

# 카카오미니는 어떻게 점점 더 똑똑해지는가

올 1월 말에서 2월 초에 걸쳐 카카오미니가 더욱 다양한 사람의 언어를 이해하기 시작했다. 이제 카카오미니로 치킨, 피자 등을 주문할 수 있을 정도가 됐다. ‘윤식당’이 몇시에 하는지, 현재 예매 순위 1위가 어떤 영화인지도 카카오미니에게 물어보면 알 수 있다. 현재의 실시간 이슈가 무엇인지, 주말에 손흥민이 프리미어리그 경기 라인업에 포함되었는지도 “헤이, 카카오!”만 부르면 알 수 있다. 집에서 회사까지 가장 빠르게 갈 수 있는 대중 교통정보를 알려주는 것은 물론, 카카오T를 통해 택시를 불러주기도 한다.

## 카카오미니의 업데이트된 서비스 명령어

• 카카오톡 주문하기  
“헤이, 카카오! 후라이드 치킨 시켜줘”  
“헤이, 카카오! 오늘 할인 메뉴 뭐야?”

• 어학/사전  
“헤이, 카카오! 냉장고가 영어로 뭐야?”  
“헤이, 카카오! 점입가경이 무슨 뜻이야?”

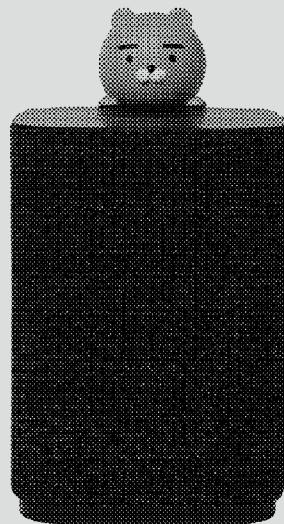
• 카카오 T 택시  
“헤이, 카카오! 강남역 가는 택시 불러줘”  
“헤이, 카카오! 택시 언제 와?”

• 프리미어리그 경기 일정/결과  
“헤이, 카카오! 맨유 경기 일정 알려줘”  
“헤이, 카카오! 어제 손흥민 잘했어?”

• TV/영화  
“헤이, 카카오! 윤식당 언제 해?”  
“헤이, 카카오! 요즘 재밌는 드라마 뭐있어?”  
“헤이, 카카오! 1987 관객수 얼마야?”

• 교통정보  
“헤이, 카카오! 9007번 버스 언제와?”  
“헤이, 카카오! 강남역까지 얼마나 걸려?”  
“헤이, 카카오! 이태원에서 명동까지 얼마나 걸려?”

• 실시간 이슈/정보  
“헤이, 카카오! 실시간 이슈 알려줘”  
“헤이, 카카오! 캐나다 수도 어디야?”  
“헤이, 카카오! 라면 칼로리 얼마야?”



## 카카오미니는 점점 똑똑해지는 중이다

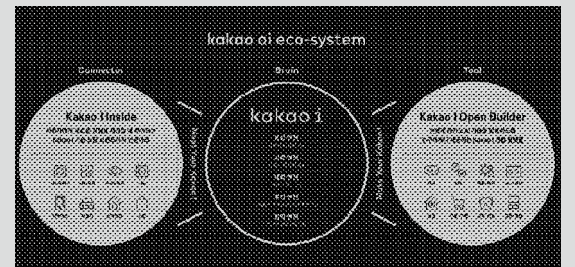
상상력 가득한 공상과학(SF) 영화를 보며 기대했던 AI의 모습은 아닐지라도 카카오미니는 점점 똑똑해지고 있는 중이다. 카카오미니를 사용하고 있는 사람이라면 아마도 한 번쯤은 카카오미니가 이전에는 알아듣지 못했던 명령을 어느 순간 이해하고 대답하는 현상을 경험했을 것이다. 그렇다면 카카오미니는 어떻게 진화하는 것일까?

<카카오 시리포트>를 통해 많이 소개된 딥러닝(deep learning) 기술은 카카오미니가 계속해서 학습하고 진화하도록 도와준다. 기술적인 내용은 이전의 <카카오 시리포트>를 통해 많이 소개돼 왔기에, 이번 글에서는 서비스 기획자의 시각으로 카카오미니가 사람의 언어를 이해하는 과정을 소개하고자 한다.

## 카카오미니는 어떻게 사람의 언어를 이해하는 것일까

카카오미니의 두뇌 역할을 하는 카카오(아이)의 코어 엔진(core engine)은 스피커 안에 있는 것이 아니다.

