



# AIQ.

# TALK

# STORY

vol.02

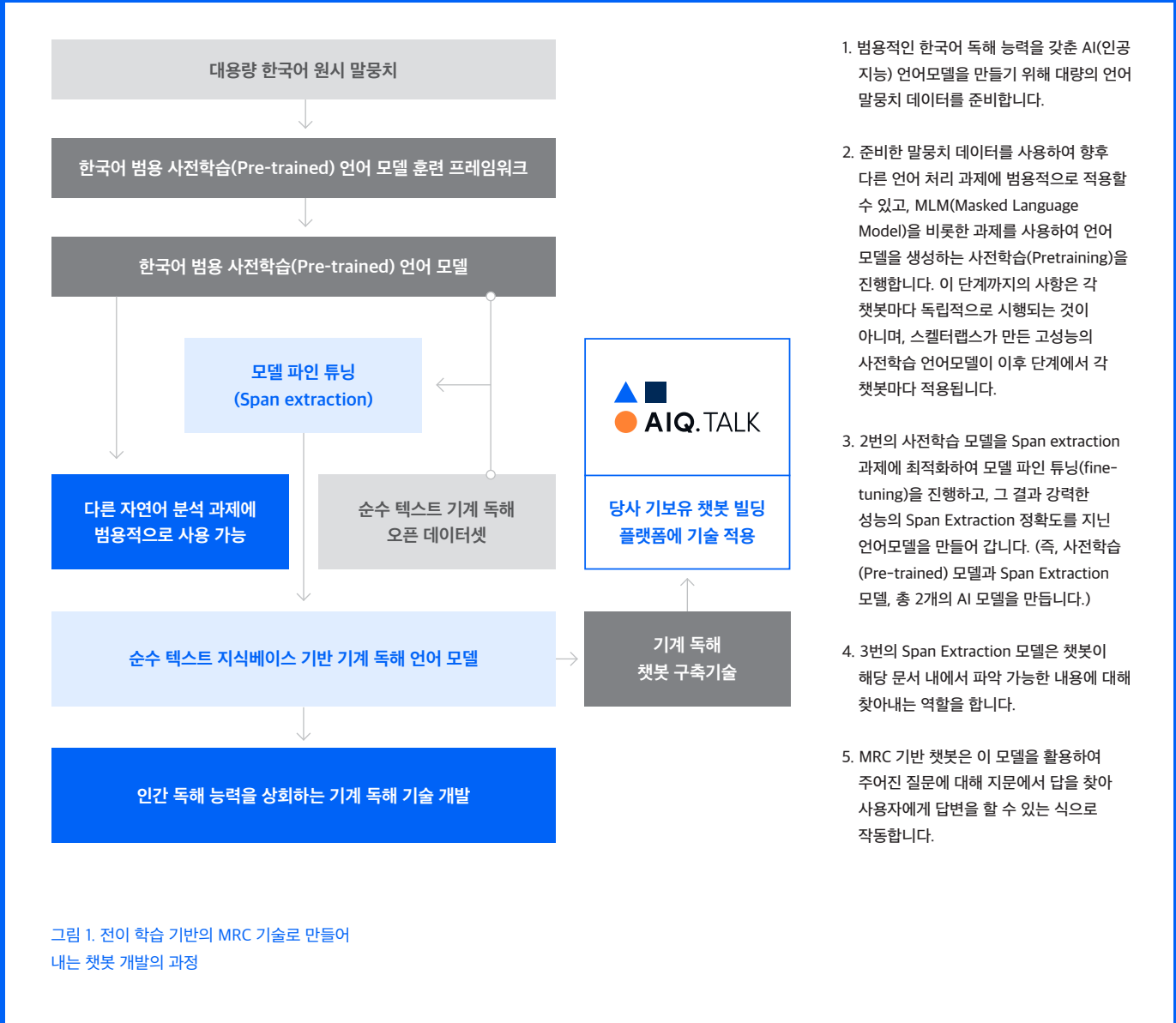
MRC(기계독해)

