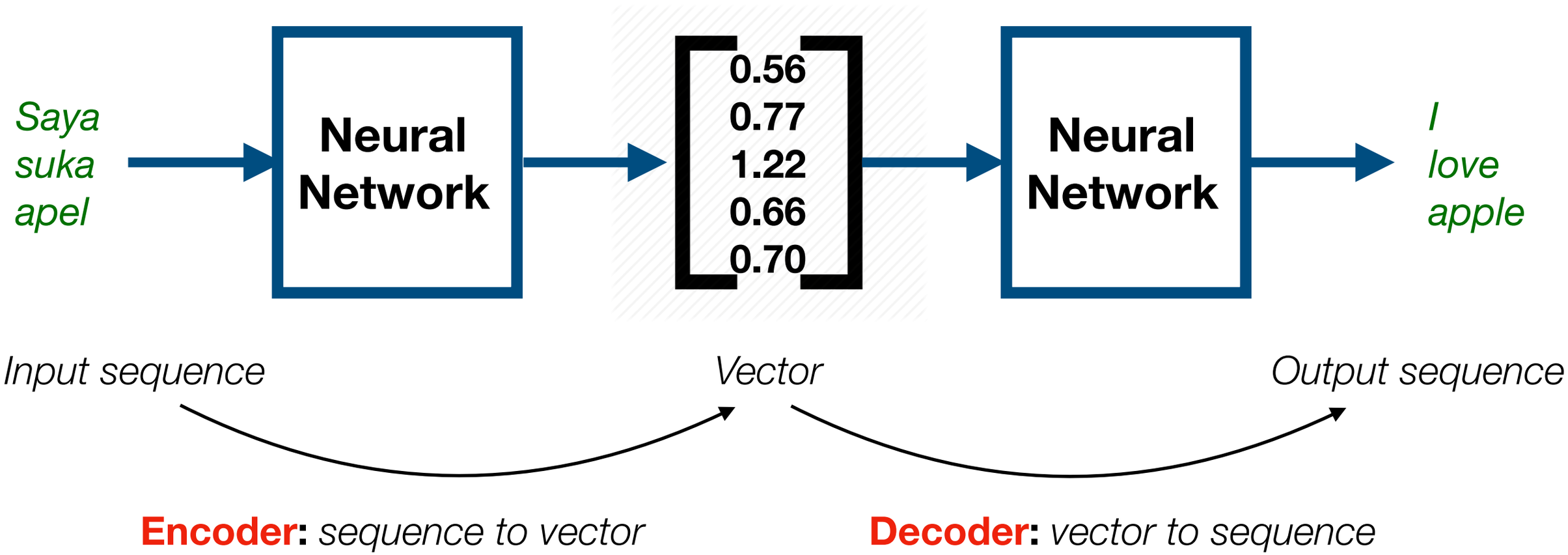
Machine Translation using Neural Network (NN)



Pasangan input-output ini disebut dengan ***samples / training data***

NN belajar rule translasi secara otomatis dengan mengamati jutaan *samples*

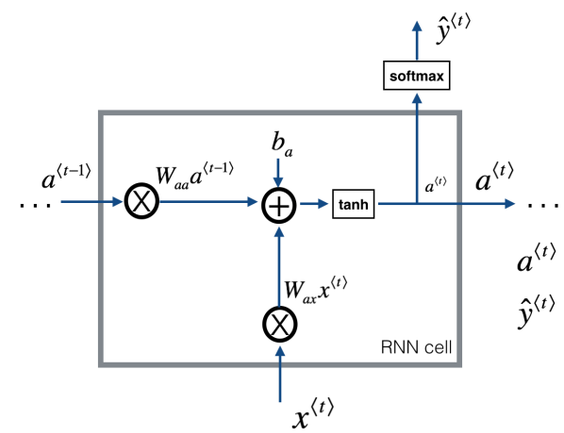
Neural Network for Machine Translation:
Encoder-Decoder architecture



Komputer/mesin tidak bisa memahami kalimat seperti manusia, jadi NN mengubah input menjadi angka (vektor) terlebih dahulu baru bisa diproses

