언어이해기술의 현황과 활용 사례

Presented by

Kanghak Kim

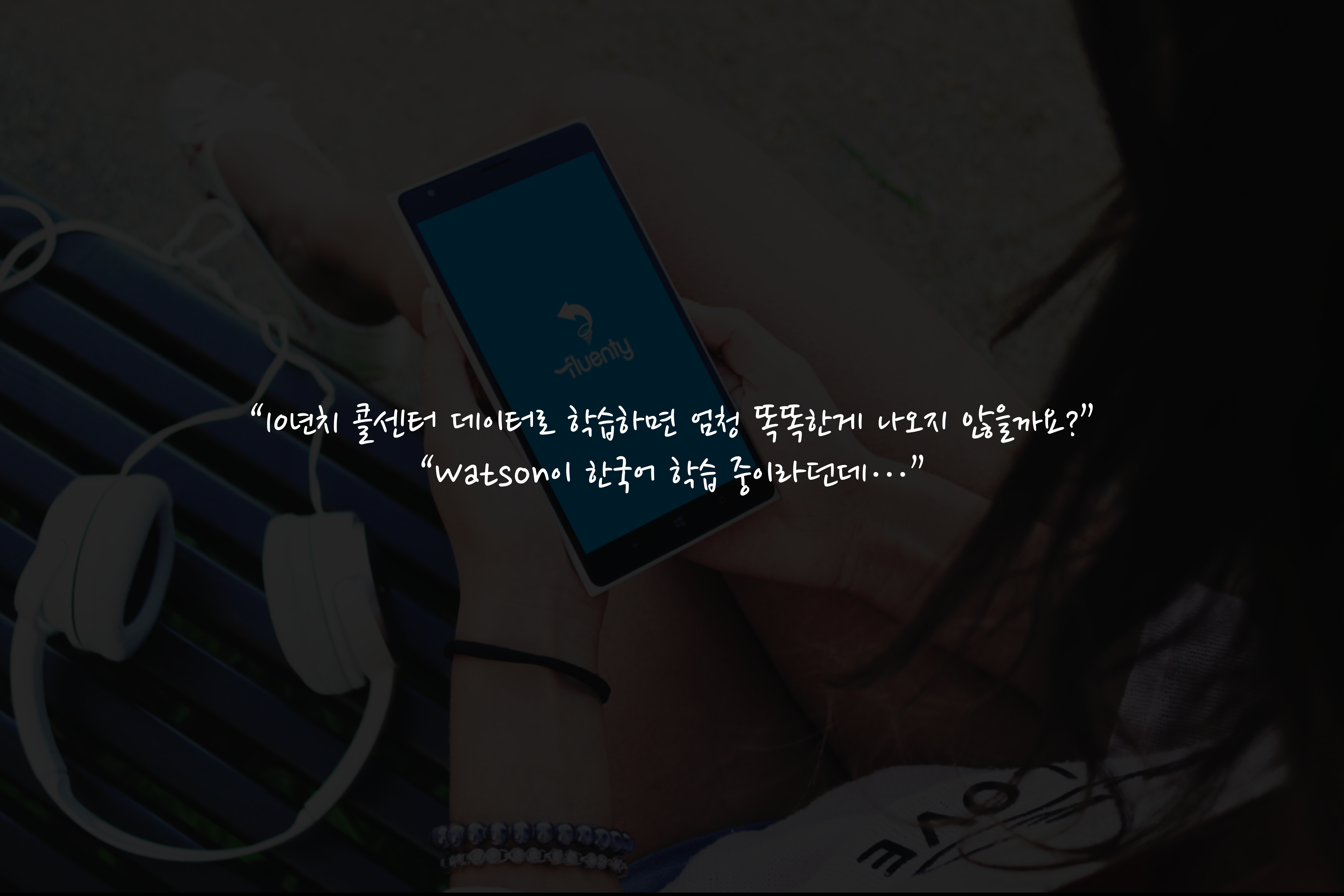
CEO, Co-founder

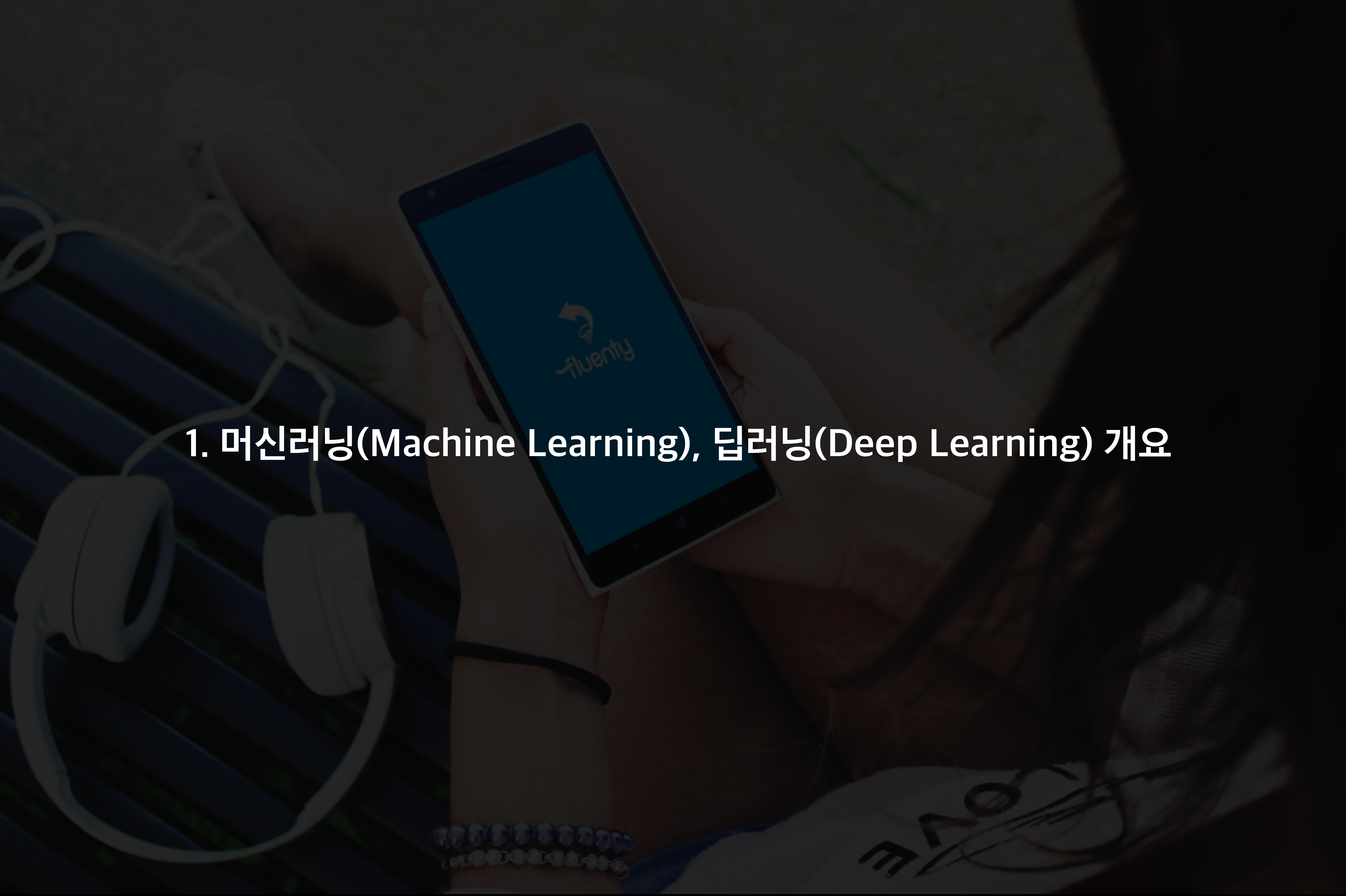
kanghak@fluenty.co

fluenty

발표 순서

1. 머신러닝, 딥러닝 개요 (Supervised learning을 중심으로)
2. 챗봇의 등장 배경:
    - 기술적 배경: 언어이해분야에서 딥러닝 발전
    - 산업적 배경
3. 대화형 인공지능 현황
    - Q&A / Small Talk
    - Goal-Based Actions
    - 대화형 인공지능에서의 자연어이해기술(NLU)
4. Demonstration
5. Q&A


fluenty
“10년치 콜센터 데이터로 학습하면 엄청 똑똑한게 나오지 않을까요?”
“Watson이 한국어 학습 중이라던데...”



1. 머신러닝(Machine Learning), 딥러닝(Deep Learning) 개요

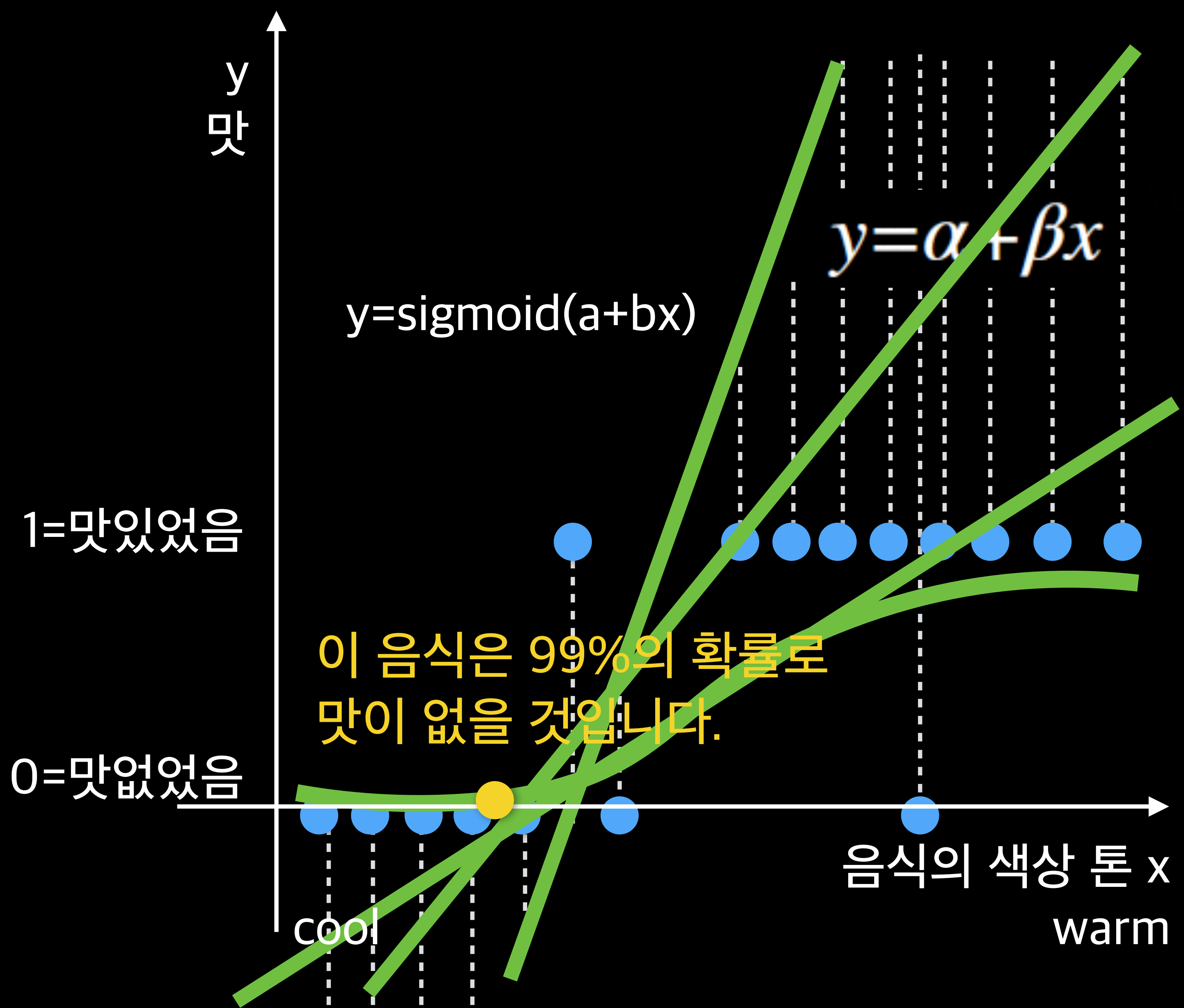
퀴즈: “알파고의 소스코드는 몇 줄일까요?”

