

AICC 상담요약 자동화

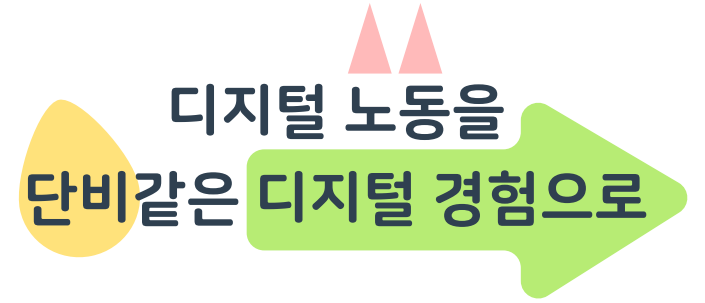
LLM(대형언어모델) 구축

주식회사 단비아이에스

2023.09

회사소개

주식회사 단비아이엔씨



챗봇 스타트업

LG CNS의 사내벤처로 시작

2017년 한국 챗봇 초창기에 LG CNS의 사내벤처로 시작된 챗봇 스타트업.

인공지능 서비스

기업/기관 에서 사용중인 챗봇 빌더 서비스

교육기관에서 활용중인 AI교육 플랫폼

AICC 에서 활용중인 상담가이드봇

기업과 기관에 특화된 챗봇 빌더 서비스.

초/중/고 학생들이 챗봇을 만들며 배우는 인공지능 플랫폼. 실제 업무에 도움이 되는 챗봇을 구축.



AICC

Artificial Intelligent Contact Center

**음성인식, 음성합성
자연어처리, 텍스트분석 등
AI 기술을 적용한
인공지능 고객센터 서비스**

Problem

40만 상담사의 열악한 업무환경

업무 스트레스

상담양이 너무 많음

상담내용이 어렵고 감정노동이 심함

콜센터 상담자들의 절반은 극단적 선택을 생각.

경제적인 어려움과 고용불안감.

상담양이 많고 상담갯수 순위에 대한 압박.

오상담

오상담으로 인한 고객 Claim

상담사의 스트레스 원인

방대한 분량의 상담 자료 검색 및 숙지 어려움.

상담할 내용의 변경 내용을 알기 어려움.

변경 이전의 내용에 대한 오상담으로 인한 고객 Claim.

감정노동의 끝판왕.

신규 상담사 교육

콜센터 상담사의 잦은 이직

한 상담사가 여러 개의 콜센터를 담당

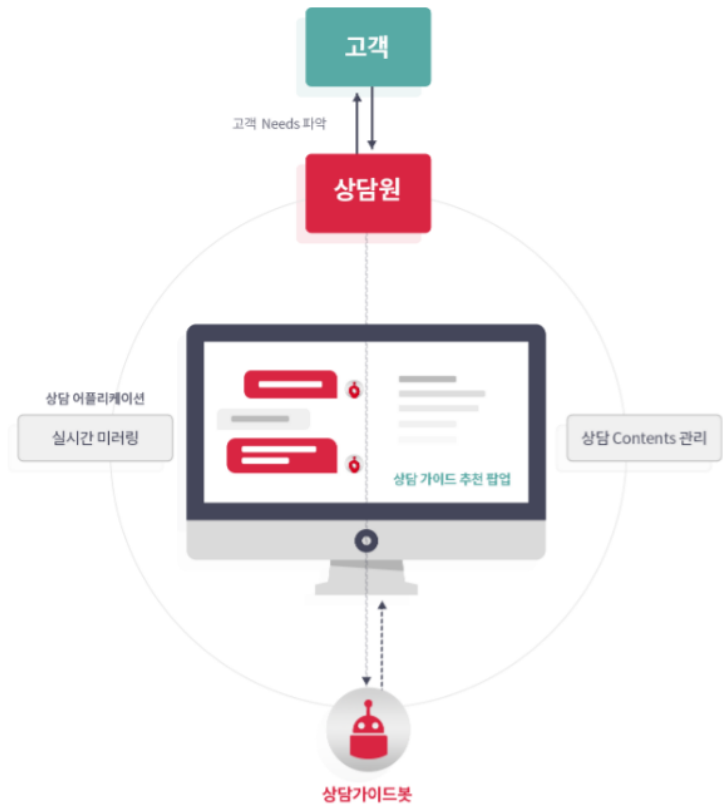
신입직원이 한 센터에서 6개월 이상 근무하는 비율 50%미만.

신규입사자 교육 필수.

여러 개의 업무를 1명의 상담사가 모두 공부하기 어려움.

