



# INTRO TO NLP

THEANO 3.34 - KASIA PRANKE

**WHAT IS  
NATURAL LANGUAGE  
PROCESSING?**

LINGUISTICS

COMPUTER  
SCIENCE

INFORMATION  
ENGINEERING

ARTIFICIAL  
INTELLIGENCE

NLP



# EXAMPLES

# MACHINE TRANSLATION

Google Translate

Dutch ▼ ↔ English ▼

De bezorger  
komt langs  
tussen 15:00 en  
19:00 uur.

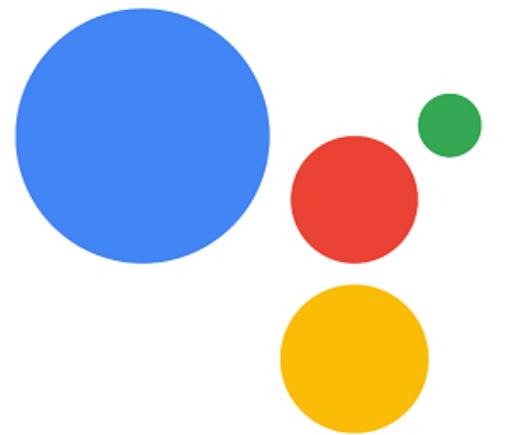
The delivery person  
comes by between  
3:00 PM and 7:00  
PM.