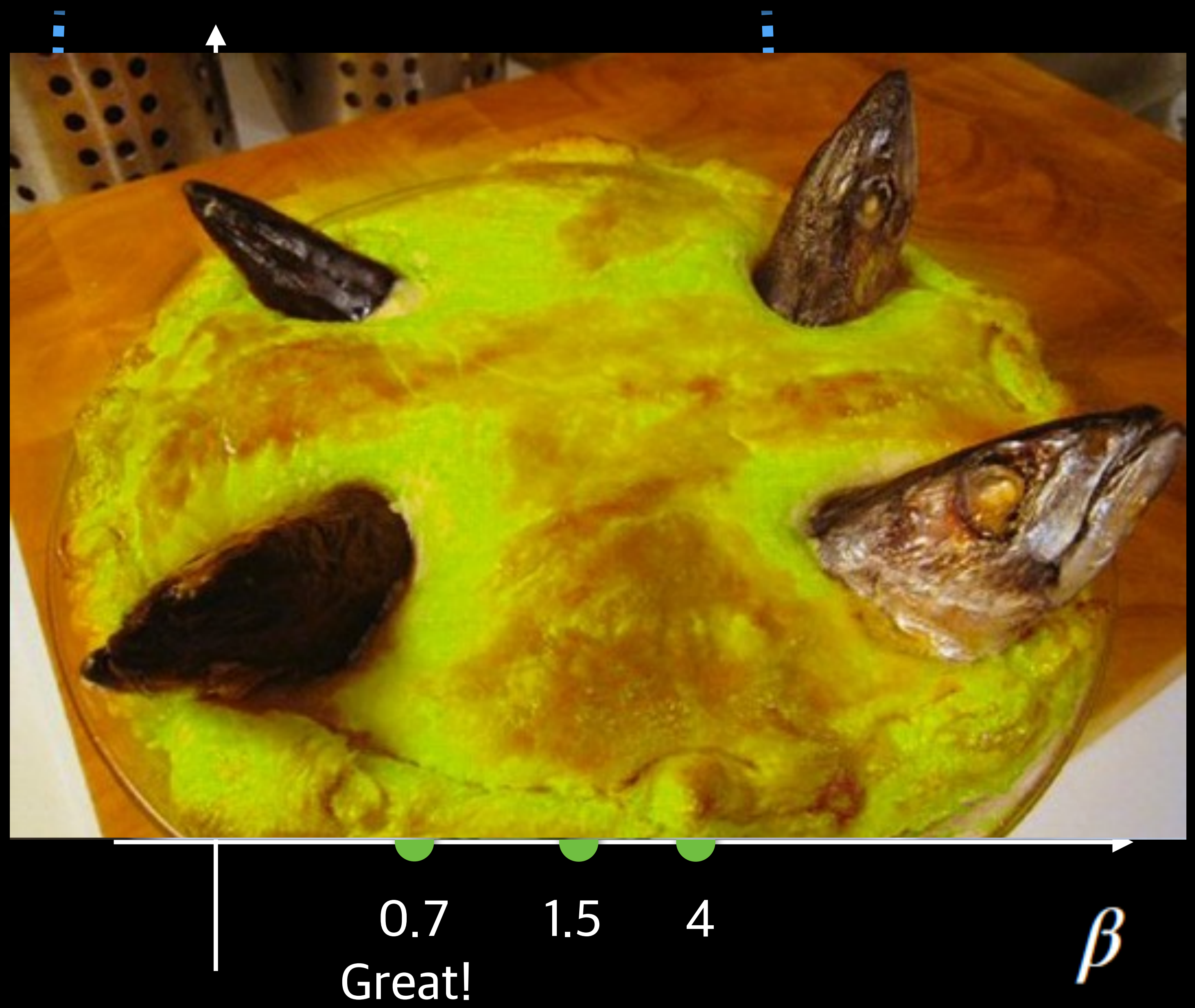
"The programming paradigm is changing. Instead of programming a computer, **you teach a computer to learn something** and it does what you want."

Eric Schmidt, Chairman of Alphabet

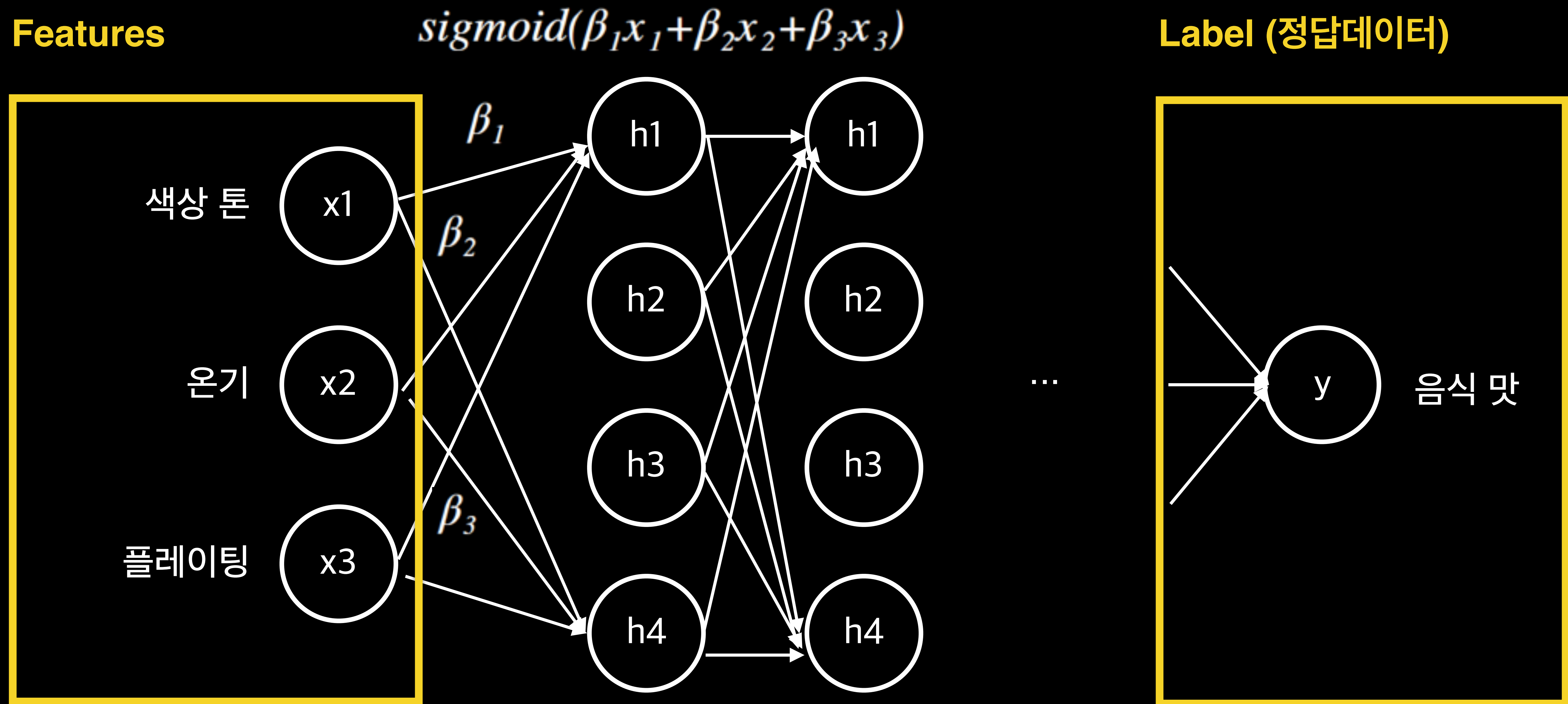
Teaching a computer to learn something?

예시로 보는 학습의 기본 구조





Deep Neural Network (DNN)의 기본 개념



Speech Recognition: Word Error Rate (WER) 30% 이상 감소

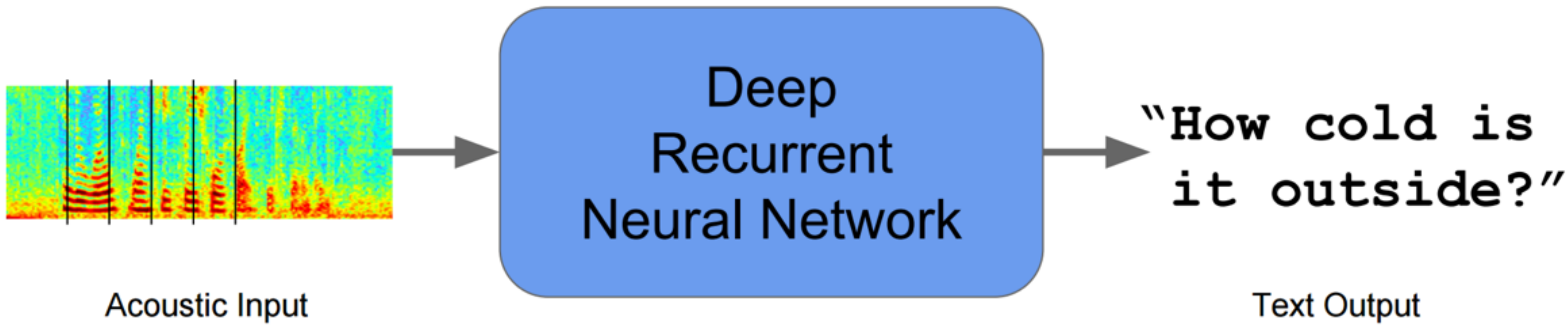


Image Recognition: 사람보다 나은 결과 달성

ImageNet Challenge

Given an image, predict one of 1000 different classes

Image credit: www.cs.toronto.edu/~fritz/absps/imagenet.pdf

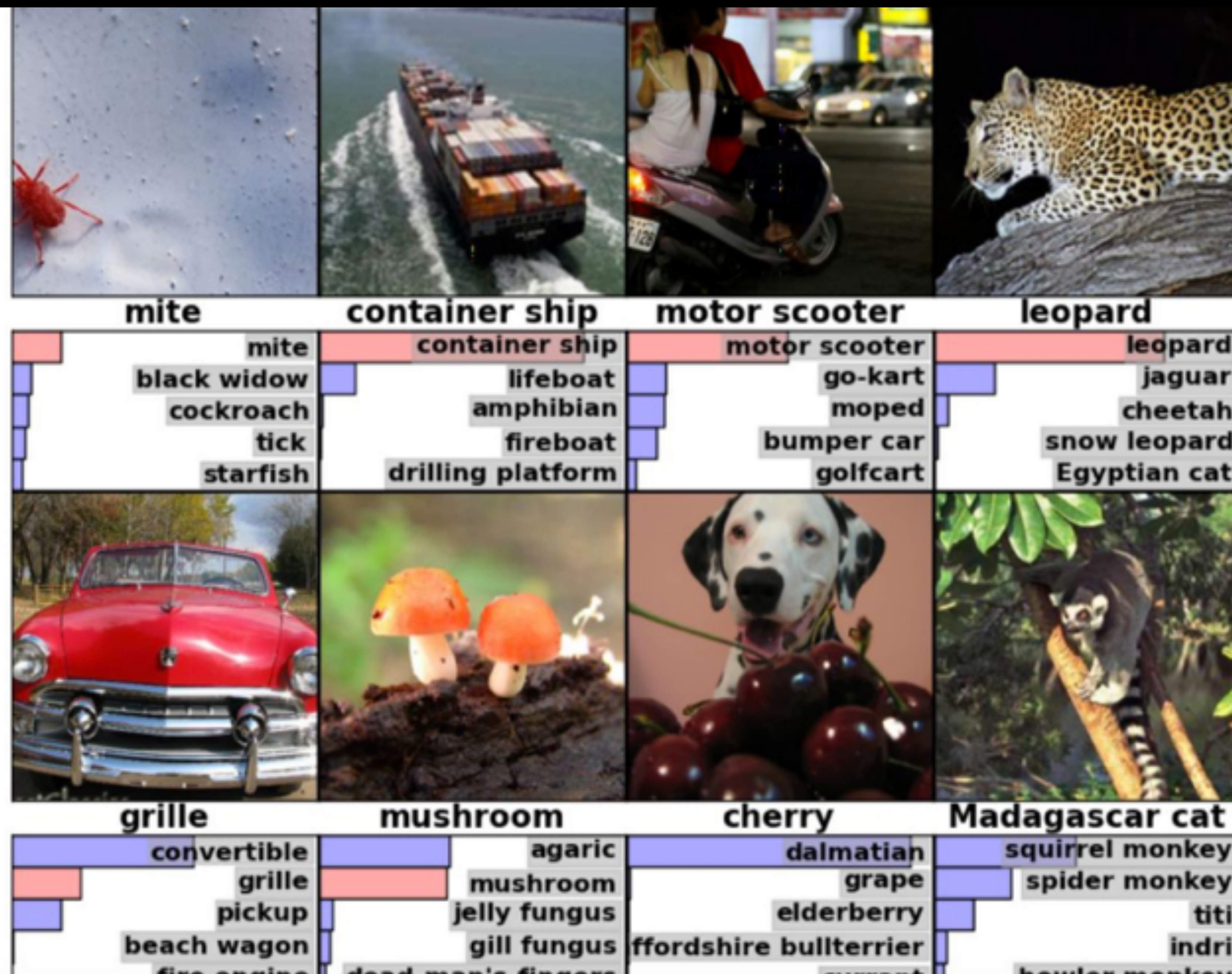


Image Recognition: 사람보다 나은 결과 달성

Team	Year	Place	Error (top-5)
XRCE (pre-neural-net explosion)	2011	1st	25.8%
Supervision (AlexNet)	2012	1st	16.4%
Clarifai	2013	1st	11.7%
GoogLeNet (Inception)	2014	1st	6.66%
Andrej Karpathy (human)	2014	N/A	5.1%
BN-Inception (Arxiv)	2015	N/A	4.9%
Inception-v3 (Arxiv)	2015	N/A	3.46%