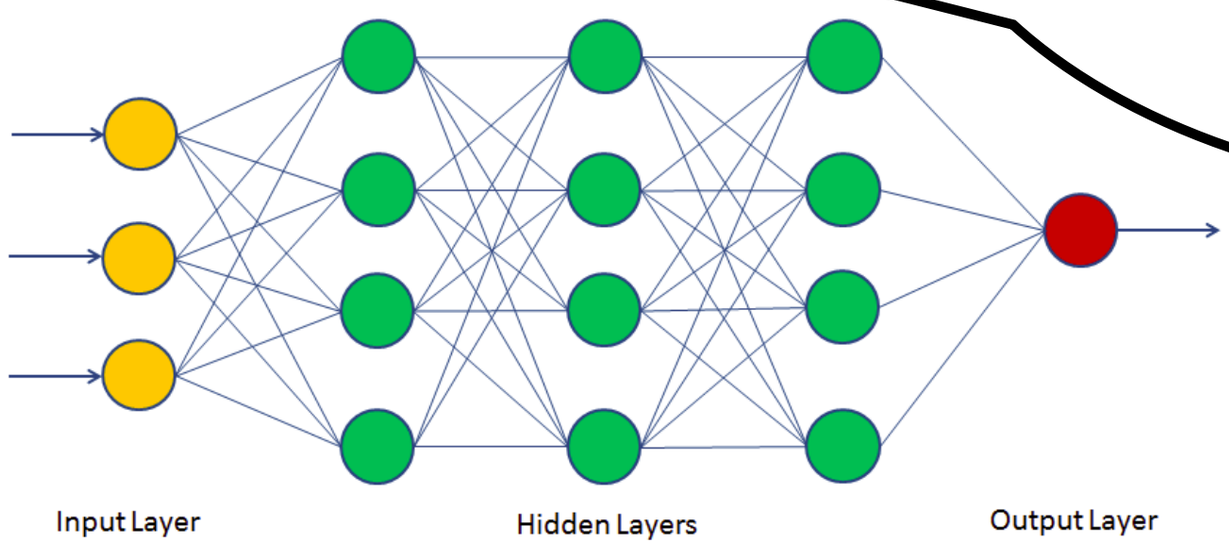
Inside Neural Network?

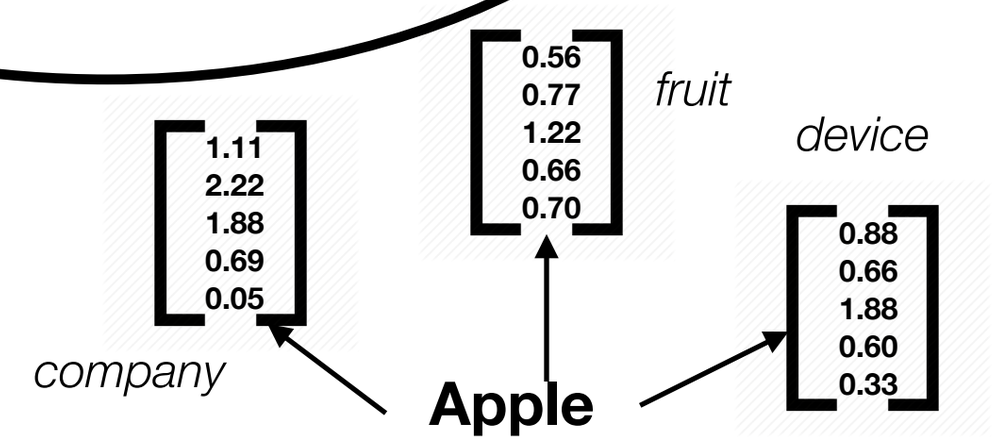
A bunch of mathematical functions!



$$a^{\langle t \rangle} = \tanh(W_{ax}x^{\langle t \rangle} + W_{aa}a^{\langle t-1 \rangle} + b_a)$$