# SPEECH RECOGNITION QUESTION ANSWERING



Hey Siri



Google Assistant

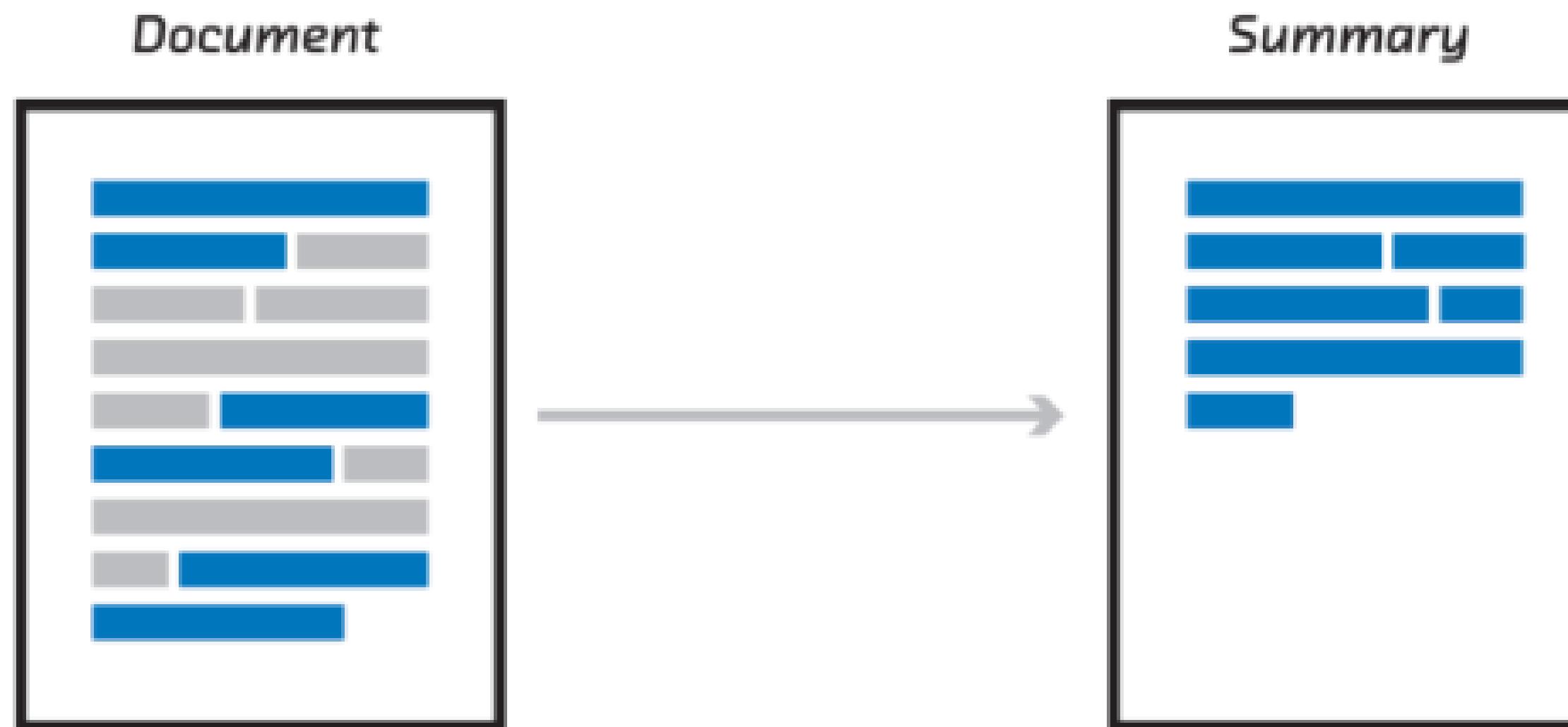


# SPEECH RECOGNITION QUESTION ANSWERING

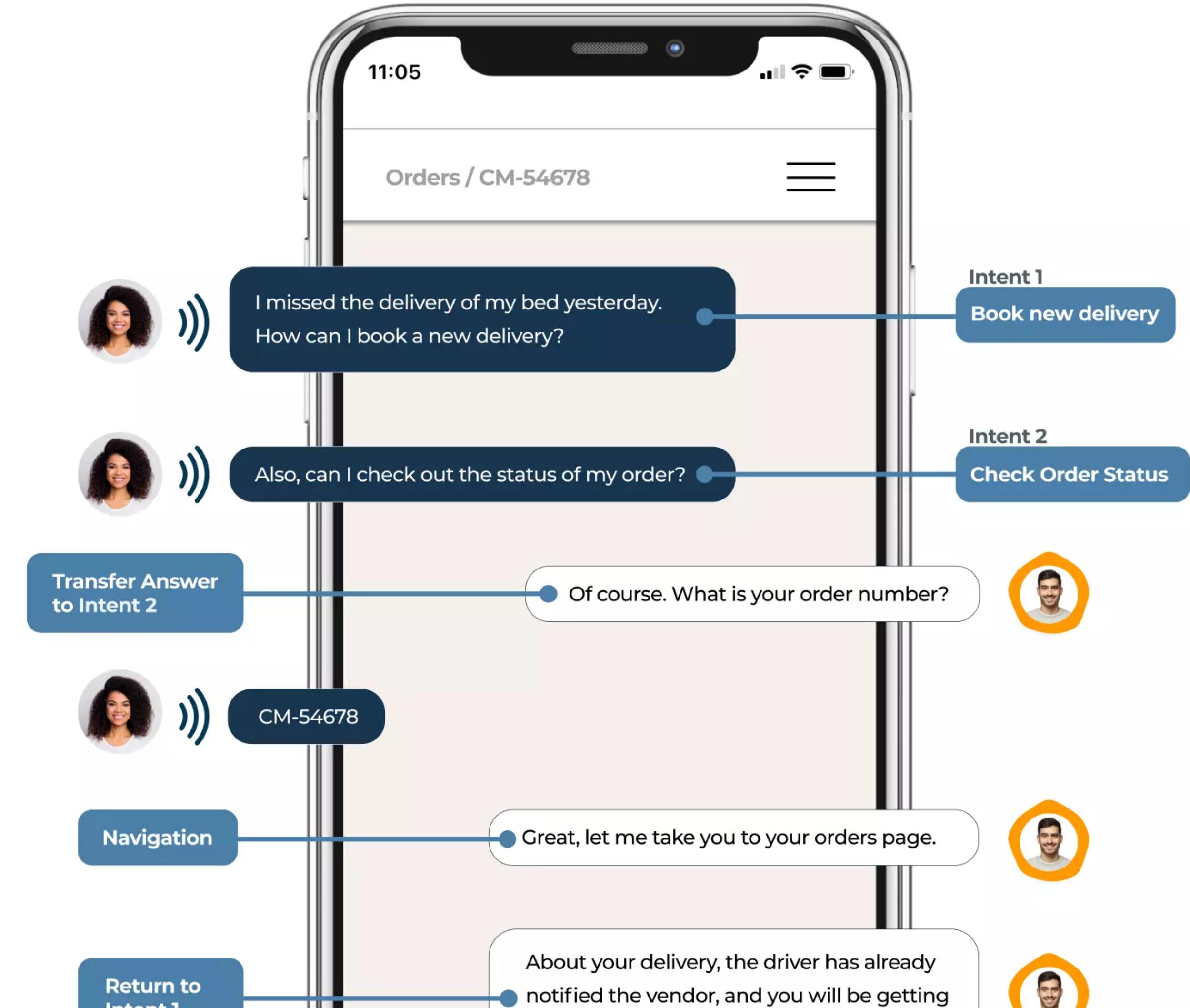