ImageNet challenge classification task

자율주행 자동차 예시

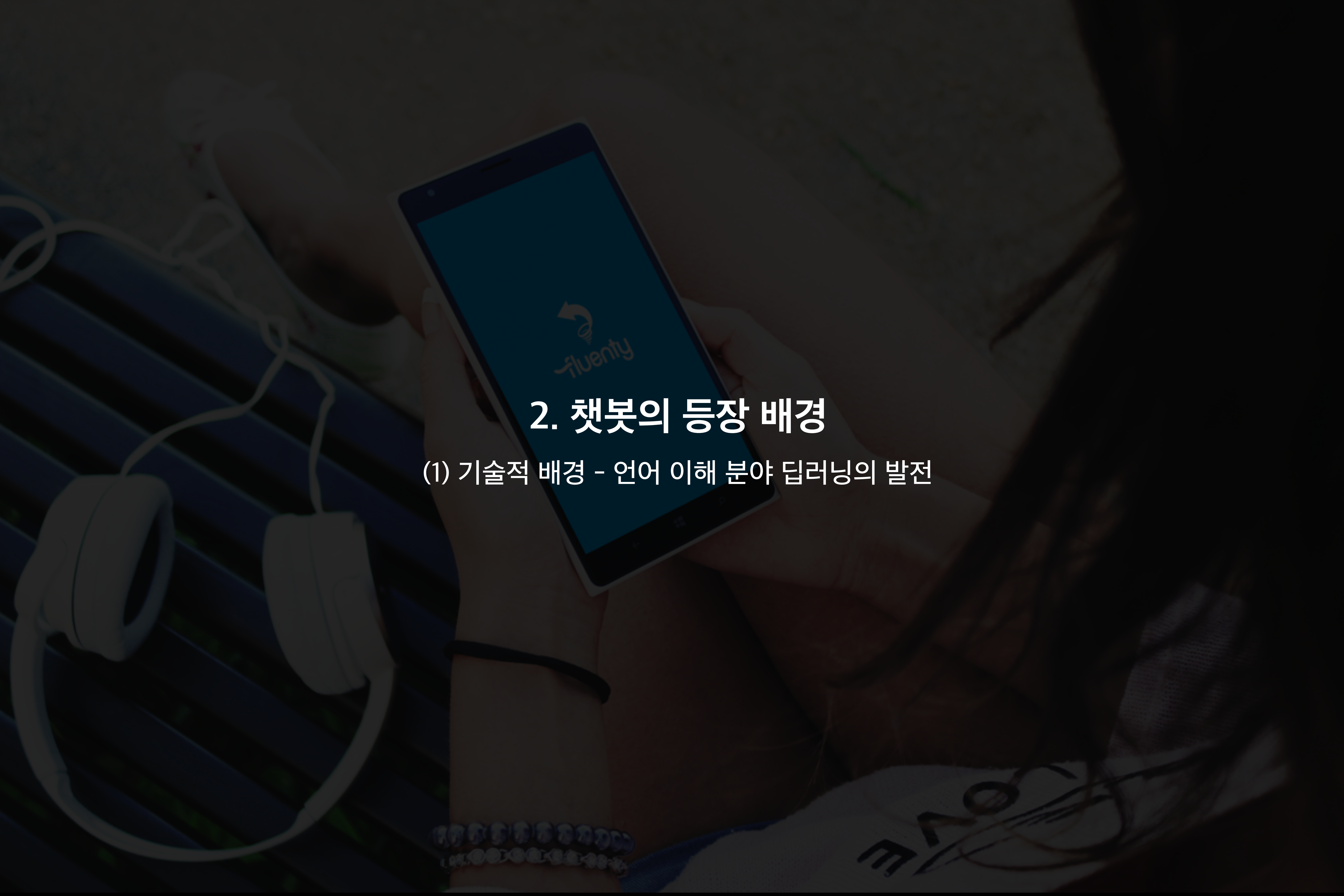

핸들의 각도
Y
Hidden
X
도로의 모양
CATERPILLAR
CAT
GM
Continental
ibeo
vector
blog.gm-korea.co.kr

Supervised Learning:

어떤 정답 데이터로 무엇을 학습할 것인지 사람이 설계해야 함

↓

알아서 학습하지는 않습니다.



2. 챗봇의 등장 배경

(1) 기술적 배경 - 언어 이해 분야 딥러닝의 발전

사람의 말을 이해한다는 것은?

1. 단어, 문장을 연속적인 공간에서 생각하고 이해하는 것

사람은 단어 간의 거리를 무의식적으로 인지

전화해

연락해

카톡해

콜미!

아이오아이

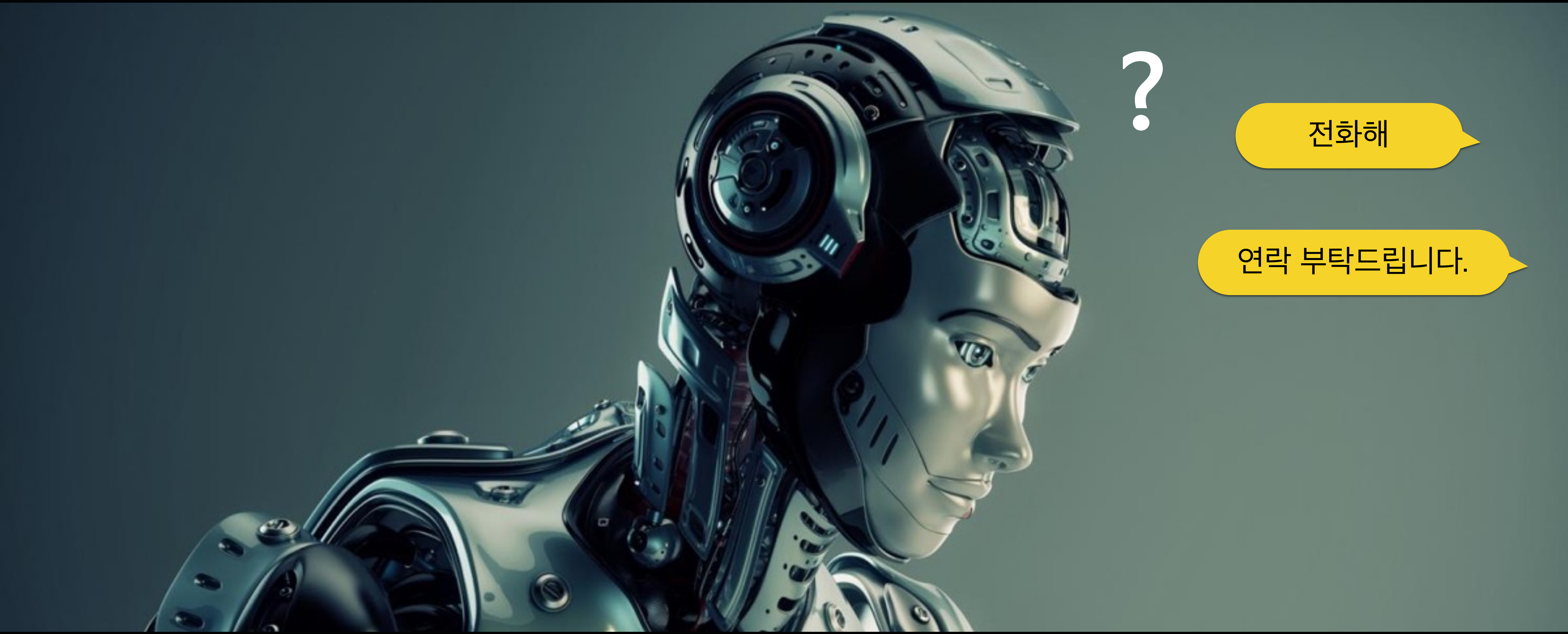
트와이스

레드벨벳

인피니트

여자친구

과거의 컴퓨터: 말이 조금만 달라도 못 알아듣는 한계



새로운 단어를 이해하는 단서: 주변의 맥락

"You shall know a word by the company it keeps"
(John Rupert Firth 1957)

요새 신곡 발표한 다른 아이돌곡은 길거리서 안들리고 트와이스만 들림. 대세긴 대세네

헉 트와이스 신곡 넘좋자나 다들 너무 예뻐

근처 맥락에 등장한 단어들이 "트와이스"를 표현함

Embedding - Word2Vec

Tomas Mikolov

(Research scientist @Facebook AI Research, FAIR)

Efficient Estimation of Word Representations in Vector Space

Tomas Mikolov
Google Inc., Mountain View, CA
tmikolov@google.com

Kai Chen
Google Inc., Mountain View, CA
kaichen@google.com

Greg Corrado
Google Inc., Mountain View, CA
gcorrado@google.com

Jeffrey Dean
Google Inc., Mountain View, CA
jeff@google.com

:1301.3781v3 [cs.CL] 7 Sep 2013

Abstract

We propose two novel model architectures for computing continuous vector representations of words from very large data sets. The quality of these representations is measured in a word similarity task, and the results are compared to the previously best performing techniques based on different types of neural networks. We observe large improvements in accuracy at much lower computational cost, i.e. it takes less than a day to learn high quality word vectors from a 1.6 billion words data set. Furthermore, we show that these vectors provide state-of-the-art performance on our test set for measuring syntactic and semantic word similarities.

1 Introduction

Many current NLP systems and techniques treat words as atomic units - there is no notion of similarity between words, as these are represented as indices in a vocabulary. This choice has several good

Distributed Representations of Words and Phrases and their Compositionality

Tomas Mikolov
Google Inc.
Mountain View
mikolov@google.com

Ilya Sutskever
Google Inc.
Mountain View
ilyasu@google.com

Kai Chen
Google Inc.
Mountain View
kai@google.com

Greg Corrado
Google Inc.
Mountain View
gcorrado@google.com

Jeffrey Dean
Google Inc.
Mountain View
jeff@google.com

Abstract

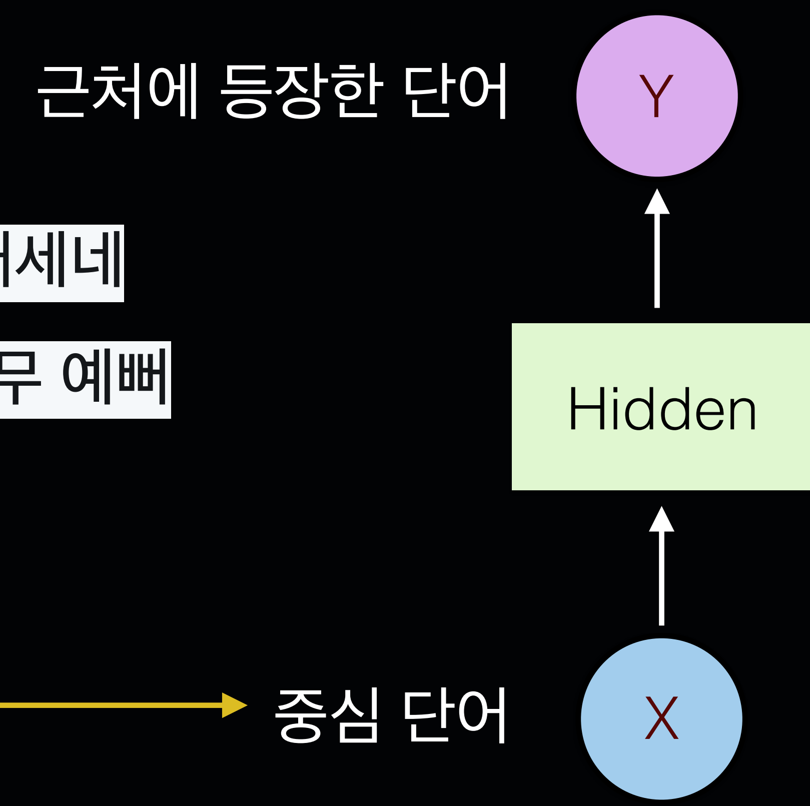
The recently introduced continuous Skip-gram model is an efficient method for learning high-quality distributed vector representations that capture a large number of precise syntactic and semantic word relationships. In this paper we present several extensions that improve both the quality of the vectors and the training speed. By subsampling of the frequent words we obtain significant speedup and also learn more regular word representations. We also describe a simple alternative to the hierarchical softmax called negative sampling.

An inherent limitation of word representations is their indifference to word order and their inability to represent idiomatic phrases. For example, the meanings of "Canada" and "Air" cannot be easily combined to obtain "Air Canada". Motivated by this example, we present a simple method for finding phrases in text, and show that learning good vector representations for millions of phrases is possible.

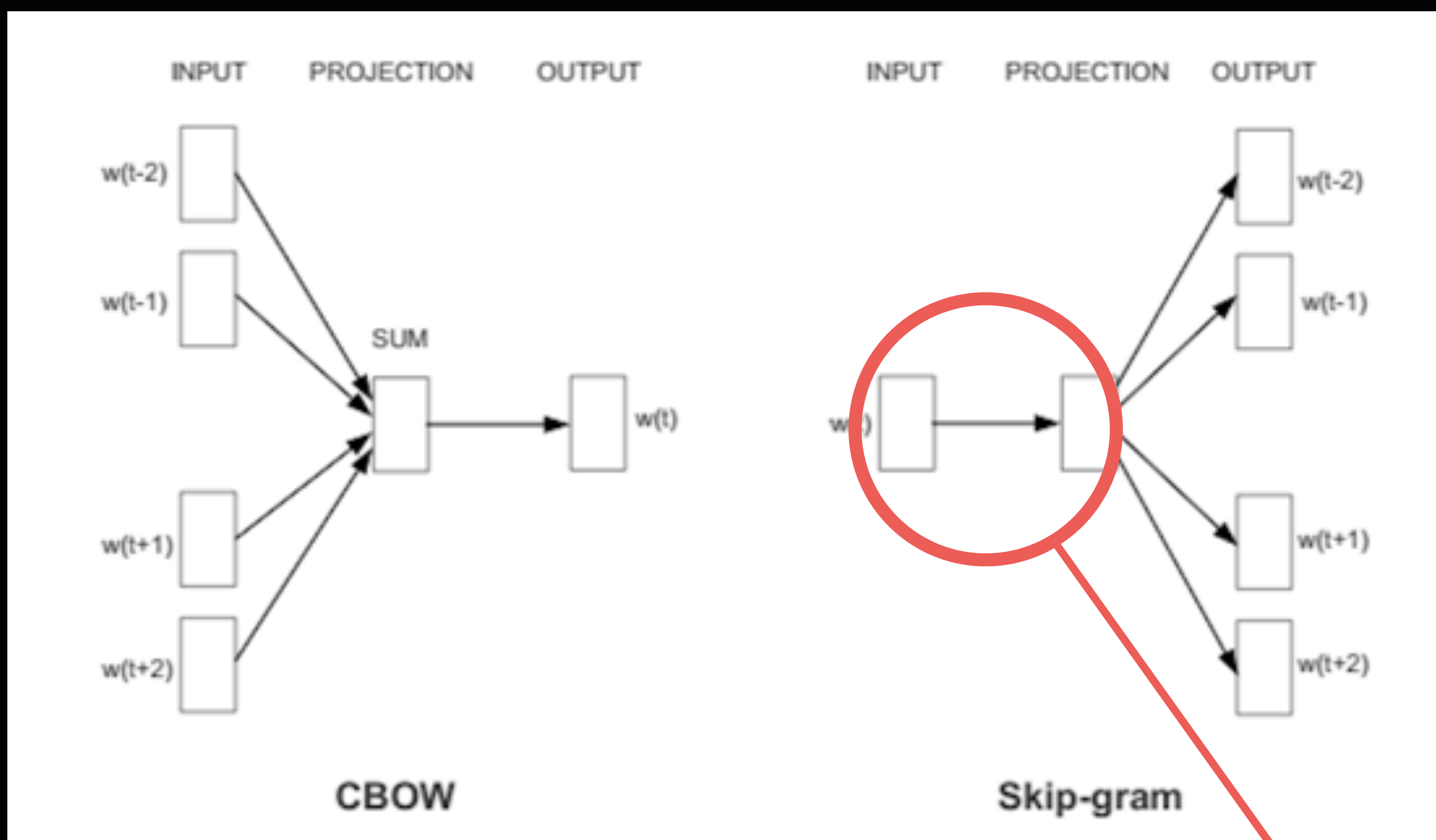
Embedding - Word2Vec (Cont’d)