# MRC는 AI가 빠르고 정확하게 대용량의 문서를 이해해 사용자의 의사결정에 유용한 정보를 추출하는 혁신적인 대화형 인공지능 기술입니다.

Skelter Labs CTO, 조성진

스켈터랩스는 AIQ.TALK Story 시리즈를 통해 일상을 편리하게 만드는 대화형 인공지능 기술에 대해 더 쉽게 소개하고 있습니다. Vol.01에서는 대화형 인공지능 솔루션을 개발할 때 근간이 되는 세부 기술 중 하나인 인텐트 분류(Intent Classification)에 대해 이야기했습니다. 이를 통해 인텐트 분류가 높은 정확도로 잘 이루어질 때 챗봇의 기술적, 제품적 가치가 높아진다는 점을 말씀드릴 수 있었습니다. 이번 Vol.02에서는 최근 각광을 받고 있는 MRC(Machine Reading Comprehension, 기계독해)에 대해 살펴보겠습니다.

Vol.01에서 집중적으로 소개한 인텐트 분류 기술이 이 세상 모든 사용자 발화를 하나하나 학습시키지 않고 그 중 일부 예문만 학습시켜도 질문의 의도를 이해하여 대답시키는 쪽이라고 했을 때, MRC 기술로 만들어내는 챗봇 개발의 과정 전반을 요약하면 다음과 같습니다.



인간이 방대한 문서의 모든 내용을 완벽하게 기억하는 것은 거의 불가능한 일일 것입니다. 하지만 인공지능 기반의 MRC 기술은 숙지해야 하는 모든 문서에 대한 학습을 빠른 시간 내에 완료해, 문서 내에서 사용자가 궁금한 내용이 적힌 위치를 빠르고 정확하게 찾을 수 있습니다. 또한 지속적인 모델의 진화에 따라 여러 개의 문서 속에서 사용자가 찾는 내용이 어느 문서에 속해 있는지, 그리고 해당 문서 속에서 또 어떤 위치에 있는지까지 찾을 수 있게 됩니다. 그리고 이를 바탕으로 사용자가 궁금한 내용에 대한 정답을 제공하는 것입니다.

# What is MRC?

## 인간 독해 능력을 뛰어넘는 독해 기술

스켈터랩스는 뛰어난 범용성을 가진 한국어 사전학습 언어모델(Korean Pre-trained Model)을 기반으로 질의응답 (Question Answering) 목적을 위한 MRC 모델을 개발해 왔습니다. 그리고 이를 우리의 챗봇 플랫폼 AIQ.TALK Chatbot에 적용해 각각 필요한 use-case에 맞게 인텐트 기반 챗봇과 MRC 기반 챗봇을 구현할 수 있는 플랫폼으로 진화시키는 것을 목표로 하고 있습니다.

앞으로의 보고서 내용은 MRC에 대해, 한국어 기준 최고의 성능을 가질 수 있게 만든 과정과 의미는 무엇인지, 앞으로 어떤 식으로 활용이 될 수 있을지에 대해 담고 있습니다.

# MRC란 무엇인가?

MRC는 독해력(문해력)을 가진 사람이 풀 수 있는 과제를 인공지능이 해결할 수 있도록 하는 기술 전반을 통칭하는 말입니다. 여기에는 Cloze Test, Multiple Choice, Open Domain 등 다양한 세부 분야가 있지만, 이 글에서는 특히 다양한 MRC 기술 중에서도 Question Answering 분야의 Span Extraction 기술에 대해 이야기하려고 합니다. Span Extraction 기술이란 기계(AI)가 주어진 지문(Context) 및 지문과 관련된 질문(Question)을 이해하여, 지문과 관련된 질문에 적절한 답변을 지문 속에서 도출하는 기술을 뜻합니다.

예를 들어 온라인 쇼핑몰에서 전자기기를 구입한 소비자가 교환 신청 기한이 궁금할 때, 쇼핑몰이나 제조사에서 제공한 길고 복잡한 문서를 전부 다 읽는 대신 MRC 기반 챗봇에게 “제품 교환 신청은 언제까지 하면 되나요?” 라고 질문을 던지면, 문서 내용을 학습한 인공지능이 문서 내에서 관련된 정보가 들어있는 구간(“.. 교환 및 환불은 주문일 이후 7일 이내에 신청되어야..”)을 찾아 “7일 이내”라고 바로 답변해줄 수 있습니다.