**Check additional resource:** <https://famousgrey.com/work/recruiting-car>

# TEXT SUMMARIZATION



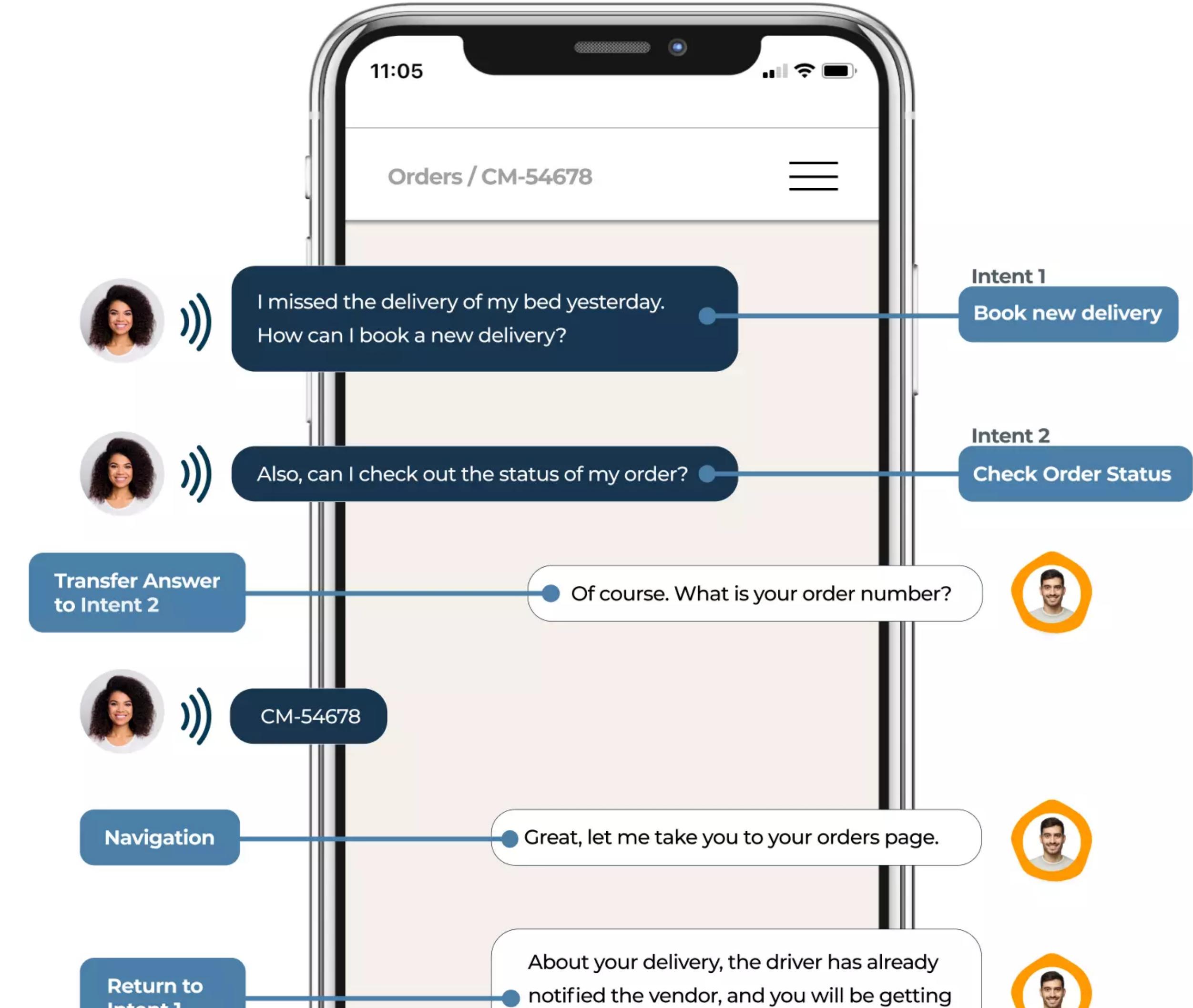
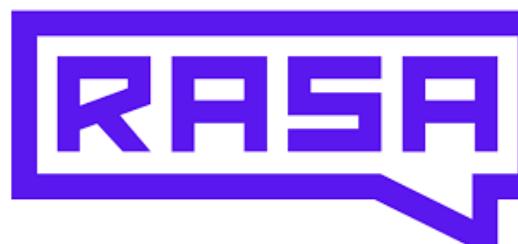
# CHATBOTS VOICEBOTS