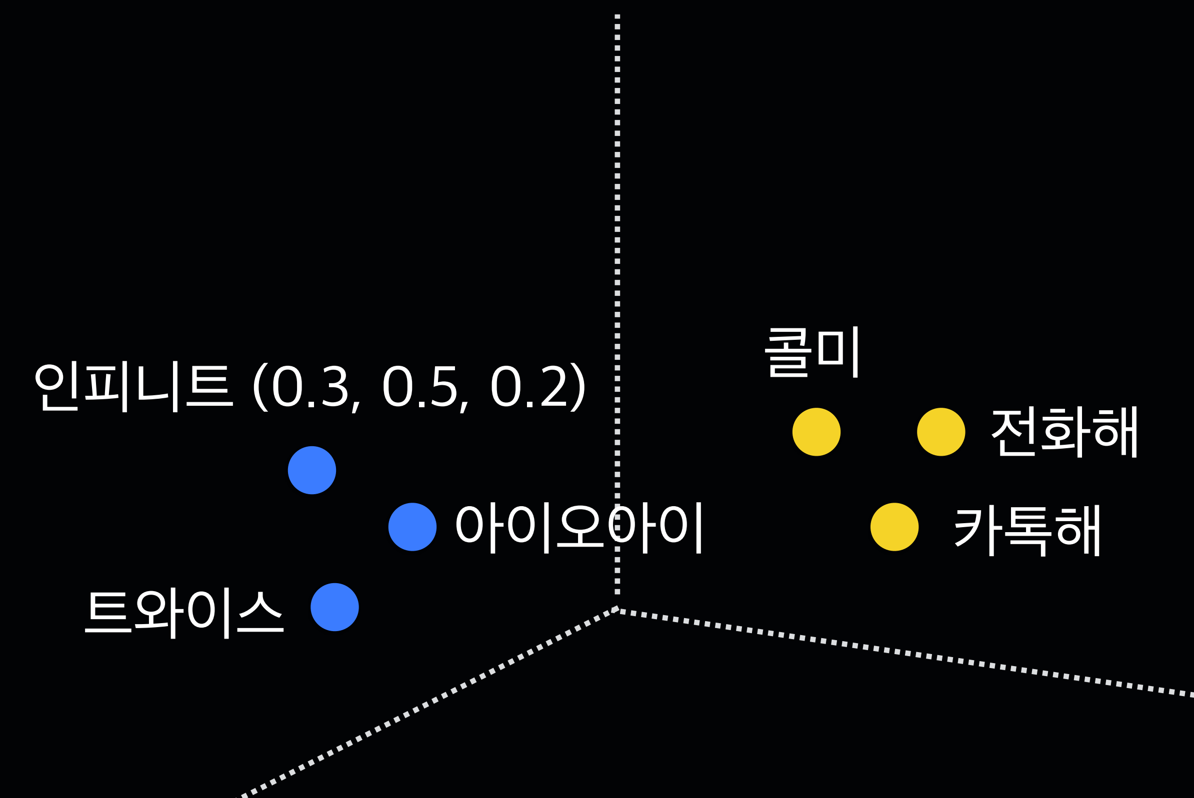
요새 신곡 발표한 다른 아이돌곡은 길거리서 안들리고 트와이스만 들림. 대세긴 대세네

헉 트와이스 신곡 넘좋자나 다들 너무 예뻐



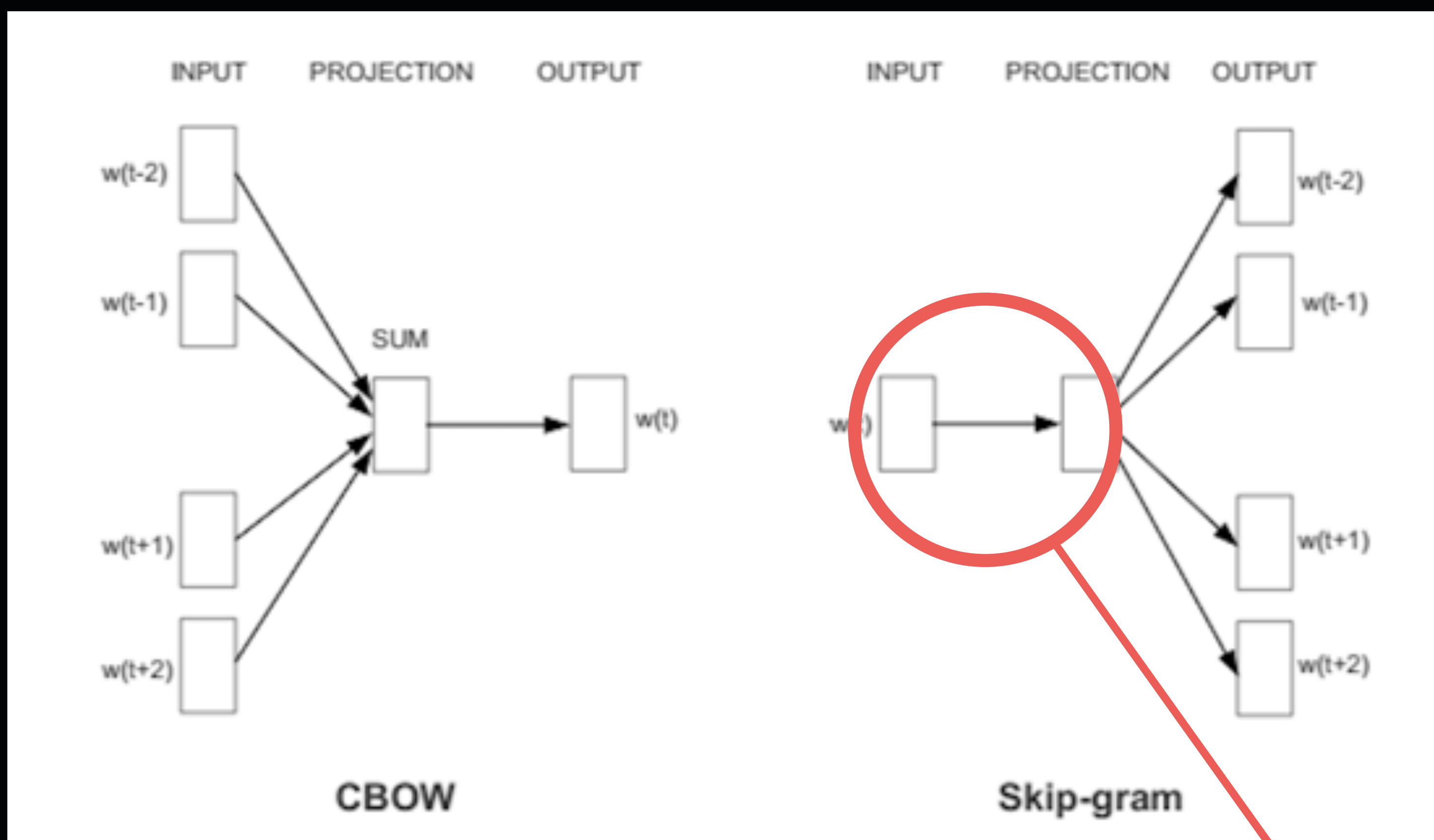
Embedding - Word2Vec (Cont'd)





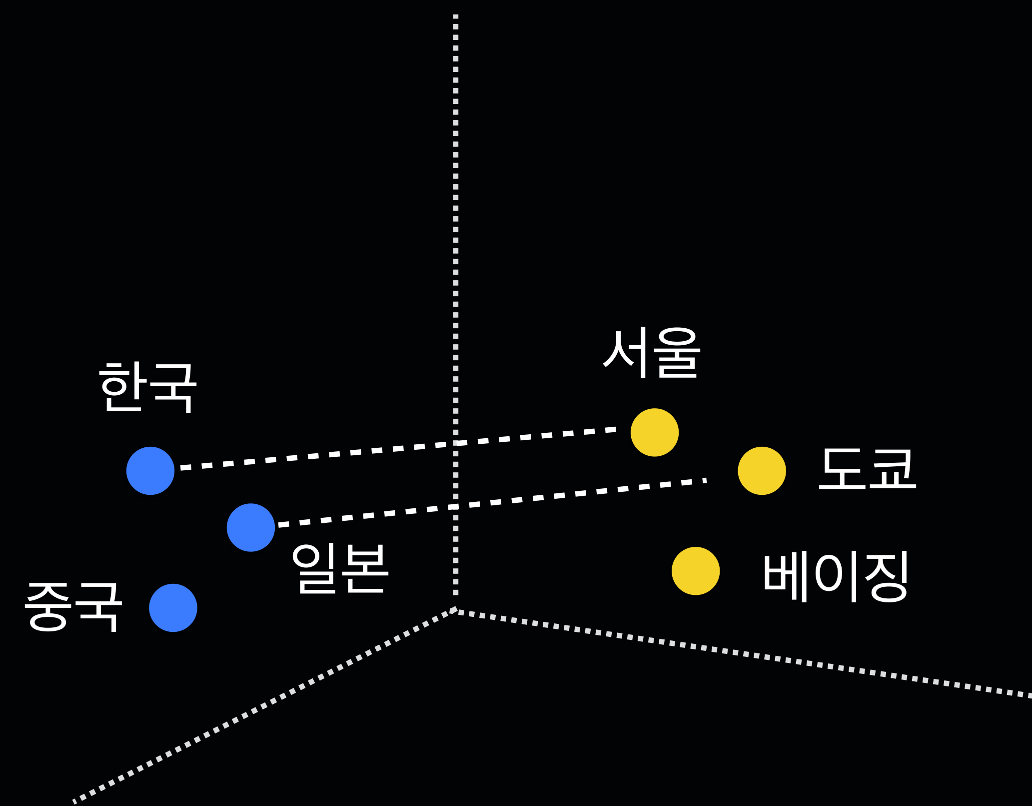
학습된 weight를 vector로 표현

Embedding - Word2Vec (Cont'd)



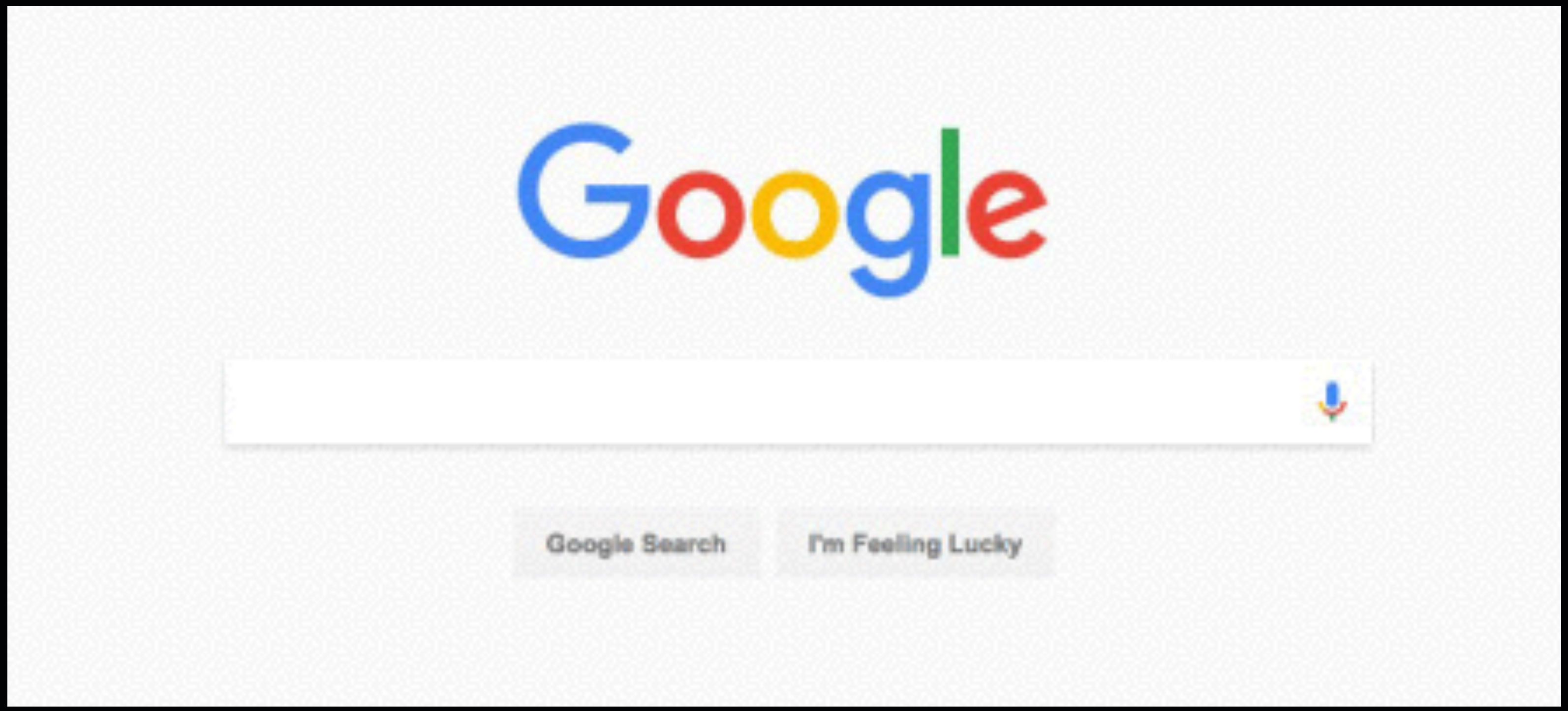
학습된 weight를 vector로 표현

Vector 연산을 통한 유추 가능



한국 - 서울이면 일본에게 뭐지?
-> 도쿄

Third Most Important Search Signal (of 100s)