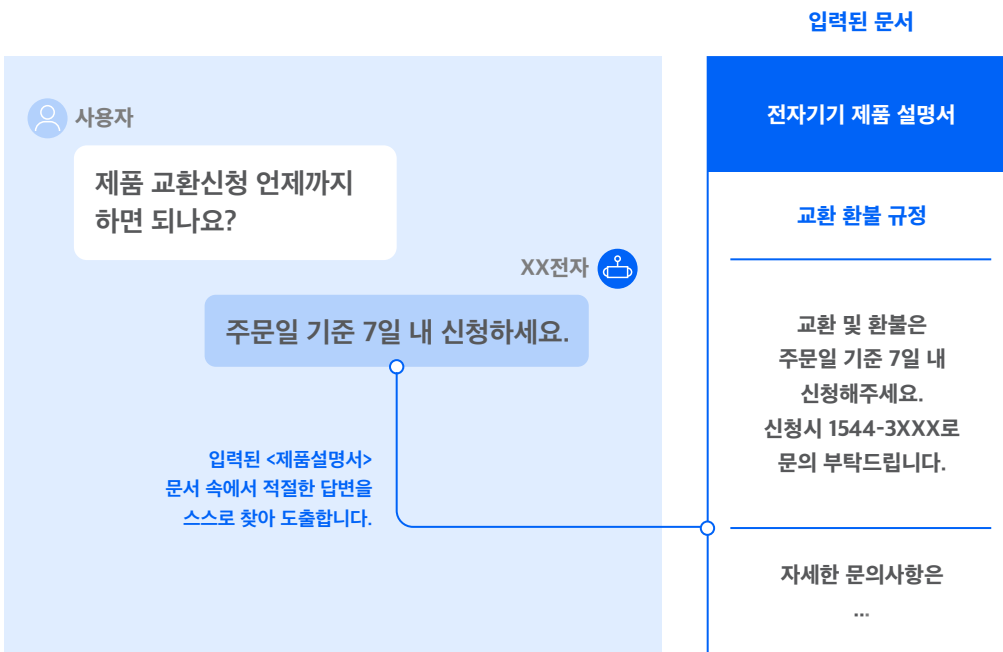


그림 2. MRC 기반 챗봇 이용 예시 1

은행에서 새로운 펀드 상품을 가입하려고 할 때, 해당 상품이 주로 어떤 기업 주식으로 포트폴리오가 구성될 예정인지 궁금하면 상품설명서를 직접 읽는 대신 “주식 운용이 어떤 포트폴리오로 구성되나요?” 라고 챗봇에게 질문을 던지고, 챗봇은 “XX전자, YY화재, ZZ카드 등 경쟁력이 높은 AA그룹에 소속되어 있는 주식 위주로 구성됩니다”와 같은 답변을 상품설명서에서 찾아 할 수도 있을 것입니다. 학계나 업계에서 새롭게 제시되는 머신러닝 및 딥러닝이 어떻게 활용되느냐에 따라 기계독해력의 수준은 향상될 수 있고, 당연하게도 사람의 질문에 대한 답변 정확도도 점차 높아질 것입니다.