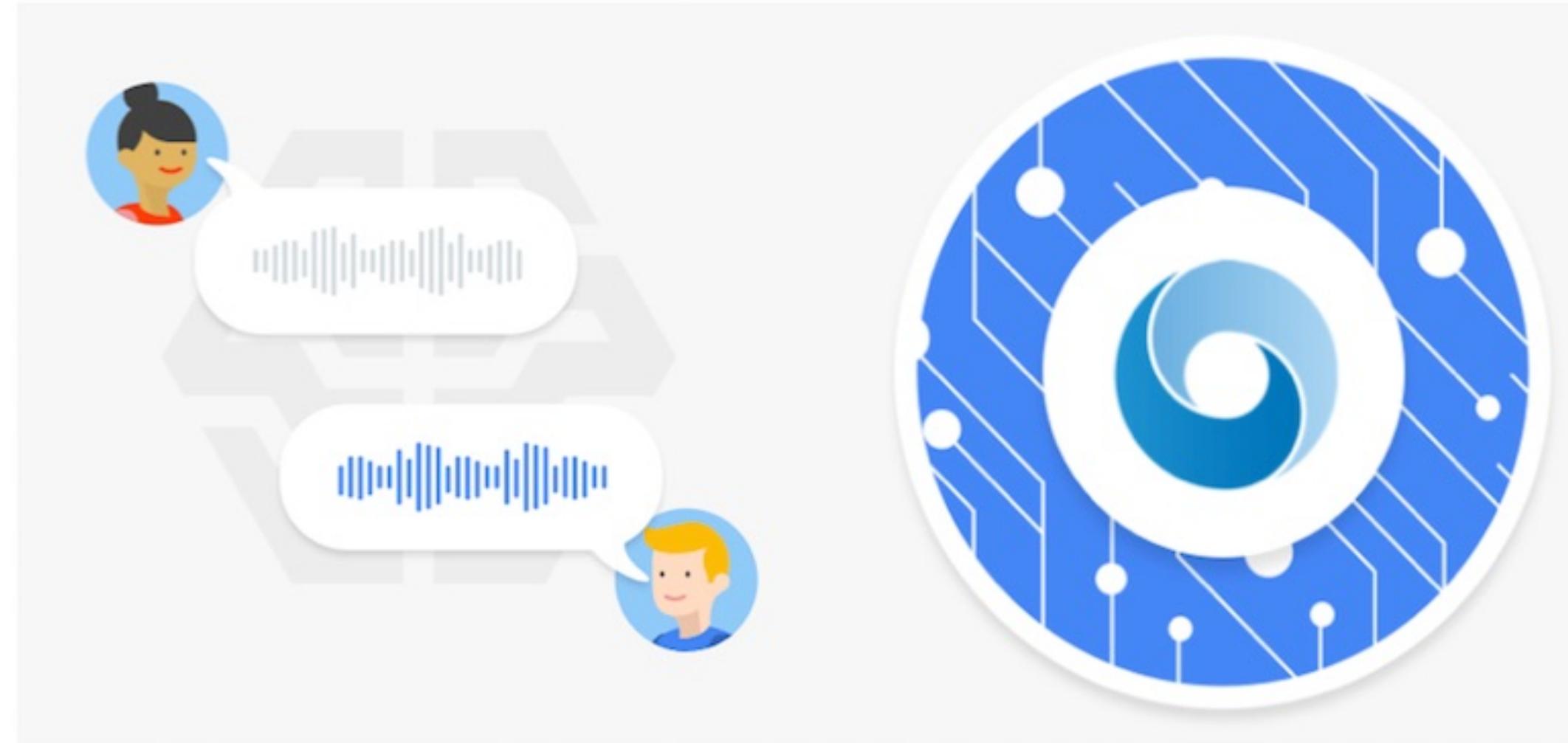
Source: <https://www.voiq.com/>

# CHATBOTS VOICEBOTS

spaCy

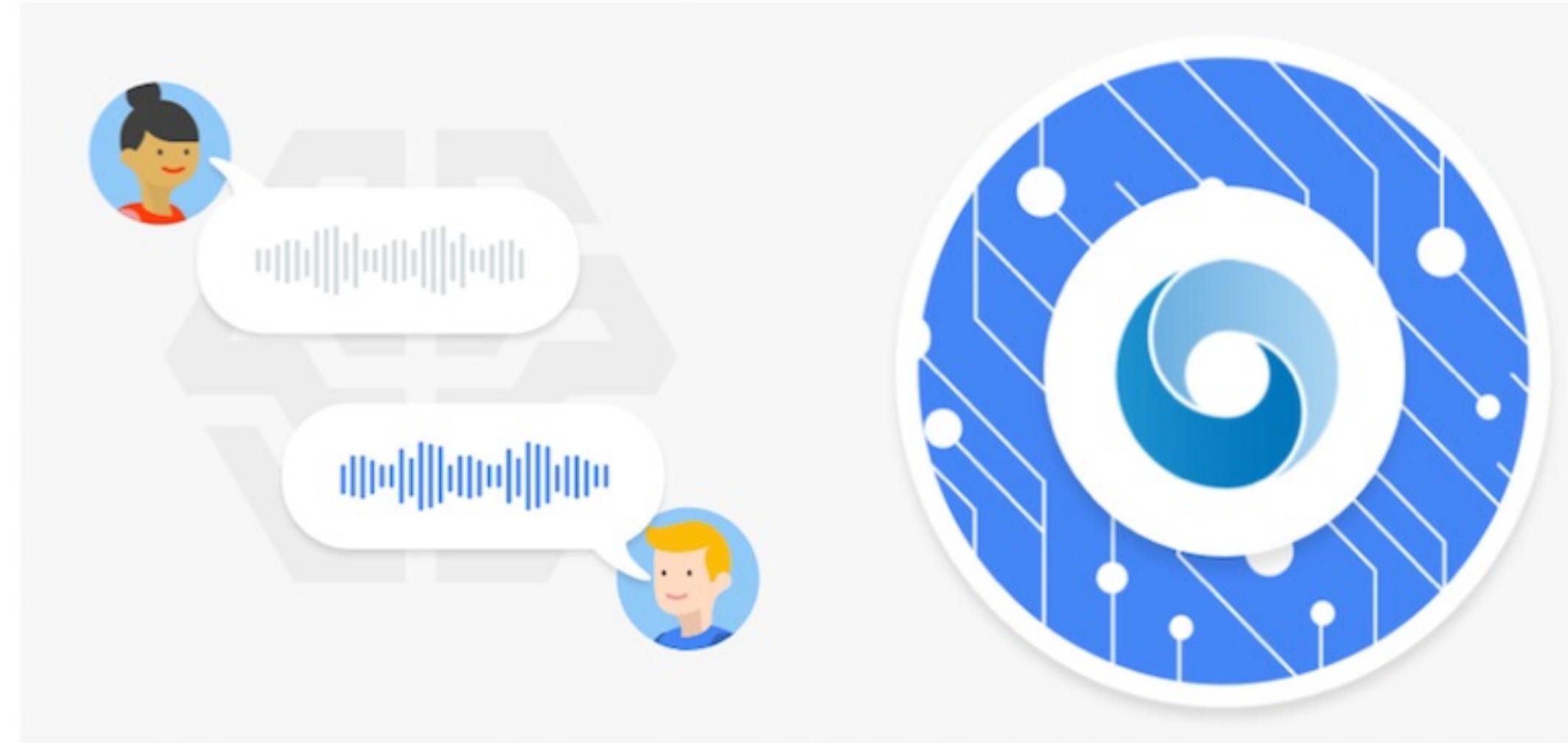


# TEXT-TO-SPEECH AND SPEECH-TO-TEXT



Google Cloud TEXT-TO-SPEECH BETA

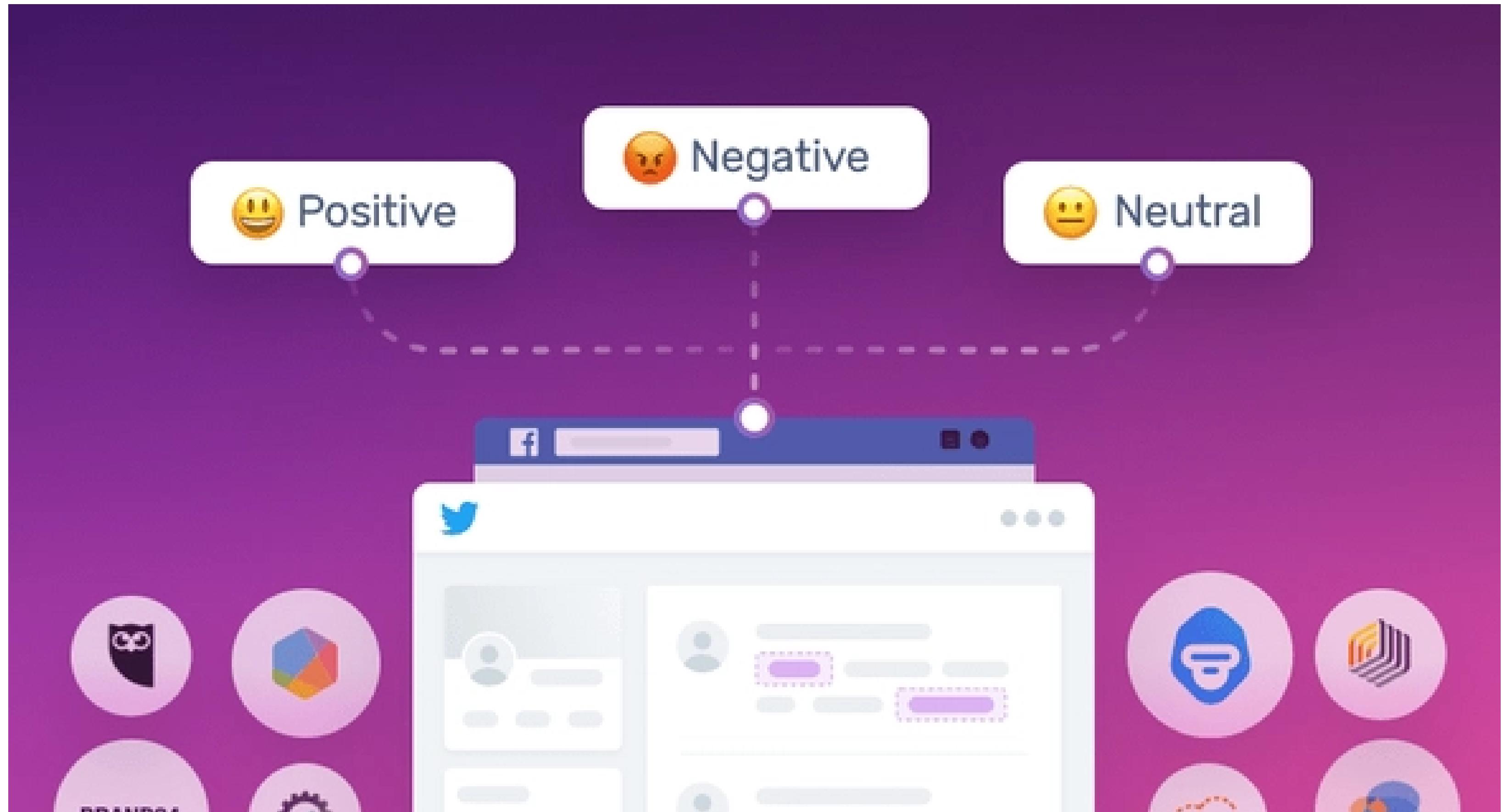