[ 그림 1 ] 카카오 AI 생태계<sup>1)</sup>



카카오미니에는 듣는 역할을 하는 ‘마이크’와 사용자에게 말을 하는 역할의 ‘스피커’가 탑재되어 있고, 사용자가 말한 명령은 인터넷망을 통해 카카오 코어 엔진 서버로 전송된다. 코어 엔진은 사용자가 인지할 수 있는 ‘업데이트’ 과정 없이 지속적으로 개선되기 때문에, 마치 사람처럼 학습한다고 느껴질 수 있다.

사용자가 “헤이, 카카오!”라고 부르면 카카오미니는 땡 소리와 함께 노란색 LED를 점등하며 대화 가능한 상태가 된다. 그 때부터 입력되는 음성은 서버로 전송되고 음성 엔진은 이를 텍스트로 변환한다. 대화 엔진은 변환된 텍스트를 사용자가 어떤 ‘의도’로 말을 했는지 ‘분류’하고, 각각의 의미 있는 단어를 ‘추출’한다. 서비스 에이전트는 추출된 단어를 이용하여 사용자 ‘의도’에 맞는 적절한 답변을 만들고, 이 답변이 다시 스피커로 전송되어 출력된다.

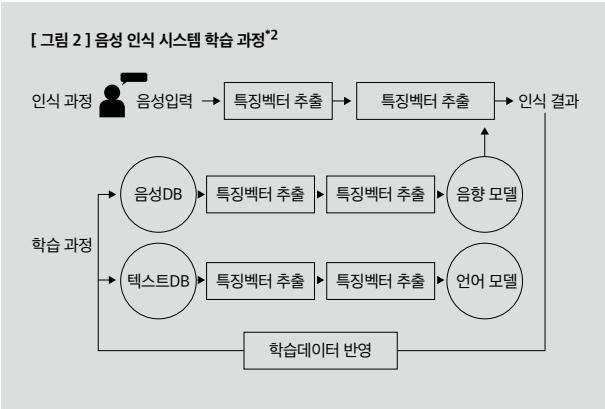
카카오미니가 언어를 잘 이해하려면 전송된 ‘음성’을 텍스트로 ‘변환’하는 과정과 텍스트를 의도에 맞게 ‘분류’하고 의미 있는 단어를 ‘추출’하는 것이 중요한데, 이때 ‘딥러닝’이라는 기술이 사용된다.

글 | 김다현 lea.kim@kakaocorp.com

쇼핑, 사전, 동영상과 같은 버티컬 서비스의 검색 기획을 해왔습니다. 2017년 초 AI 서비스팀에 합류한 후 카카오미니의 탄생을 함께 했습니다. 특히 카카오(아이)가 좀 더 똑똑해져서 많은 사람들에게 더욱 필요한 서비스가 되었으면 합니다.

### 카카오미니가 학습하는 방법

카카오미니를 많이 사용할수록 음성 인식 성능은 점점 더 좋아진다. 그 원리는 무엇일까? 아래 그림은 카카오 음성 엔진의 학습 방법이다. 음성 인식 시스템은 크게 ‘인식’ 과정과 ‘학습’ 과정으로 나눌 수 있다.



‘인식 과정’은 사용자의 음성을 프로그램이 인식하여 텍스트 결과를 내어주는 것으로, 실제로 사용자의 눈 앞에서 이루어지는 과정이다. 이때 음성 엔진은 미리 학습된 음향 모델과 언어 모델을 이용한다.

‘학습 과정’은 사용자의 음성과 인식된 텍스트를 다시 학습하여 음향 모델과 언어 모델을 만드는 단계다. 실제 사용자 데이터는 다양한 연령, 목소리 특성, 주변 환경을 포함한 데이터이므로 이를 학습하여 새로운 모델을 만드는 것은 성능 개선에 필수적인 요소이다. 일반적으로 음향 모델은 깊은 신경망(deep neural network, DNN) 방식을 적용하면 대용량의 사용자 데이터를 이용하여 더욱 정교한 모델링이 가능하다.