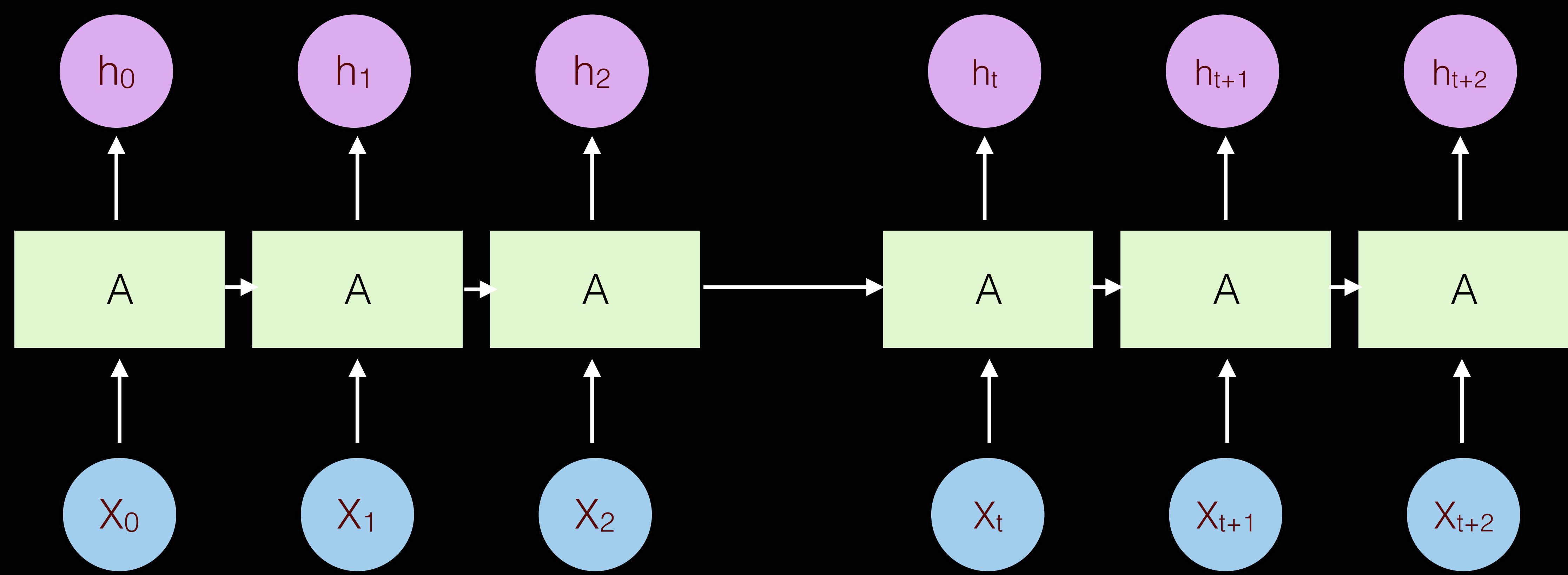
사람의 말을 이해한다는 것은?

2. 시간 순으로 이어지는 상대방의 말에서 기억할 것과 기억하지 않을 것을 구분하고 내가 할 행동, 말을 생성하는 것

"팀장님, 제가 다음주에 일이 있어서
금주 중에 지금 진행 중인 것
마무리하고 휴가 써도 될지요?"

Recurrent Neural Network(RNN):

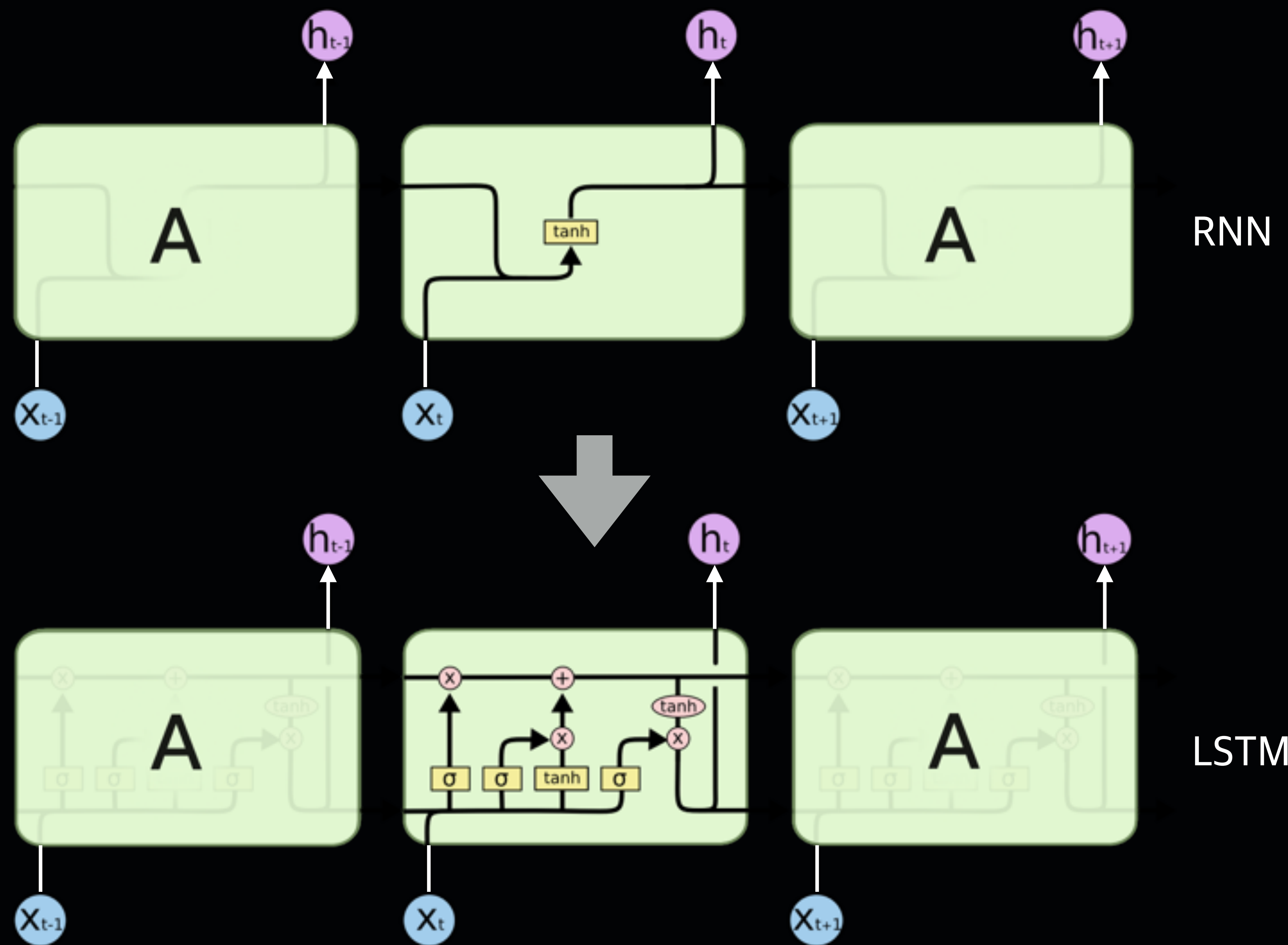
순서가 있는 (sequential) 데이터에서 과거의 정보를 모두 기억하여 다음 행동을 예측



실제로는 Long-term dependency를 잘 반영하지 못함

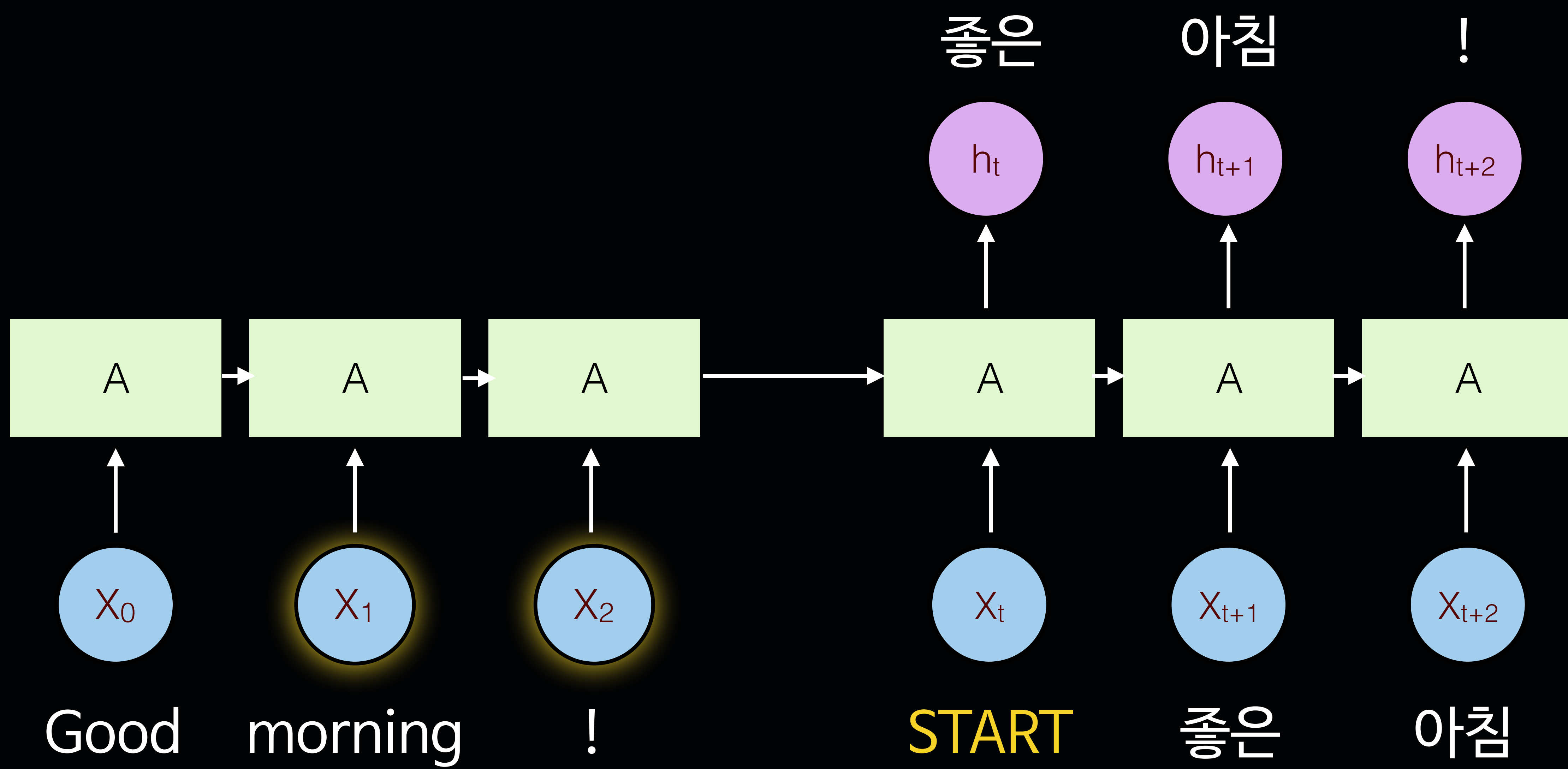
Long Short Term Memory (LSTM):

과거의 데이터로부터 흘려보낼 정보를 구분



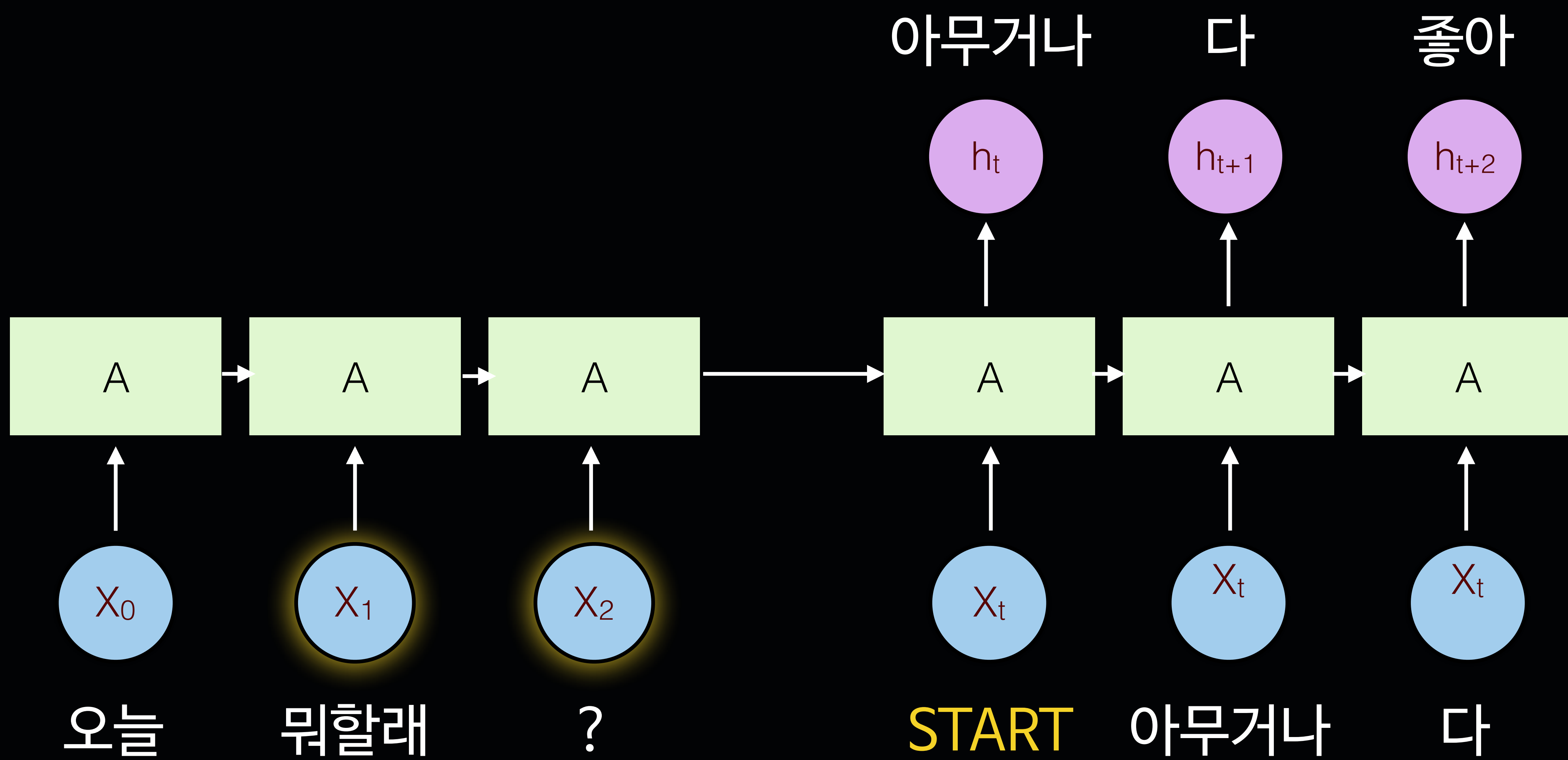
Recurrent Neural Network(RNN), LSTM 예시:

기계번역 (Machine Translation)



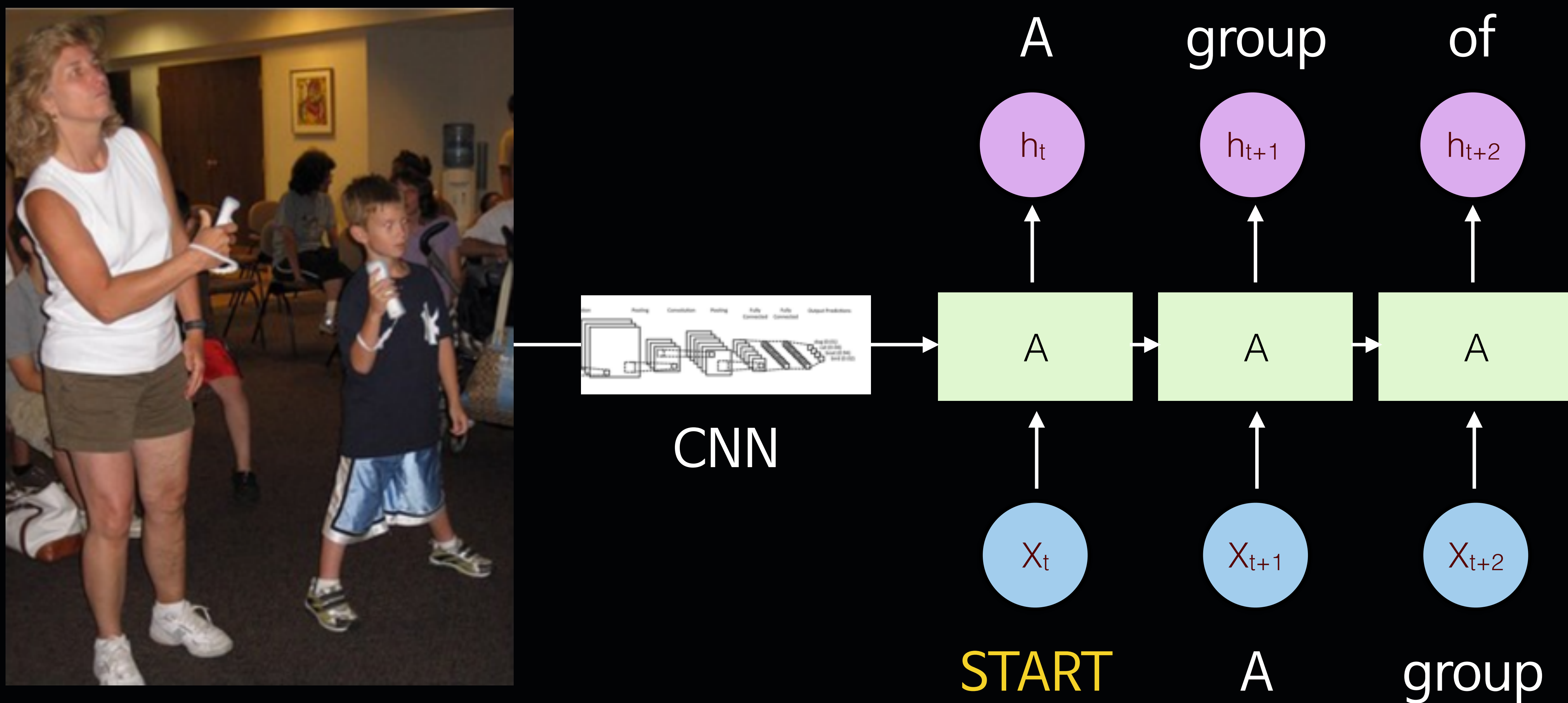
Recurrent Neural Network(RNN), LSTM 예시:

스마트 리플라이 (Smart Reply)



Recurrent Neural Network(RNN), LSTM 예시:

이미지 설명달기 (Image Captioning)



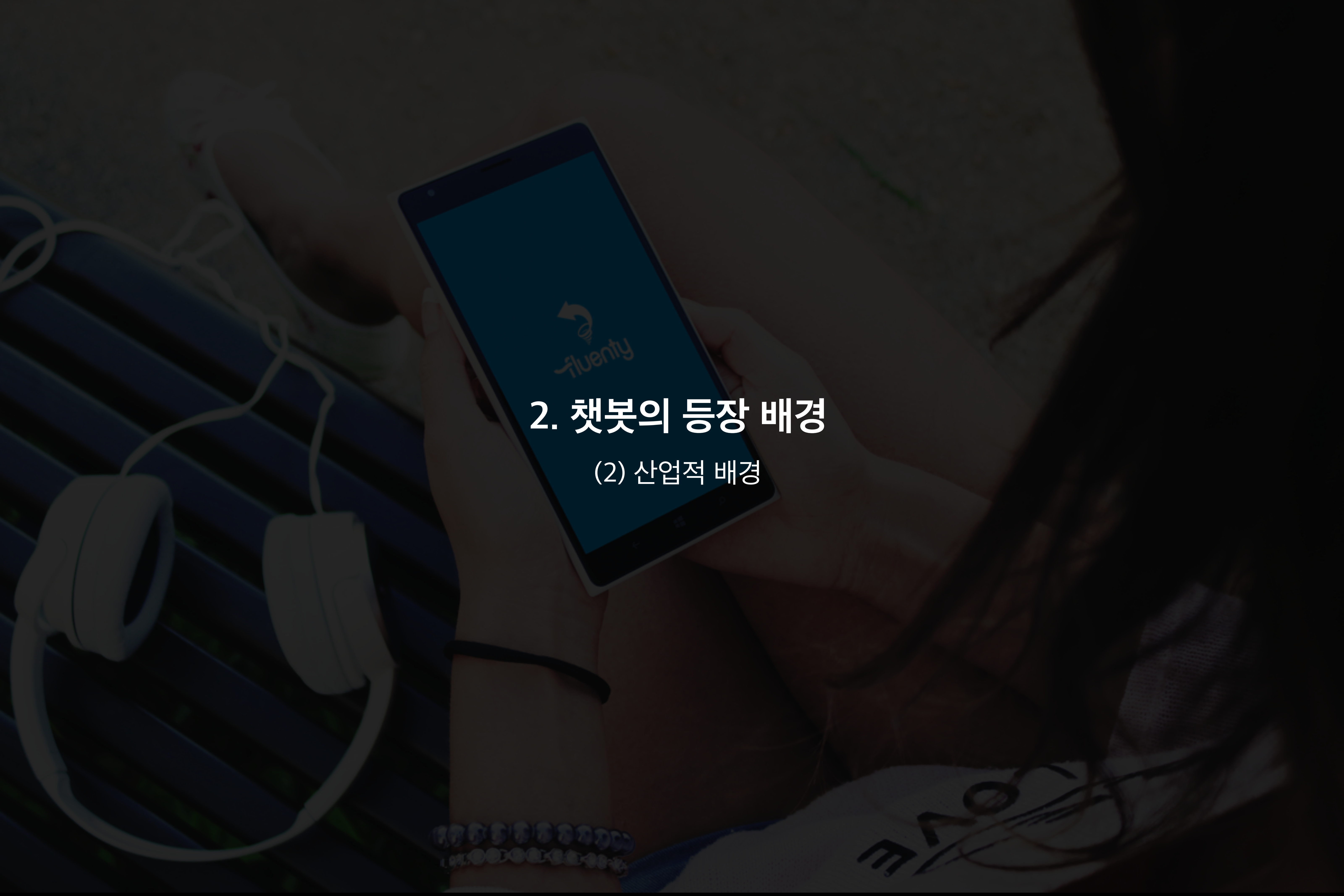
Recurrent Neural Network(RNN), LSTM 예시:

이미지 설명달기 (Image Captioning)

"A cow is standing in the middle of a street"

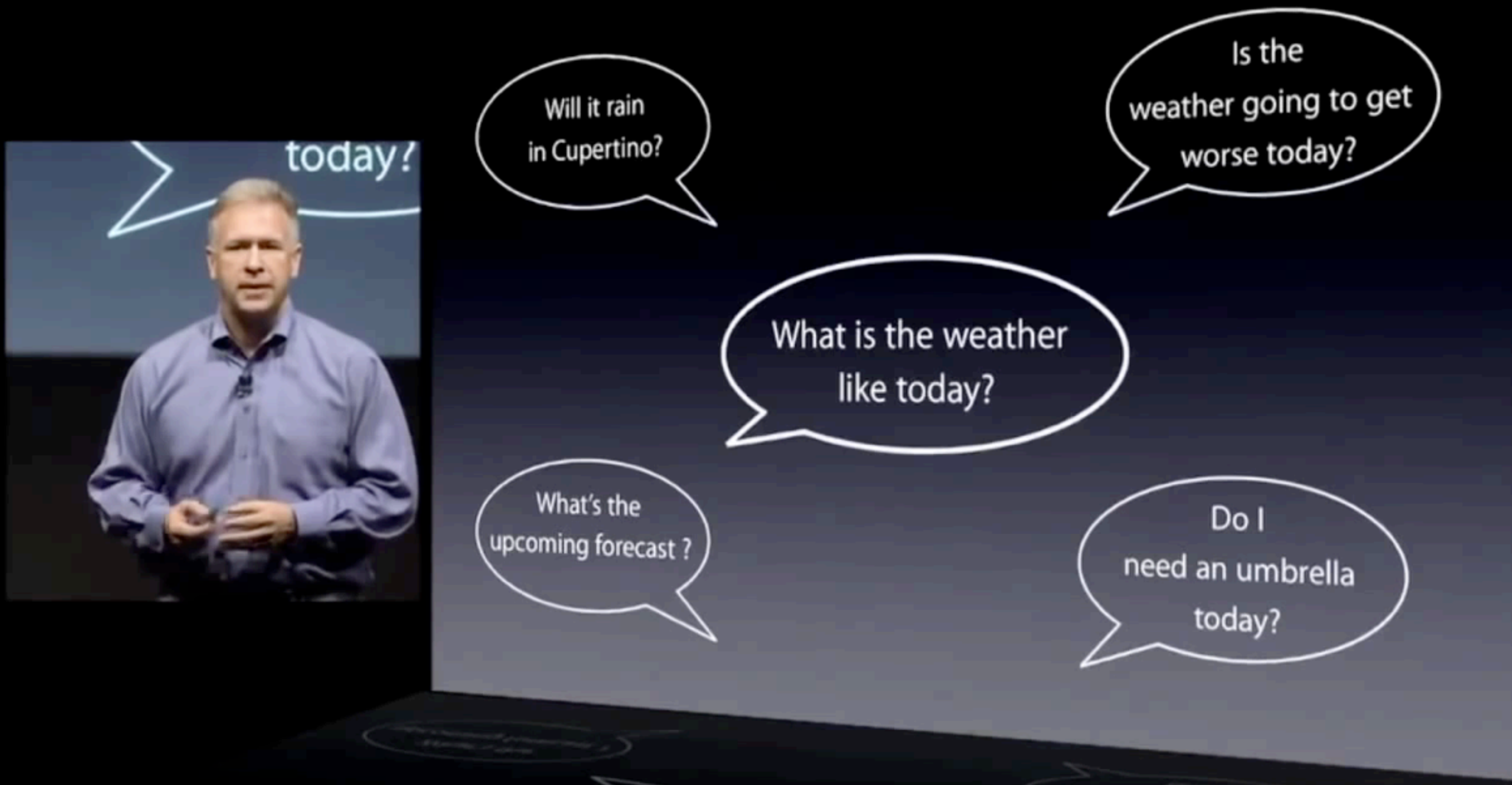
"A group of people sitting at a table with wine glasses"

"A cat is sitting on a toilet seat"



2. 챗봇의 등장 배경

(2) 산업적 배경

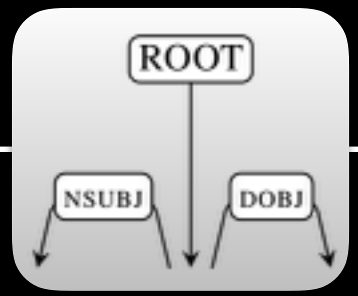

today?
Will it rain in Cupertino?
Is the weather going to get worse today?
What is the weather like today?
What's the upcoming forecast ?
Do I need an umbrella today?