# TEXT-TO-SPEECH AND SPEECH-TO-TEXT



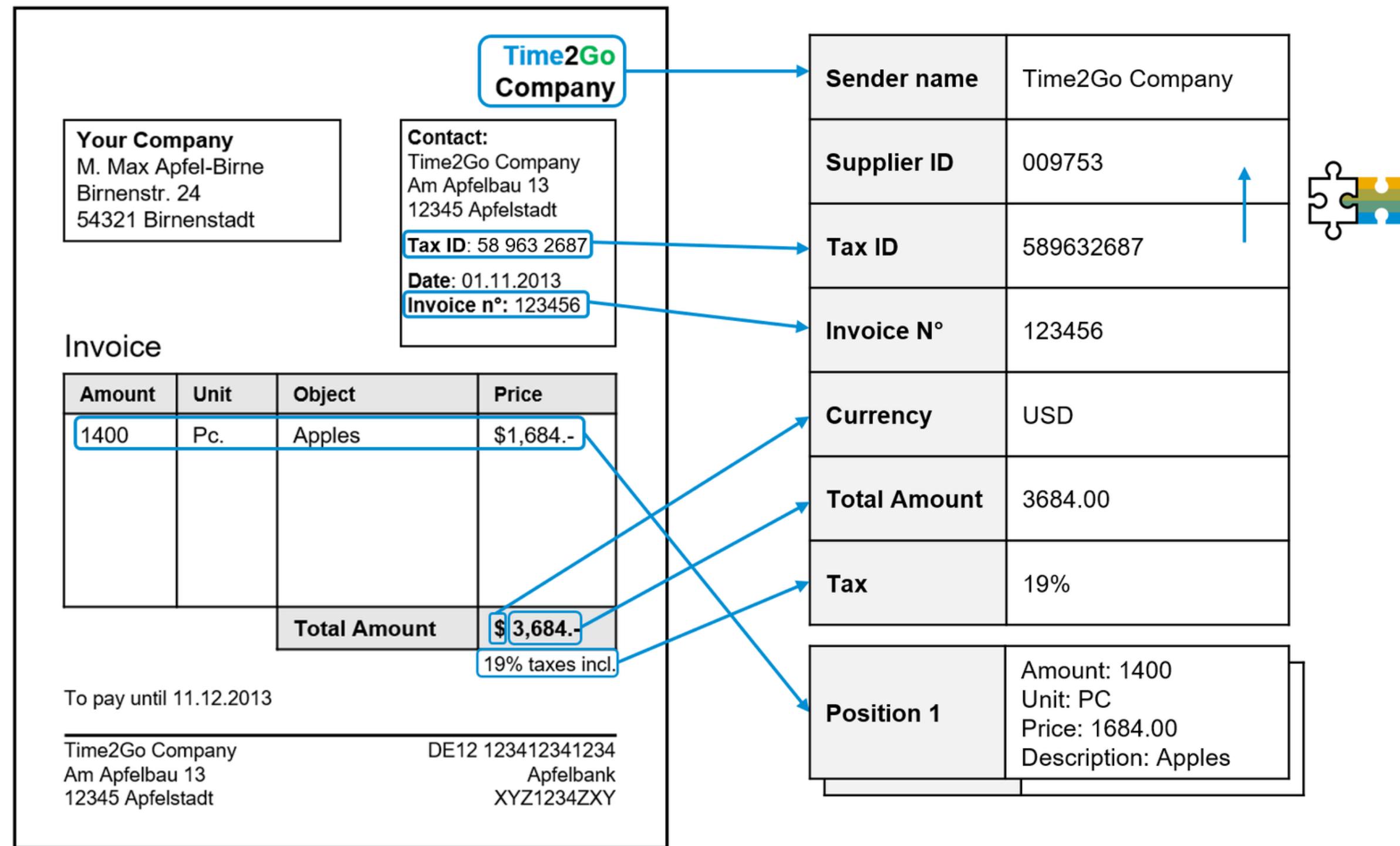
 Google Cloud TEXT-TO-SPEECH BETA

**Check additional resource:** <https://cloud.google.com/text-to-speech>

# SENTIMENT ANALYSIS



# INFORMATION EXTRACTION



**FIRST THINGS**

**FIRST...**

# TOKENIZATION

spaCy

```
import spacy

nlp = spacy.load("en_core_web_sm")

doc = nlp("Apple is looking at buying U.K. startup for $1 billion")
for token in doc:

    print(token.text)
```