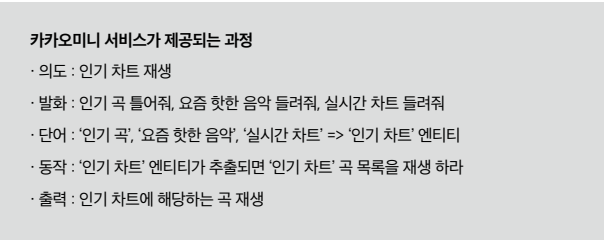
‘의도’ 분류를 학습하기 위해서는 ‘충돌’하는 명령어가 없는지 검토하는 데이터 전처리 과정이 중요하다. 유입된 명령어를 ‘의도’에 따라 분류한 후 ‘공통’으로 등장하는 것들은 ‘공통 명령어’로 묶어 사용자의 이전 상태에 따라 적절한 ‘의도’로 분류한다.

예를 들어 ‘그만’이라는 명령어가 유입되었을 때를 생각해 보자. 이 경우 사용자가 이전부터 ‘음악’을 듣고 있었다면 ‘그만’은 ‘음악 종료’에 해당하는 ‘의도’이고, ‘라디오’를 듣고 있었다면 ‘라디오 종료’에 해당하는 ‘의도’로 분류하는 방식이다. 정제된 데이터는 말뭉치에 등장하지 않은 단어 중 말뭉치에 존재하는 것과 유사한 단어를 학습하기 위해 단어 임베딩(word embedding) 방식을 적용하여 모델링 한다. 최근에는 GloVe(Global Vectors for Word Representation)를 활용하여 학습하고 있다. 특히 사용자들이 카카오미니와 더 많은 대화를 할수록 카카오미니는 점점 더 똑똑해진다.

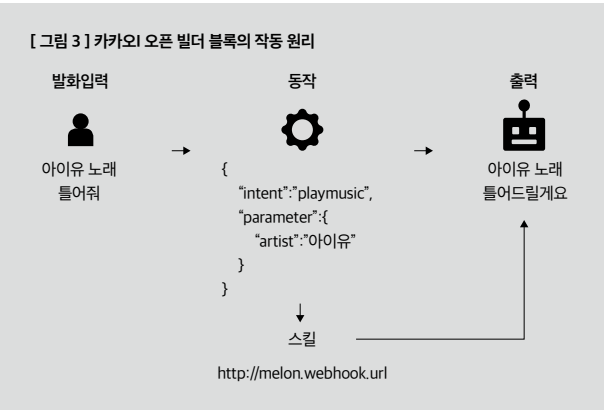
### 카카오미니 서비스가 추가되는 과정

카카오미니 서비스가 만들어지는 과정은 기존의 웹서비스와는 조금 다르다. 어떤 서비스를 만들지 결정하면 서비스를 사용하는 ‘의도’부터 분류한다. 예를 들어 ‘음악 서비스’를 만든다고 하자. 음악을 듣는 형태는 다양하다. 인기 차트를 듣고 싶을 수도 있고, 분위기에 맞는 음악이나 특정 가수의 곡을 듣고 싶은 의도도 있다. 음악을 듣다가 마음에 안들면 다음 곡 재생을 요청할 것이고 이것 저것 요청하는게 귀찮다면, 그냥 ‘음악 좀 틀어’라고 말할 수도 있다.

사용자의 ‘의도’를 분류한 이후에는 그 의도를 이루기 위해 사용자가 명령할 만한 ‘발화’를 만든다. ‘발화’는 의미 있는 ‘단어’의 조합으로 구성되고 특정 단어의 조합에 따라 서비스 에이전트가 어떤 동작을 수행할지 정의한다. ‘동작’ 수행은 카카오미니의 캐릭터에 맞게 출력 문장을 다듬는 과정을 거치고, 최종 결과물이 스피커를 통해 ‘출력’된다.



이 모든 과정은 카카오의 개발 플랫폼인 ‘카카오 오픈 빌더(Kakao I Open Builder)’<sup>\*3</sup>를 통해 이루어진다.