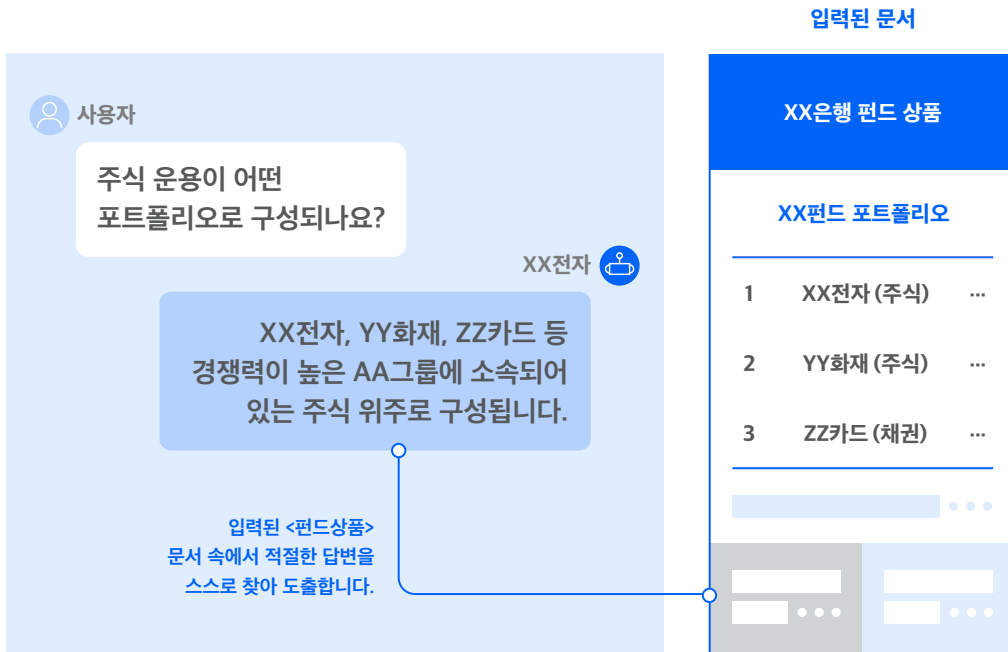


그림 3.  
MRC 기반 챗봇 이용 예시 2

MRC기반 챗봇을 적용할 수 있는 산업 분야는 무궁무진합니다. 특히 앞서 제시한 예시에서 처럼 금융, 의료, 법률, 특허 등 방대한 정보를 담고 있는 장문의 문서를 반드시 활용해야 하는 분야에서 마치 사람에게 물어보듯 자연스럽게 질문하는 것만으로 특정한 정보를 찾아야하는 경우가 대표적입니다.

# MRC 개발을 위한 데이터셋

MRC 기술을 지속적으로 발전시켜가기 위한 훈련을 위해서, 그리고 그렇게 만들어진 기술의 성능을 검증하고 개선점을 찾기 위해서 가장 중요한 것은 충분한, 의미있는 양의 지문, 즉 데이터셋입니다. 사실 영어 기반의 MRC를 개발하고 성능을 검증할만한 오픈 데이터셋은 이미 쉽게 찾아볼 수 있고 종류도 다양합니다. 대표적인 것이 미국 스탠포드 대학에서 시작된 영어 AI 언어지능 연구용 질의응답 데이터셋인 SQuAD(The Stanford Question Answering Dataset)입니다. 현재 많은 개발자들이 활용하고 있는 SQuAD 2.0 버전은 기존 SQuAD 1.1에서 보유한 100,000개의 위키피디아(Wikipedia) 기반 질문에 50,000개 새로운 질문들을 더해 총 150,000개의 데이터로 구성되어 있습니다. 2.0에서 추가된 50,000개의 질문들은 문서에서 절대 찾을 수 없는 내용들에 대한 질문들로, 엉뚱한 오답을 내지 않고 '알 수 없다(unanswerable)'고 대답할 수 있는지를 검증할 수 있는 용도로 만들어졌습니다. 그 외에도 구글이 질의응답용으로 위키피디아 기반의 30 여만개 문장 데이터로 구성하여 제공하는 NQ(Natural Questions), 마이크로소프트 AI & Research에서 검색엔진 Bing의 검색 데이터를 바탕으로 만든 MS MARCO(Human Generated MACHine Reading Comprehension Dataset) 등이 널리 쓰이고 있습니다.

하지만 이처럼 기존의 MRC 오픈 데이터셋은 영어 기반이 대부분이었기 때문에 한국어 MRC 기술을 개발하고 성능을 검증할 만한 데이터셋이 부족했습니다. 이러한 한계점을 극복하기 위해 국내에서도 MRC 개발에 활용될 수 있는 한국어 데이터셋 구축의 필요성이 대두되었고, LG CNS가 한국어 위키피디아로부터 추출한 표준데이터 7만개로 구성된 KorQuAD(The Korean Question Answering Dataset) 1.0을 2018년 12월 무료 개방하며 본격적인 한국어 MRC 기술개발의 물꼬가 트였습니다. 그 후 표준데이터를 10만개로 확대, 문장의 나열로 구성된 문단은 물론 표나 개조식 계층 구조를 지닌 문서, 혹은 문단 전체를 답변 내용으로 지정할 수 있는지, 그리고 얼마나 빠르게 답변할 수 있는지의 여부 등을 복합적으로 평가할 수 있도록 업그레이드한 KorQuAD 2.0이 2019년 9월 공개됐습니다.