0	1	2	3	4	5	6	7	8	9	10
Apple	is	looking	at	buying	U.K.	startup	for	\$	1	billion

Source: <https://spacy.io/usage/linguistic-features#tokenization>

# LEMMATIZATION

```
import spacy

# English pipelines include a rule-based lemmatizer
nlp = spacy.load("en_core_web_sm")
lemmatizer = nlp.get_pipe("lemmatizer")
print(lemmatizer.mode) # 'rule'

doc = nlp("I was reading the paper.")
print([token.lemma_ for token in doc])
# ['I', 'be', 'read', 'the', 'paper', '.']
```

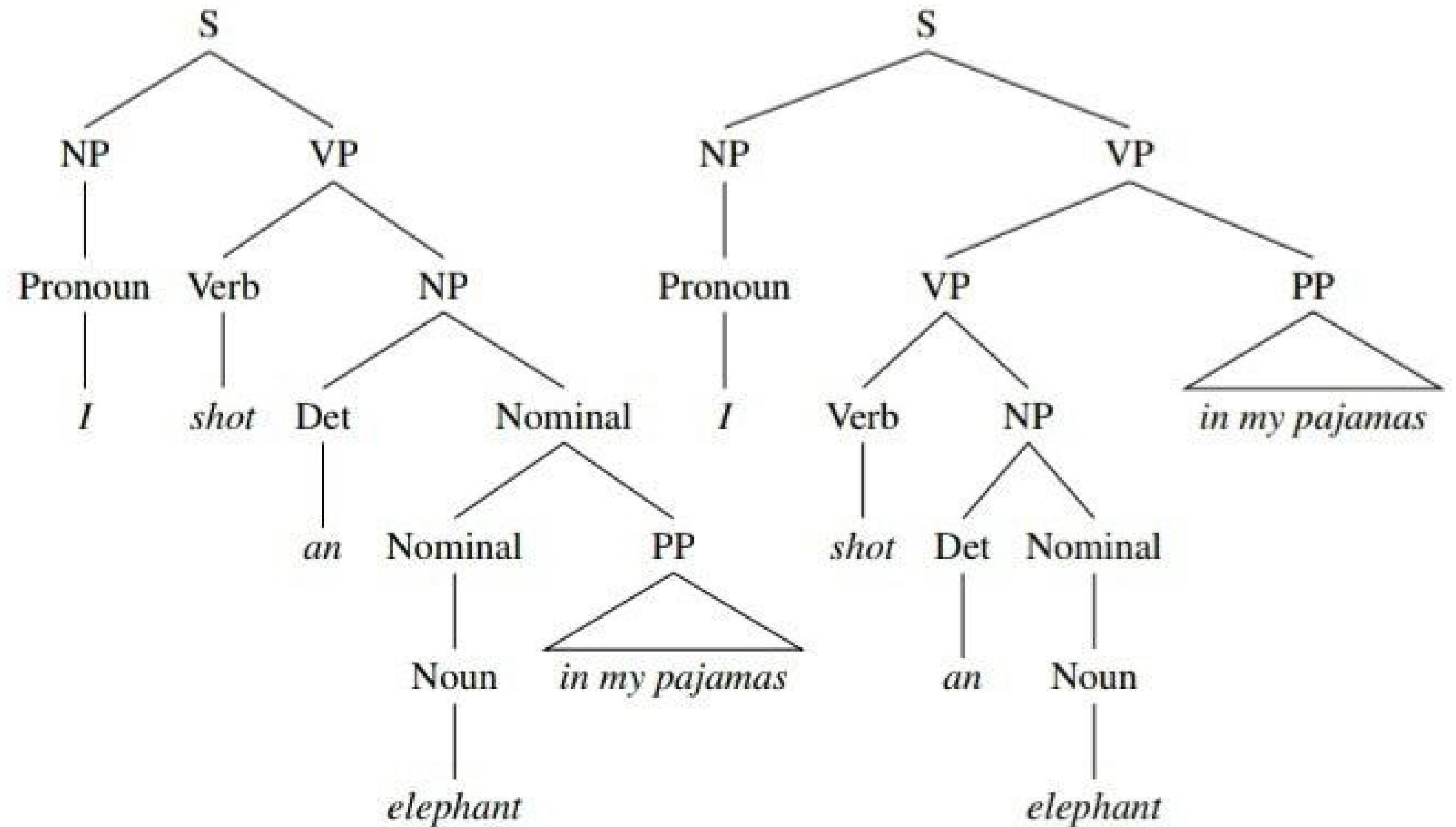
# PART-OF-SPEECH TAGGING (POS)



TEXT	POS
Apple	PROPN
is	AUX
looking	VERB
at	ADP
buying	VERB
U.K.	PROPN
startup	NOUN
for	ADP
\$	SYM
1	NUM
billion	NUM