$$\hat{y}^{\langle t \rangle} = soft\max(W_{ya}a^{\langle t \rangle} + b_y)$$

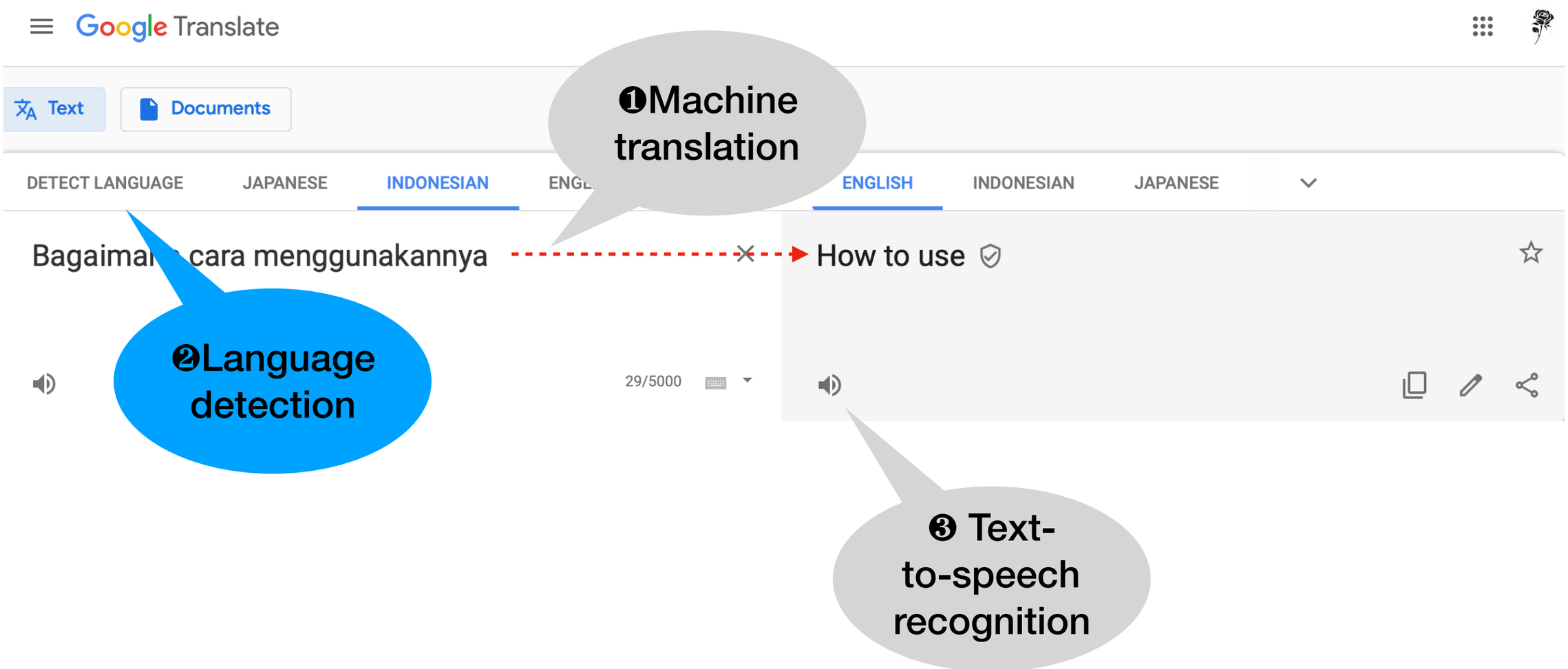




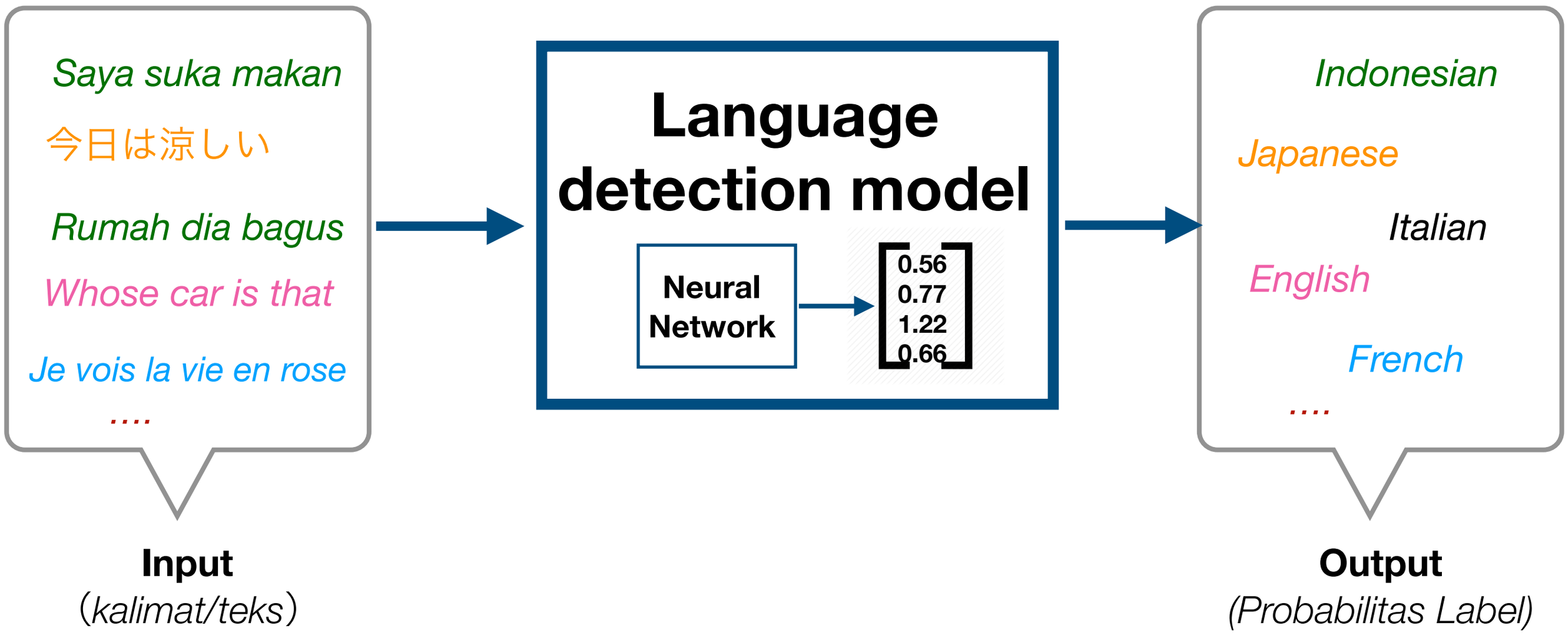
NN mengubah (*mathematical transformation*) vektor dari satu vektor ke vektor yang