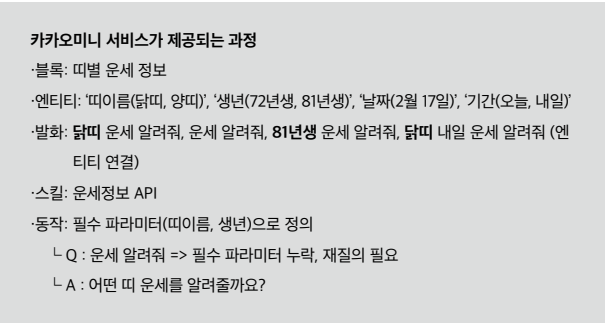


카카오 오픈 빌더는 ‘엔티티(entity)’, ‘블록(block)’, ‘스킬(skill)’로 구성되어 있다. ‘엔티티’는 사용자의 의도에 맞는 동작을 수행하기 위해, 주요 단어를 ‘추출’하도록 규정한 데이터 사전이다. ‘블록’은 사용자 명령으로 부터 의도를 파악하여 서비스 에이전트가 수행할 액션과 응답할 내용을 형식에 맞게 설계하고 정의하는 모듈(module)이다. ‘스킬’은 ‘블록’에서 사용자의 ‘의도’를 수행하기 위해 다양한 API(Application Programming Interface)를 호출하여 동작을 수행할 수 있도록 한다.

기획자는 위에서 정의한 ‘의도’를 빌더(builder)의 ‘블록’으로

생성하고 사용자 발화에 포함되는 ‘단어’를 ‘엔티티’ 사전으로 등록한다. ‘블록’에서 ‘발화’를 입력하면, 빌더는 입력된 발화의 단어를 사전에 등록된 ‘엔티티’와 연결한다.

‘블록’에서는 설정한 ‘스킬’을 연동할 수 있도록 ‘동작’ 파라미터(parameter)도 정의한다. ‘엔티티’가 제대로 추출되었는지 검증하고, 필수 ‘엔티티’를 정의해서 필수 값이 비어 있다면 사용자에게 ‘재질의’하도록 설정할 수 있다.



### 카카오미니가 매일 조금씩 좋아지는 이유

카카오는 자주 유입되는 발화와 의도를 분석할 수 있는 통계 시스템을 가지고 있다. 서비스 기획자는 이를 통해 매일 유입되는 데이터를 모니터링할 수 있다. 자주 들어오는 명령은 어떻게 더 스마트하게 답변할 수 있을지 검토하고, 실패한 명령은 왜 실패를 했는지 그 원인을 파악한다.

실패의 원인은 앞에서 설명한 ‘음성을 텍스트로 변환’ 하는 과정일 수도 있고, 텍스트를 ‘분류’하거나 ‘추출’하는 것이 잘못 되었을 수도 있다. 실패의 원인이 파악되면 이는 카카오 코어 엔진을 업데이트하거나 카카오 오픈 빌더에서 엔티티, 블록, 스킬을 수정해서 해결한다.

수정된 내용은 지속적으로 서버에 업데이트되기 때문에 카카오미니가 사용자의 의도를 한번에 이해하지 못했다고 해서 너무 실망하지 않아도 된다. 차츰 시간이 지나면서 내가 알지 못하는 사이에 끊임없이 업데이트될 것이기 때문이다.

### 앞으로 카카오는 어떻게 진화하게 될까

카카오는 카카오미니 외에 다양한 기기에 탑재 될 수 있다. 카카오미니가 아닌 다른 형태의 스피커나 냉장고, 에어컨과 같은 가전 제품이 될 수도 있다. 자동차나 집안의 스마트홈 시스템에도 카카오가 들어갈 수 있다. 그렇게 되면 운전하면서 “헤이, 카카오!”를 불러 길을 찾거나 음악을 들을 수 있고 음성 명령만으로도 집안의 불을 켜거나 끌 수도 있을 것이다.

또한 다양한 서비스를 추가할 계획도 가지고 있다. 카카오 통계 시스템 분석을 통해서 자주 유입되는 명령어 중 ‘서비스’가 없어서 대응하지 못하는 기능부터 업데이트 할 계획이다. 카카오 번역<sup>\*4</sup>, 보이스톡, 장소 정보 제공 등의 서비스를 준비 중이고, 외부 제휴를 통해 다양한 콘텐츠 서비스 또한 제공할 예정이다.

카카오 시스템은 기존 카카오미니 사용자들에게도 동일한 업데이트를 지원할 예정이다. 향후 카카오가 탑재되는 자동차, 스마트홈, 스피커 등은 비록 기기의 형태와 기능은 다르지만 사용자들에게 모두 동일한 경험을 제공 할 것이다.

<sup>\*1</sup> 참고 | <https://kakao.ai/about> <sup>\*2</sup> 참고 | 김훈, “음성 인식 방법과 카카오의 음성형 엔진,” 카카오시리포트 Vol.6, 2017 <sup>\*3</sup> 참고 | <https://i.kakao.com/openbuilder/#/login> <sup>\*4</sup> 참고 | <https://translate.kakao.com>