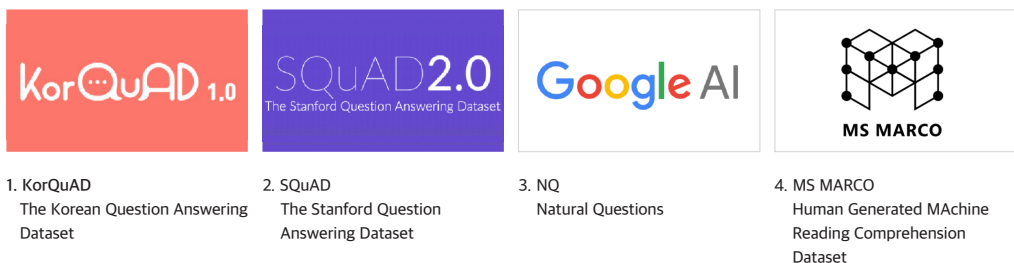


그림 4. 대표적인 MRC 개발 및 성능 검증용 오픈 데이터셋

# KorQuAD 데이터셋

현재 스킨터랩스가 리더보드 1위를 기록하고 있는 KorQuAD에 대해 더 자세히 알아보겠습니다. KorQuAD를 운영하고 있는 LG CNS는 각 기업 혹은 연구기관이 KorQuAD 공개 훈련 데이터셋을 이용하여 만든 AI 모델을 제출하면, 비공개 데이터셋을 이용하여 성능을 측정하고 그 순위를 갱신하는 리더보드를 운영하고 있습니다. 이는 해당 표준 데이터를 활용한 MRC 연구 성과를 공유하고, 객관적인 기준으로 알고리즘의 성능을 평가하기 위함입니다. 그리고 2020년 3월 기준, KorQuAD 1.0의 비공개 데이터셋으로 평가한 약 70여개의 언어모델 중에서 스킨터랩스의 AI MRC 언어모델은 F1 스코어 95.15점을 받아 1위를 기록하고 있습니다. (KorQuAD 1.0 기준 사람이 직접 문제를 풀었을 때의 F1 스코어는 91.2점)

KorQuAD 1.0 데이터셋을 살펴보면, 대한민국에 대한 문단 1-2개 정도 길이의 지문이 주어졌을 때 “대한민국의 수도와 그 면적은?”이라는 질문에 대해 “서울특별시, 605.25km<sup>2</sup>”라고 답변을 하는 AI 모델을 개발할 수 있도록 구성되어 있습니다. 즉 순수 텍스트만을 기반으로 단어 혹은 구 단위의 답을 할 수 있습니다. KorQuAD 2.0은 이를 기반으로 좀 더 긴 분량의 여러 문단이나 표, bullet과 같이 구조화 된 문서에 대해서도, “서울특별시의 특징은?”이라는 질문과 본문 속에 등장하는 “도시 중앙으로 한강이 흐르고 북한산, 관악산, 도봉산 등의 여러 산들로 둘러싸인…”이라는 장문의 답까지 학습시킬 수 있도록 구성되었습니다.

## KorQuAD Leaderboard

KorQuAD 1.0의 Test set으로 평가한 Exact Match(EM) 및 F1 score입니다.

Rank	Reg.Date	Model	EM	F1
-	2018. 10. 17	Human Performance	80.17	91.20
1	2020. 01. 08	SkERT-Large (single model) Skeller Labs	87.66	95.15

그림 5-1.  
KorQuAD 1.0 리더보드 1위 결과

KorQuAD 2.0의 Test set으로 평가한 Exact Match(EM) 및 F1 score입니다.

Rank	Reg.Date	Model	EM	F1
-	2019. 09. 05	Human Performance	68.82	83.86
1	2020. 06. 24	SkERT-Large (single model) Skeller Labs	76.64	88.09

그림 5-2.  
KorQuAD 2.0 리더보드 1위 결과

스킨터랩스의 AI모델은 현재 KorQuAD 1.0 기반으로 성능을 향상시켜가고 있지만, 추후 KorQuAD 2.0, 그리고 영어와 일본어 같은 다른 언어에 대해서도 완성도 높은 기계독해용 언어모델을 개발할 계획입니다.