lain (berkali-kali), bertujuan untuk mencari "representasi" yang paling tepat

“Bedah” Google Translate (2)

Feature “Detect Language” di Google Translate



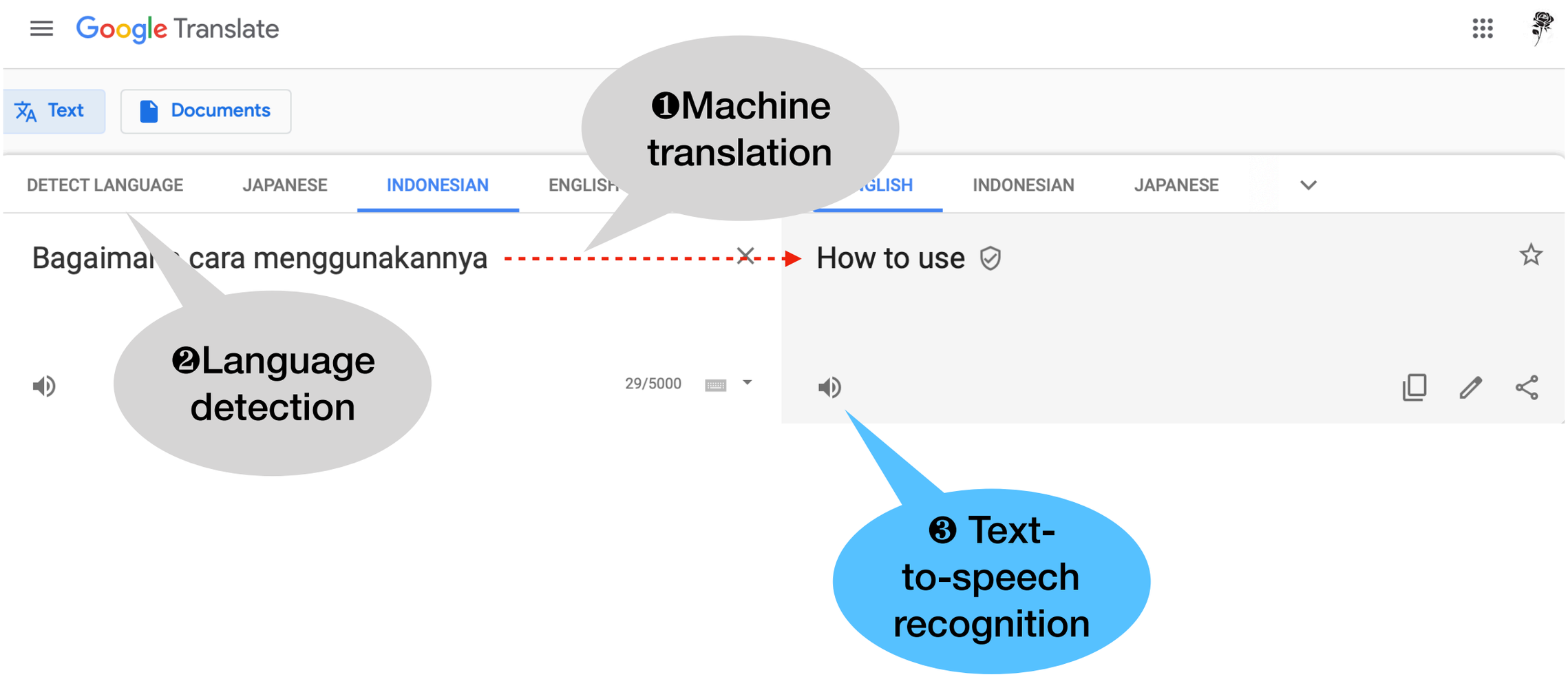
Language Detection: Classification Problem