Source: <https://spacy.io/usage/linguistic-features#pos-tagging>

# DEPENDENCY PARSING



# NAMED ENTITY RECOGNITION (NER)

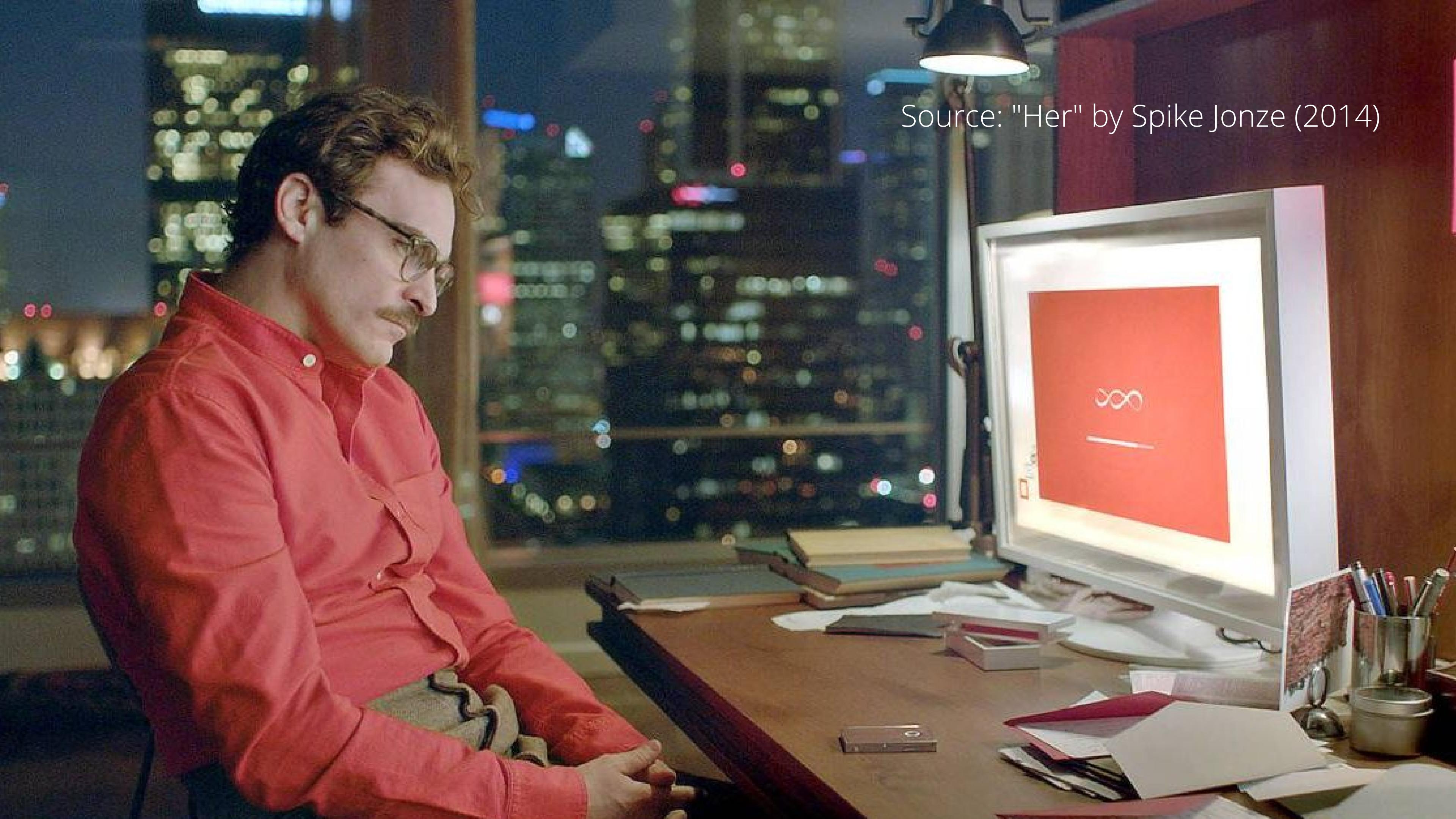
When Sebastian Thrun **PERSON** started working on self-driving cars at Google **ORG** in 2007 **DATE**, few people outside of the company took him seriously.

Apple **ORG** is looking at buying U.K. **GPE** startup for \$1 billion **MONEY**

# HOPE FOR THE FUTURE

A composite image with a dark, moody atmosphere. On the left, a man in a dark spacesuit with a glowing blue visor looks towards the center. In the center, a woman with short blue hair sings into a microphone on a stage. To the right, a large, detailed Earth-like planet hangs in a dark space with wispy clouds. Below the planet, a futuristic city with glowing buildings and roads stretches across the horizon under a dark, cloudy sky.

Source: [https://www.youtube.com/watch?v=163\\_C5UVU-I](https://www.youtube.com/watch?v=163_C5UVU-I)

A man with curly hair and glasses, wearing a red shirt, sits at a desk in a room decorated with Christmas lights. He is looking at a computer monitor displaying a red interface with a white logo. The desk is cluttered with papers, books, and a lamp.

Source: "Her" by Spike Jonze (2014)



**THANK YOU  
ANY  
QUESTIONS?**

**THEANO 3.34 - KASIA PRANKE**