J.A.R.V.I.S.

음성 기반 네이티브 챗봇의 등장 및 한계점

음성인식
Speech-to-Text

"Tell me a joke"



Natural Language Understanding
(NLU)

- Tagging
- Syntactic Parsing
- Named Entity Recognition

실행가능한 데이터
{intent:"greetings", …}

Apps

DBs

Chatting Bot

여기부터 시작

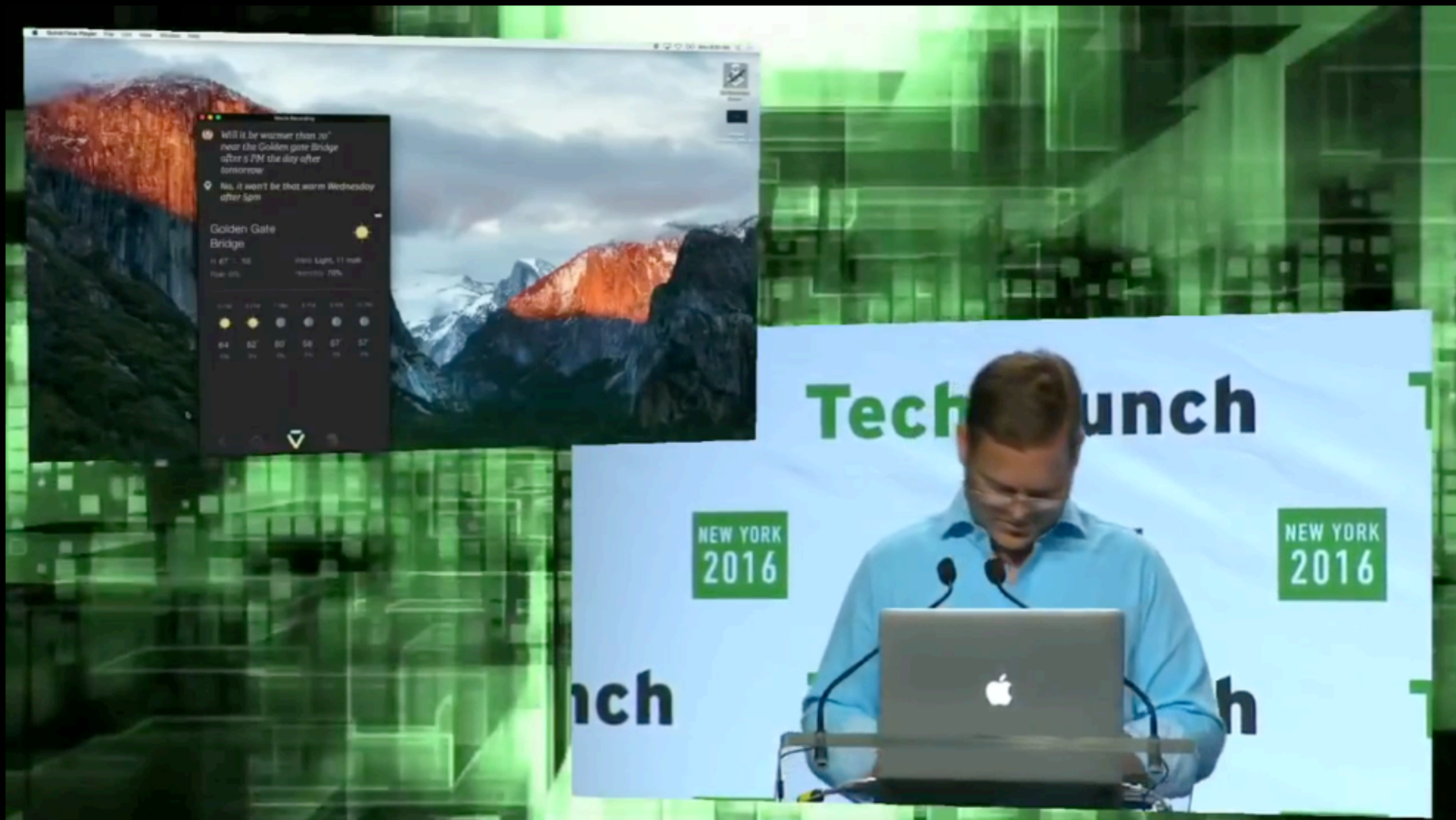
스마트폰에서의 음성입력은 "편리"하지만 "편안"하지 않음

자연어이해 후 가능한 액션이 부족

2. 스마트 홈, 텍스트 기반

1. 앱 생태계를 활용하는 플랫폼

스마트폰 음성비서 발전 양상 - VIV Labs (Acquired by Samsung)



음성인식 스마트 홈 디바이스의 성공 - Amazon Echo

amazon echo

8.6M

Units sold by 2016

26%

설문조사 결과 Alexa 기반 기기 사용자의 26%는 "아주 자주(very often)" 혹은 "꽤 자주(somewhat often)" 음성을 통해 쇼핑을 하는 것으로 답변함
(Source: BI Intelligence)

$7B

Amazon Echo를 통한 거래액은 2020년까지 약 7조원에 이를 것으로 예측됨
(Source: BI Intelligence)

CES 2017에서는 Alexa가 냉장고, TV, 스마트워치, 전구 등 다양한 제품에 결합됨으로써 인공지능이 플랫폼이 될 수 있음을 증명함

Amazon's Alexa is everywhere at CES 2017

by Jacob Kastrenakes | Jan 6, 2017, 3:20pm EST

음성인식 스마트 홈 디바이스의 성공 - Amazon Echo

10,000 Skills

"Alexa, ask Kayak where I can go for $500."

"Alexa, ask airport security for the wait time at LGA,"

"Alexa, start 7-minute workout,"

"Alexa, open Domino's and place my Easy Order,"

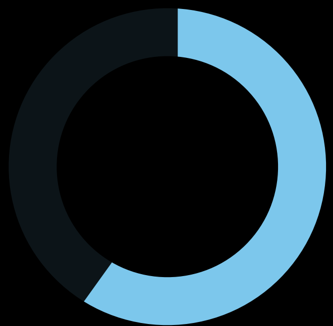
"Alexa, ask Uber to request a ride,"

"Alexa, ask Board Games who goes first in Scrabble."

"Alexa, ask Wine Gal to recommend a wine for a pepperoni pizza."

"Alexa, start my car"

대화형 인터페이스가 새로운 서비스 플랫폼으로 등장



65%

스마트폰 사용자

신규 어플리케이션을 설치하지 않음

$3.39

1개의 다운로드

평균 앱 설치 광고 단가는 2013년 $2.07에서 2015년 $3.39로 급격하게 증가하였음[1]

$2.1B

인공지능 스피커 시장 규모[2]

가트너에 따르면 인공지능 스피커 시장은 2020년에 약 21억달러에 이를 것으로 예측됨

현재 10,000개 이상의 스킬셋이 입점하여 있음

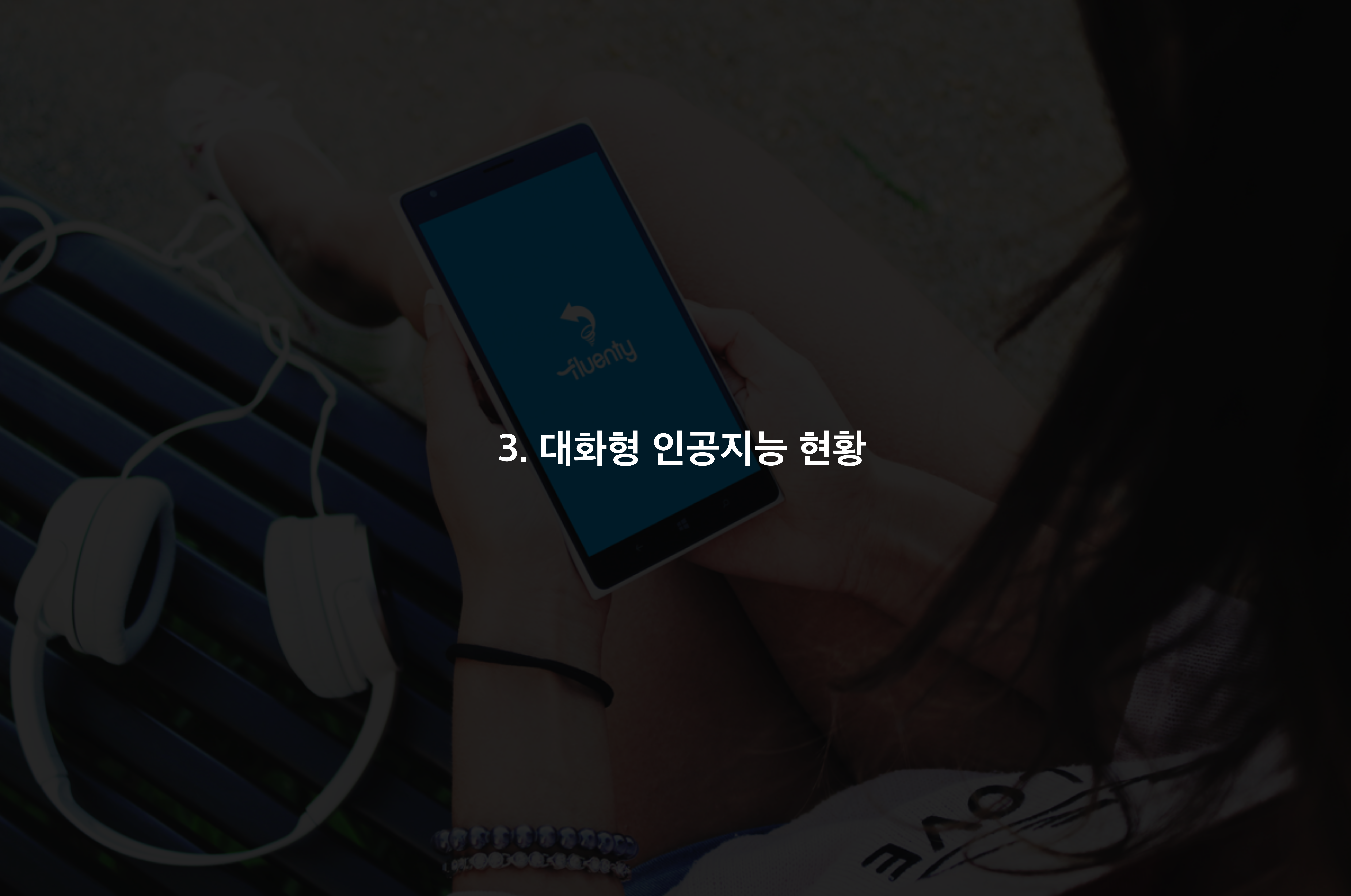
$32B

메신저 챗봇 시장 규모

메신저 챗봇은 기존 어플리케이션보다 높은 Retention을 보이며, 최대 320억 달러의 매출을 발생시킬 것으로 예측됨[4]

1. Source: QUARTZ 2. Source: Salesforce 3. Source: BI Intelligence 4. Source: BI Intelligence

3. 대화형 인공지능 현황



기술도 발전했고, 우리는 데이터가 많이 있으니 (주로 상담원 STT 데이터)
딥러닝이 **알아서** 대화형 인공지능을 만들어주나요?

Most of the value of deep learning today is in **<u>narrow domains</u>** where you can get a lot of data. Here's one example of something it cannot do: **<u>have a meaningful conversation</u>**.

There are demos, and if you cherry-pick the conversation, it looks like it's having a meaningful conversation, but if you actually try it yourself, it quickly goes off the rails.

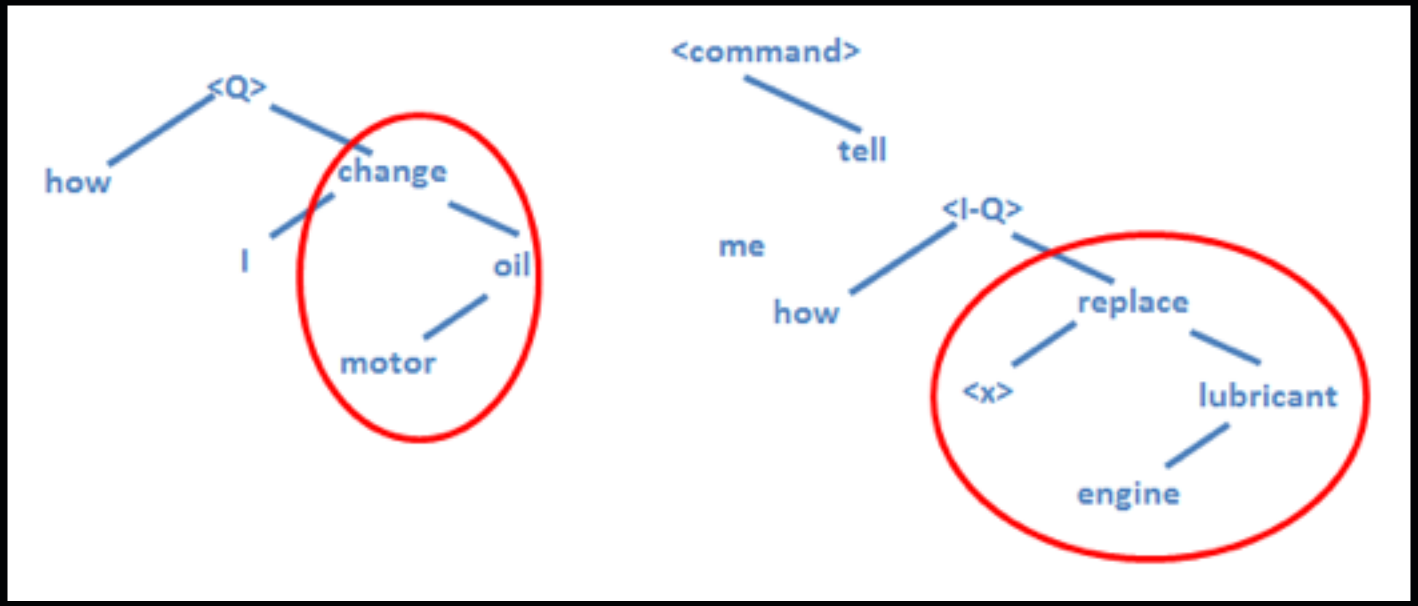
Question answering, small talk
vs.
Goal-based actions

Question answering, small talk (비목적성 대화)

Corpus-based

- Information retrieval (심심이, bigram 기반)
- Deep neural nets (RNN, LSTM 등)

Q1. How do I change the motor oil?
Q2. Tell me how the engine lubricant gets replaced.