Language Detection adalah sebuah classification problem, i.e. dimana input sistem adalah kalimat/text dengan output berupa label (kategori) dari input

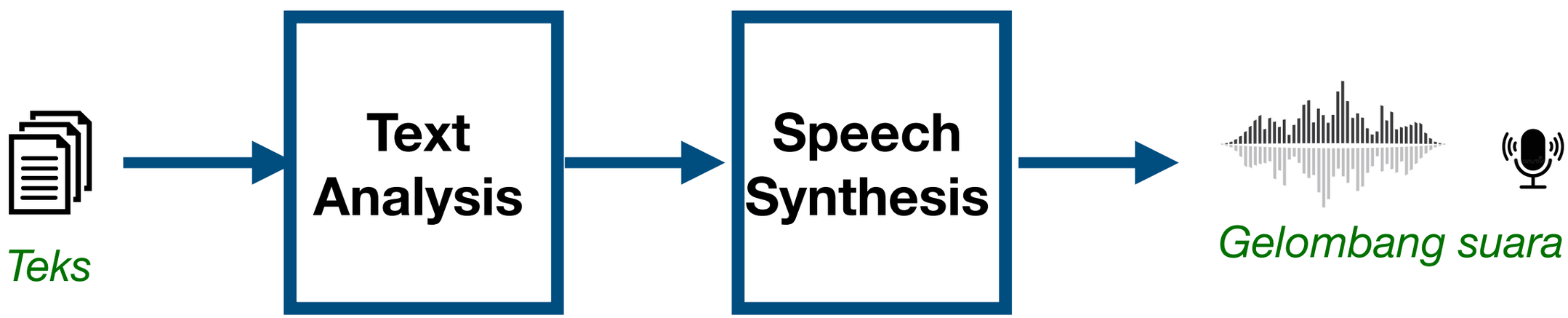
“Bedah” Google Translate (3)

NLP dipadukan dengan task AI yang lain?