# Why MRC?

스켈터랩스는 두꺼운 보험 약관에서 나에게 맞는 정보를 찾아 헤맨 경험, 어려운 법률 문서에 애를 먹었던 경험을 상기했습니다. 챗봇이 활용되고 다양한 분야를 꼼꼼히 살피고, 챗봇 데이터 학습에서 MRC의 가능성을 찾았습니다.

우리는 인텐트 분류 기반의 챗봇에 이어, Question Answering 기법 기반의 챗봇 개발이 가능하도록, MRC 언어 모델 고도화를 진행하고 있습니다.

# 그렇다면 어떤 경우에 Question Answering 기법을 쓰고 어떤 경우에 인텐트 분류 기법을 써야할까요?

의도 분류와 Question Answering 기법 모두 각자 맞는 챗봇 대화의 경우가 있기 때문에, 챗봇 설계 시 그 사용 목적에 따라 적합한 기술을 택하는 것이 중요합니다.

<b>Question Answering</b> 기법을 적용하기 좋은 챗봇 use-case	<b>인텐트 분류</b> 기법을 적용하기 좋은 챗봇 use-case
콘텐츠의 규모가 커서 인텐트를 만들기에 너무 많은 자원과 노력이 들어가야 하는 경우	콘텐츠의 규모가 크지 않은 경우
사용자의 질문에 1회성 답변으로 끝나며, 이후에 이어질 대화를 위해 사용자의 말에서 정보를 수집할 필요가 없는 경우	1회성 질문 및 답변으로 끝나는 것이 아닌, 사용자의 말로부터 정보를 수집하고 챗봇과 주고 받는 대화를 2세트 이상진행하는 경우
자신이 필요한 정보를 기업이 가진 정보 중에서 찾아내는 것이 챗봇 사용 목적인 경우	사람들이 물어보는 질문의 범위가 예측 가능한 경우

\*인텐트 분류에 대해서는 AIQ.TALK Story vol.01의 내용을 참고 바랍니다.

즉, 경우에 따라 인텐트 분류 기법이 적절한 경우, Question Answering 기법이 적절한 경우가 나뉘기 때문에, 사용 목적에 따라 적합한 기술을 택해 챗봇을 설계해야 합니다.

# 언어모델이 사용자가 원하는 답을 제공한다는 것은?

인텐트 기반 챗봇 구축은 사용자가 챗봇과 어떤 대화를 나누고 싶어 하는지, 그 대화를 어떤 질문들로 이끌어 나갈지를 챗봇 설계자가 파악하고 있다는 전제가 있을 때 효과적입니다. 예를 들어, “다음 중 어떤 것이 궁금하신가요?”라는 질문을 던지고 응답 가능한 항목을 보여주는 것과 같이, 사용자에게 전달할 수 있는 대화 콘텐츠를 기반으로 챗봇이 사용자와의 대화를 이끌어 나갈 수 있습니다.

MRC 기술은 범용적인 한국어 독해 능력을 갖춘 언어 모델을 바탕으로 만들어진다는 점에서, 챗봇 설계자가 챗봇이 답변 가능한 사용자의 의도를 미리 준비하지 않고, 답변의 출처가 될 지 문만을 기반으로 판단하지 않는 것이 특징입니다. 물론 이것이 가능하려면 인공지능이 사람 만큼, 혹은 그 이상의 독해 능력을 갖추어야 하므로 성능 개선의 재료가 되는 한국어 대용량 말뭉치 데이터 확보가 무엇보다 중요합니다. 특히, 사전학습 모델에 다양한 분야의, 다양한 패턴의, 실제 우리가 사용하는 자연스러운 말뭉치가 많이 학습될수록 실제 사람과 같은 독해 능력을 만들어낼 수 있습니다.

스켈터랩스는 MRC와 인텐트 분류 두가지 기술을 적재적소에 사용해서 더 높은 정확도로 사용자의 니즈를 만족시키는 방법을 지향합니다. 챗봇 제작자가 파악하고 있는 사용자 대화에 대한 니즈는 인텐트 기반 챗봇으로 구현하고, 특정 분야의 문서에서 사용자가 원하는 내용을 찾아 간단한 답변으로 충분한 영역은 MRC로 구현해서 양 기술의 장점을 모두 누릴 수 있는 것이 궁극적으로 사람을 돕는 챗봇의 형태라고 생각합니다.