Question answering, small talk의 한계

“비행기 예약해줘”

“네 예약해드리겠습니다.”

“…”

Goal-Based Action은 언어의 의도를 이해하고 이에 맞는 행동을 해야 함

“비행기 예약해줘”

출발지: “출발하실 도시명을 알려주세요”

도착지: “도착지 도시명을 알려주세요”

탑승인원: “탑승인원은 몇명이신가요?”

“김포”

“제주”

“두명이요”

Query

항공조회 DB

Goal-Based Action 에서의 자연어이해 (NLU)

언어이해분야 딥러닝을 통한 Intent Analysis, Entity Recognition, Slot 채우기

"김포에서 제주 가는 다음주 화요일 비행기 찾아줘"

Intent: 항공 검색
출발지: 김포
도착지: 제주
출발일: 2017-04-04
탑승인원: None

의도인식 후 빈 정보 파악,
Prompt를 통해 추가 정보 수집

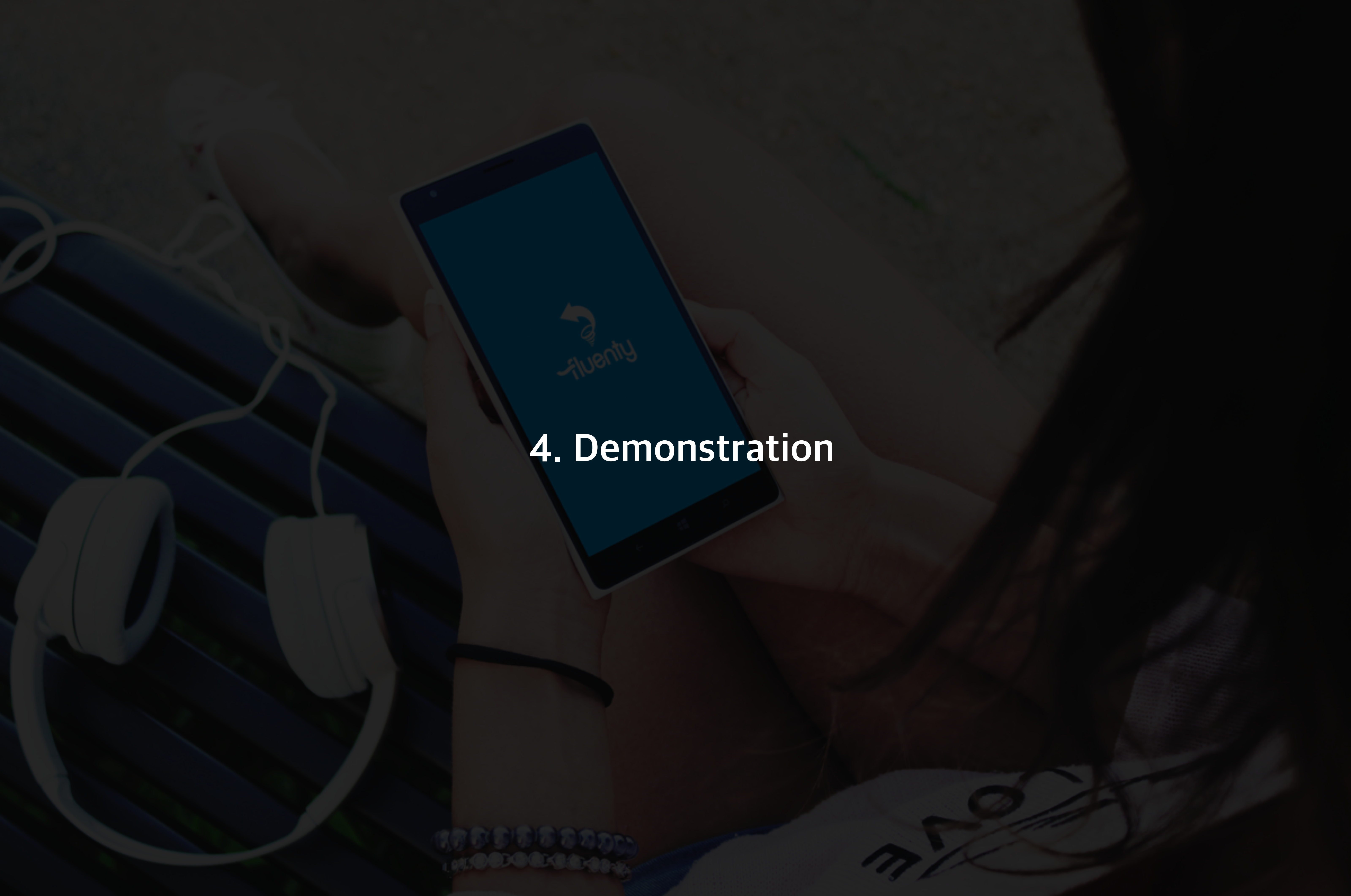
"몇명이 가시나요?"

기대했던 것과 다른 의도의 발화가
입력되어도 처리 가능

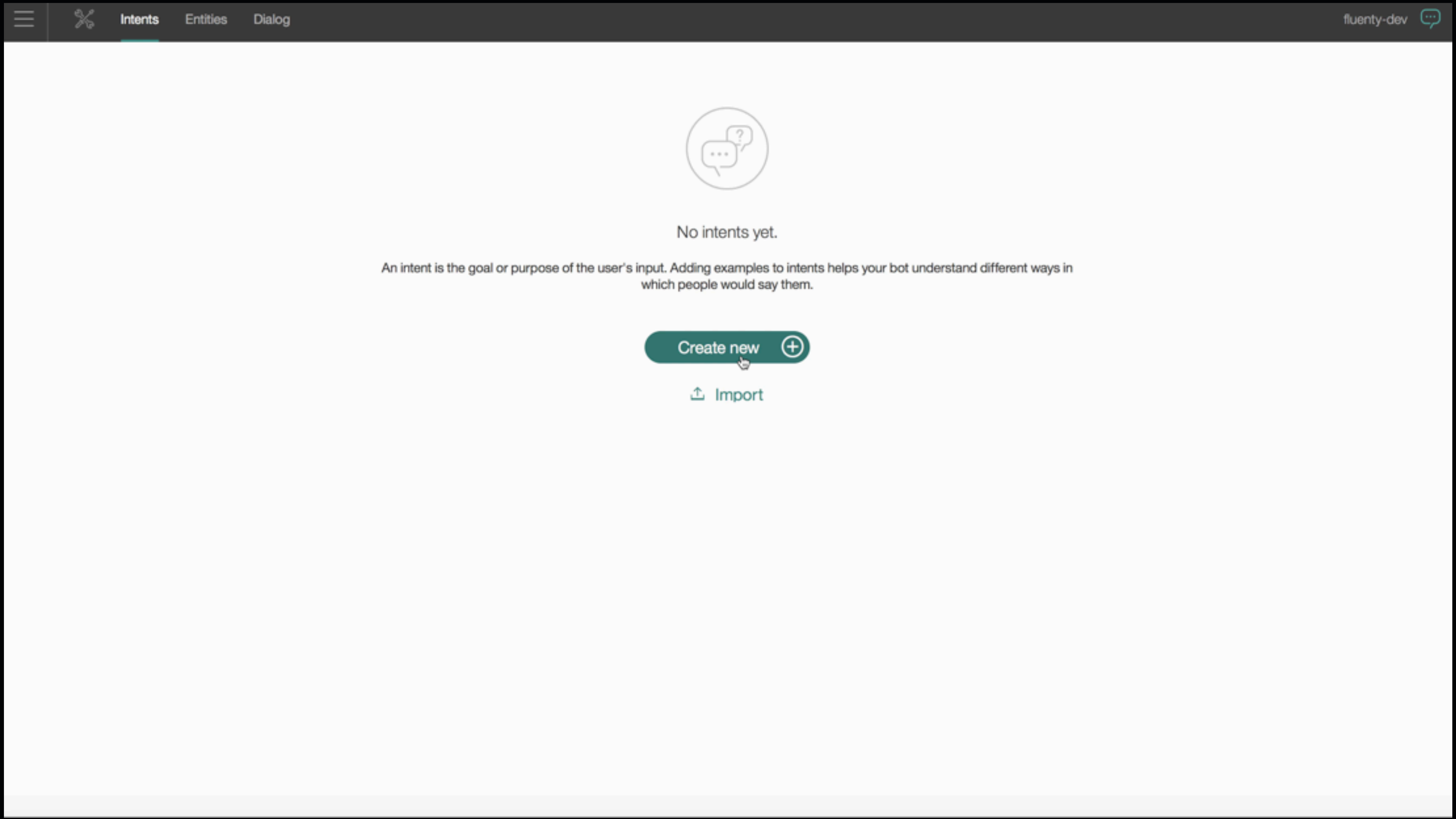
"제주에 3월에 가기 좋은 여행지 추천해줄래?"

Intent: 여행지추천
장소: 제주
시간: 3월

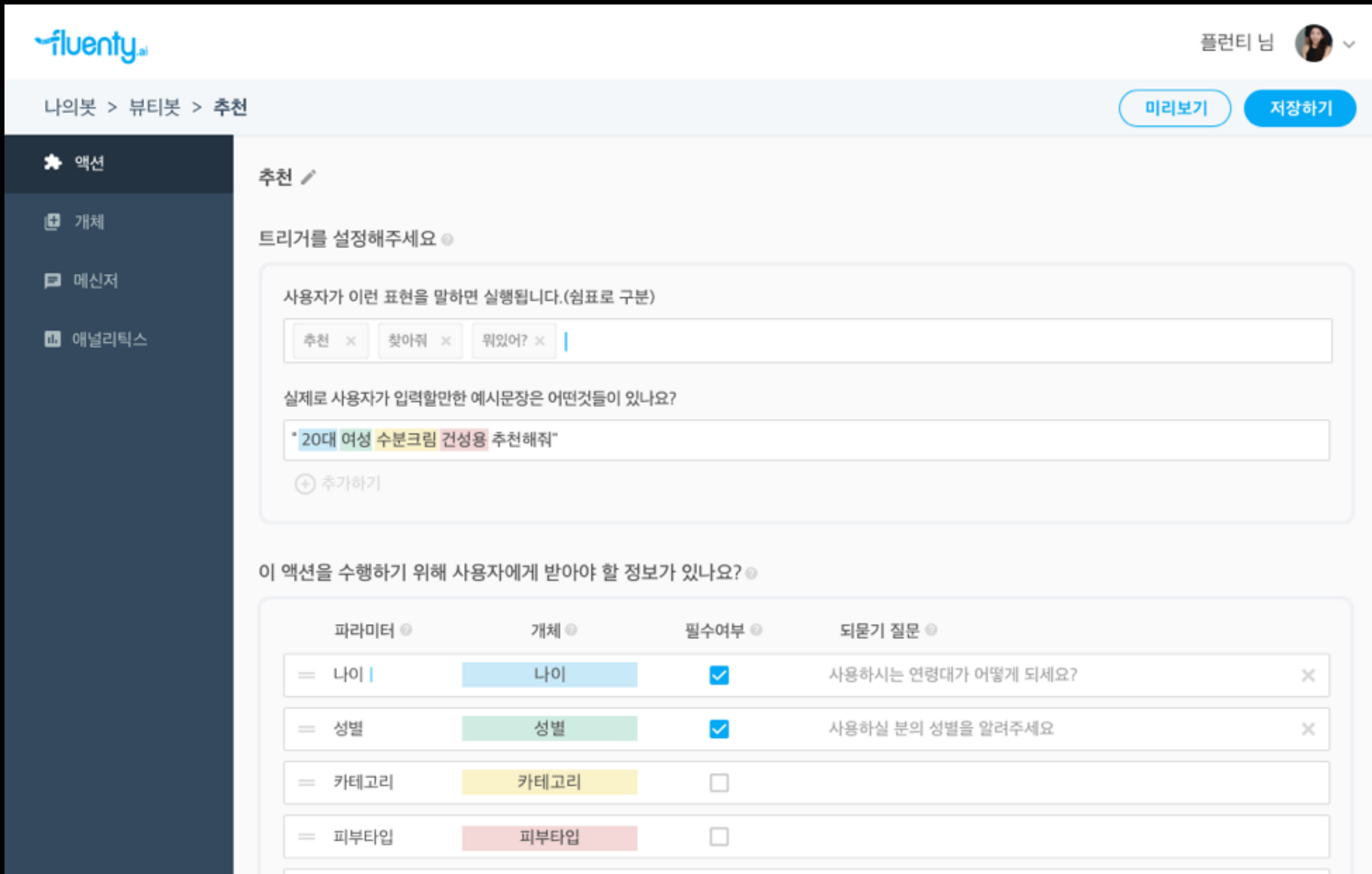
"3월에 제주 추천 여행지는 다음과 같습니다."


fluenty
4. Demonstration

봇 빌더 - IBM Watson Conversation



봇 빌더 - Fluenty.ai



Demonstration

감사합니다.

Presented by

Kanghak Kim

CEO, Co-founder

kanghak@fluenty.co