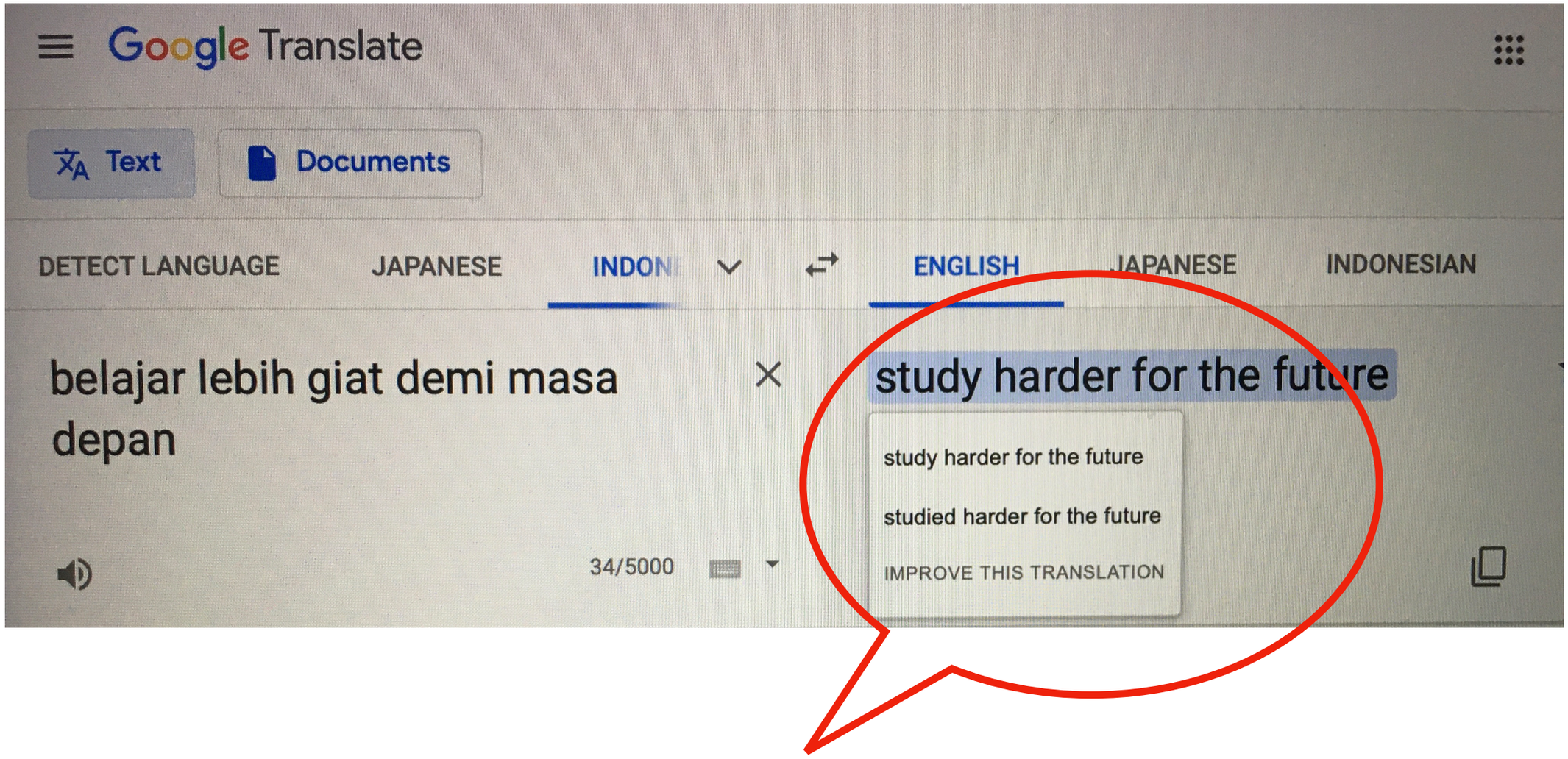


Text-To-Speech Recognition

- Sistem Text-To-Speech mengubah bahasa natural (bahasa manusia) ke dalam bentuk ucapan (suara)
- Perpaduan dari **NLP dan Speech Processing**