# How to develop?

스켈터랩스는 기존의 기계 학습 모델과 달리, 범용 사전학습 모델을 구축했습니다. 전이학습 기반의 범용 모델은 특정 과제를 위한 학습 데이터의 양이 적더라도 빠르게 고품질의 결과물을 획득할 수 있습니다.

# 전이학습 기반 사전학습(Pre-trained) 언어 모델 구축

기존 머신러닝 방법론은 주어진 과제마다 과제에 맞게 정리된 학습 데이터만을 사용하여 학습해야 했고, 학습 데이터들은 해당 과제에 맞는 정답(Label)이 무엇인지 명확하게 알고 있어야 했습니다.

전이 학습에서도 개별 과제에 대해서는 해당 과제에 맞게 정리된 학습 데이터를 사용하는 점은 같습니다. 다만 전이 학습은 특정 과제에 대한 훈련(Task-Specific Fine-tuning)을 진행하기 전에, 범용적인 과제에 대해 미리 학습을 진행하고(Pre-training) 이렇게 도출된 고범용성의 사전학습 모델(Pre-trained model)을 기초로 다시 특정 과제에 대한 훈련(그림에서 Learning Task 1 혹은 Learning Task 2에 해당)을 한다는 점에서 차이가 있습니다. 이에 따라 특정 과제에 대한 학습 데이터(그림에서 Dataset 1, Dataset 2에 해당)의 양을 늘리지 않아도, 사전학습 모델의 성능을 향상시키는 것만으로 파인튜닝 훈련 결과물의 성능을 향상시킬 수 있습니다.

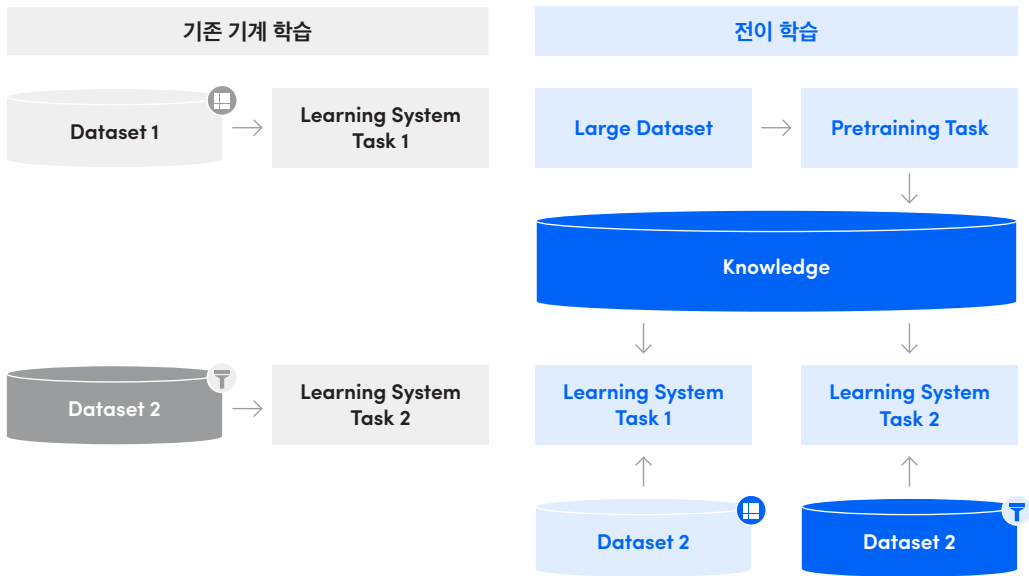


그림 6. 기존 기계 학습과 전이 학습의 비교

최근 연구에서는 방대한 양의 말뭉치로부터 자동으로 학습 데이터를 생성할 수 있는 형태의 다양한 범용 Task를 적절히 활용하면서 더욱 많은 양의 말뭉치를 학습시킬 때 더욱 강력한 사전학습 모델을 만들 수 있음을 확인할 수 있습니다. 이렇게 더욱 고도화된 사전학습 모델을 사용하면 개별 과제에 대한 성능도 함께 향상되는 식입니다.