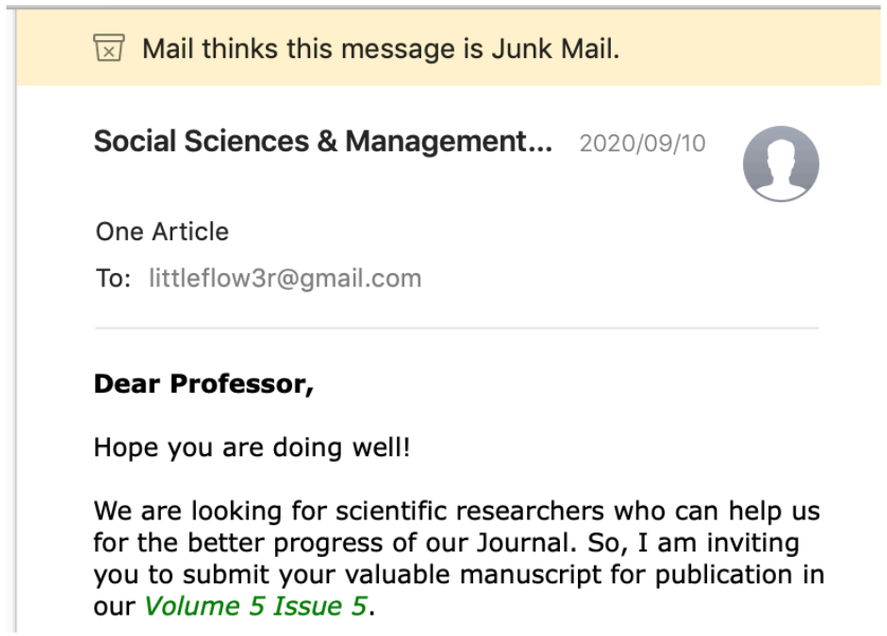


Masih tentang Google Translate



Google Translate effortlessly menggunakan user untuk “memperbaiki” data mereka, untuk meningkatkan akurasi translasi

Other NLP applications:
Email SPAM detector



input

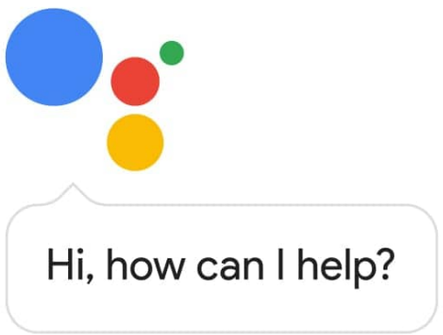
Text email

output

SPAM?
NON-SPAM?

Probabilitas spam/non-spam

Other NLP applications:
Virtual Assistant



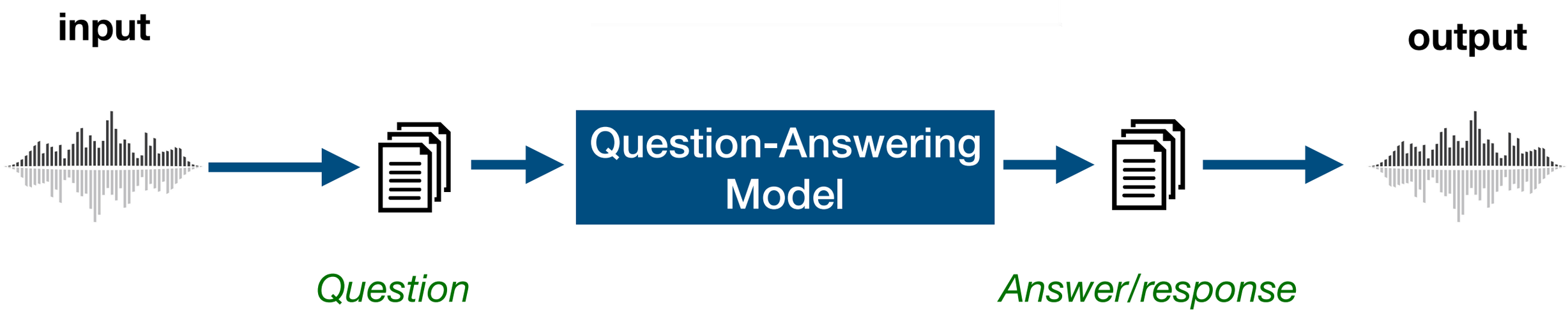


*image taken from the respective brand's website

Other NLP applications: Automatic Question Generation

Topik thesis saya!

(2) reading passage

As the demands of communication across diverse communities have been developing in recent years, the use of English as the main international language has increased to enable interaction between different societies both in business and academic settings. Owing to this, English proficiency tests such as TOEFL and TOEIC are imperative in measuring English communication skill of non-native English speakers. In addition, from teachers' perspective, having know … can personalise the material … udy by Cotton (1988), classro … conducting questioning/t … as a teaching method.

Despite the importance … d English tests mentioned … learners can only rely on … oks. Manual construction of such questions is expensive and requires high level of skill. Recent research has investigated how natural language processing can contribute to automatically generate questions, and this kind of research has received immense attention lately. Automatic question generation can be a breakthrough since it can produce as many questions as possible automatically. Producing the reading passage from Internet materials enables us to provide learners with fresh, updated, and high-quality English reading passages.

(1) target word

She was a bright young PhD graduate from Yale University, and …

The word "bright" in paragraph 2 is closest in meaning to

(A) smart ← (3) correct answer

(B) cheerful and lively

(C) dazzling

(D) valuable

(4) distractors

-end-