스켈터랩스는 이미 KorQuAD 1.0 과제를 통해 한국어에 대해 가장 강력한 성능을 보이는 MRC 언어모델을 구축하였고, 이 모델의 구축에 활용된 범용 사전학습 모델을 확보하고 있습니다. 이를 보다 고도화하기 위해, 다양한 한국어 말뭉치의 확보 및 사전학습 범용 Task와의 조합, 개선된 한국어 사전학습 모델을 위한 자동화된 프레임워크 구축, 이렇게 구축된 모델을 AIQ.TALK Chatbot / AIQ.TALK MRC 모두에 적용하는 기술 고도화를 진행했으며, 발전시켜나가고 있습니다. 더불어 개별 과제에 대해서도 보다 향상된 파인튜닝 학습 결과물을 도출할 수 있도록 지속적으로 개발 중입니다.

# 전에 없던 정보검색의 시작

필요한 정보를 빅데이터 속에서 빠르게 찾아내는 능력이 중요한 경쟁력으로 떠오르는 만큼, MRC는 현재 정보 검색이 필요한 분야에서 다방면으로 활용이 가능한 유망기술입니다. 특히 일반인이 이해하기에 진입 장벽이 높은 법률, 의료, 특허, 금융, 정부 정책과 같은 지식에 대해서 누구나 쉽게 필요한 정보를 찾을 수 있도록 관련 검색 엔진에 MRC를 활용할 수 있는 것이 대표적인 예입니다.

챗봇 설계 솔루션을 제공하는 스퀘터랩스에서는 특히 MRC를 기존 챗봇의 보완해줄 수 있는 기술로서 주목합니다. 챗봇 제작자가 설계한 시나리오 내에서 사용자와의 대화가 가능한 것이 현재 산업계에서 주로 쓰이고 있는 챗봇이라면, MRC는 챗봇 제작자의 시나리오 설계가 없이도 특정 주제나 문서에 대한 데이터만을 기반으로 챗봇을 만들 수도 있게 합니다. 다만 시나리오 기반과 MRC는 어느 쪽이 낫다기보다, 각자 활용될 수 있는 분야가 다르기 때문에 각 기술이 챗봇 설계의 목적에 맞게 적절히 적용되는 것이 중요합니다.

스퀘터랩스는 다양한 활용 목적을 가진 챗봇에 범용적으로, 높은 성능으로 적용될 수 있는 자연어이해 엔진을 만들기 위해서 MRC 기술을 계속해서 고도화시켜 가고 있습니다. KorQuAD 1.0 데이터셋처럼 순수 문장으로만 이루어진 텍스트 뿐 아니라 표와 목록 등 마크업 HTML tag를 포함한 복합 텍스트를 독해하는 과제에 있어서도 사람의 독해 능력을 상회하는 언어모델을 구축해 나가고자 합니다. 그리고 궁극적으로는, 우리의 미션인 “언제 어디서나 우리의 일상을 이해하고, 도와주고, 더 나아지게 하는 머신 인텔리전스의 혁신을 이룬다”에 걸맞게 사람들이 정보 검색에 쓰는 시간을 보다 유용하고 창의적인 일에 활용할 수 있도록 필요한 정보를 효율적으로 습득하게 도울 수 있을 것이라 믿습니다.

# See you soon

본 보고서는 스텔터랩스의 MRC 기술에 대한 기본적인 이해를 돕기 위해 작성되었으며, 이를 통해 독자 여러분의 대화형 인공지능에 대한 관심을 높이고자 합니다. 앞으로도 스텔터랩스의 인공지능 기반 대화 엔진을 완성하는 다양한 기술 컴퍼넌트에 대해 소개드릴 예정이니 지속적인 관심 부탁드립니다.

## AIQ.TALK STORY

vol.02

MRC(Machine Reading Comprehension)

발행일 2021. 10. 13  
발행처 스텔터랩스  
발행인 스텔터랩스 마케팅 팀  
디자인 차지연  
메 일 [contact@skelsterlabs.com](mailto:contact@skelsterlabs.com)