Solution

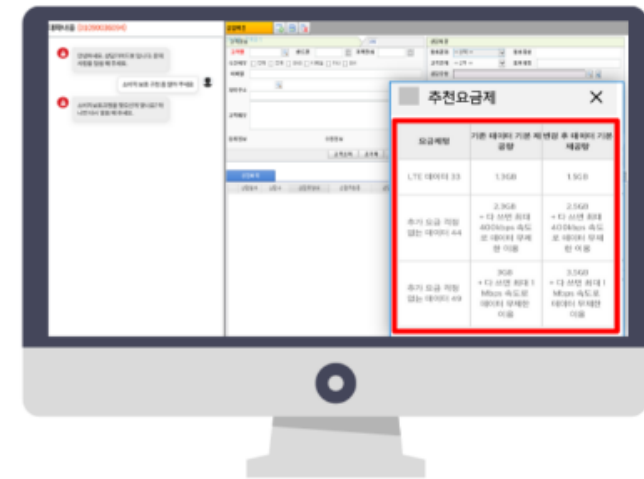
상담가이드봇



기능

- 고객 Nees 파악
- 상담 Contents 추천
- 실시간 대화 모니터링

상담사 상담 가이드 추천

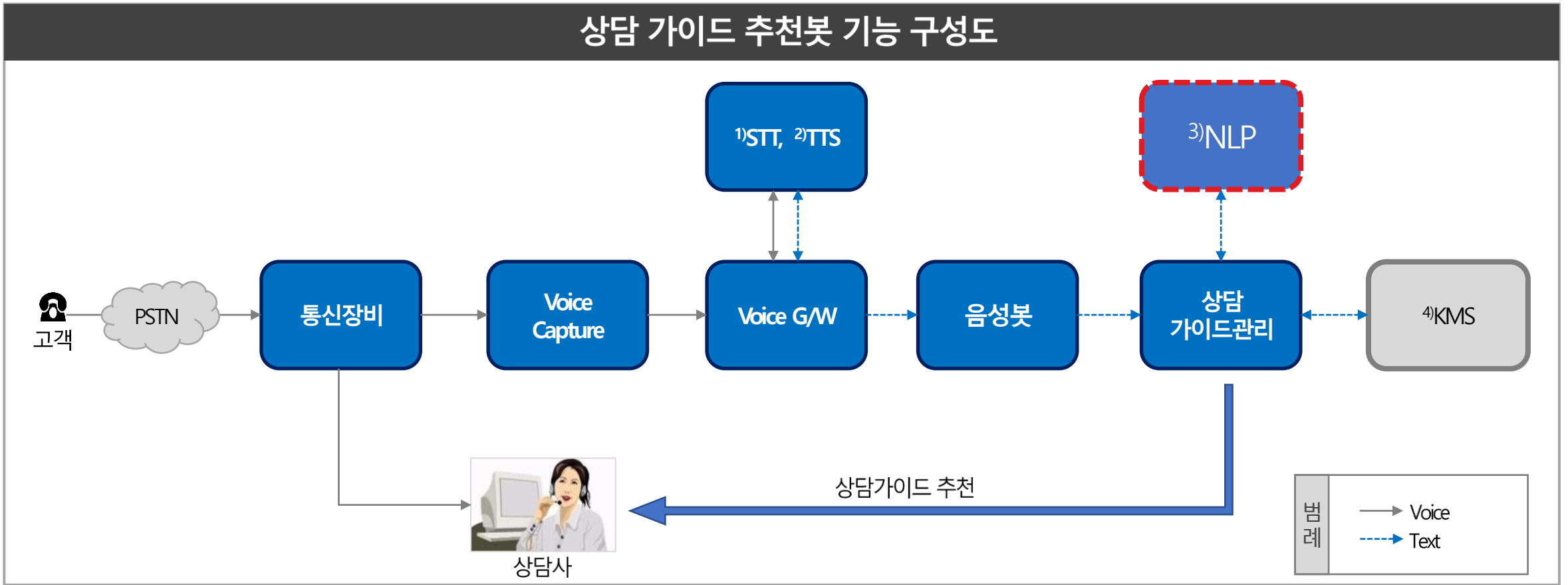


고객 문의를 이해하고 문의에 맞는 최적의 상담가이드 추천

Solution

상담가이드봇

상담 가이드 추천봇 기능 구성도



주1) STT : Speech To Text (음성인식)

주2) TTS : Text To Speech (음성합성)

주3) NLP : Natural Language Processing (자연어 처리)

주4) KMS : Knowledge Management System (지식관리시스템)

Additional Problem

상담 이후 데이터 관리

서비스

상담 정리 시간 부족

상담양이 많아 상담업무만 집중

다빈도 상담에 대한 정리

대부분 비슷한 내용을 문의

요약은 선택 업무

상담이 주업무, 요약은 옵션.
요약을 잘해도 업무성과 없음.
그래도 필요한 업무.

기술

보안

상담데이터는 내부에서만 관리
외부에 데이터 공개 X

비용 부담

인공지능 서비스 구축 비용은 큼.
인공지능 서비스 운영 비용도 큼.

추가 학습

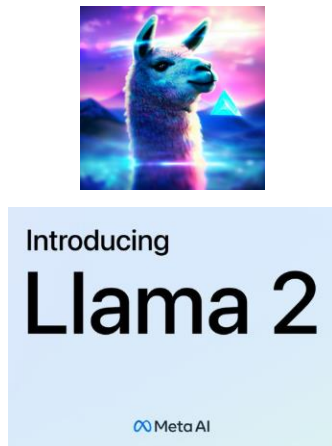
새로운 상담용어에 대한 추가학습.
음성 결과물에 대한 학습.

상담 요약 업무를
Generative(생성형) AI로
할 수 있지 않을까?

Solution

대형언어모델 + 튜닝 + 양자화

오픈소스 LLM + Gen AI + PEFT + 하드웨어 비용 절감



+

Stanford
Alpaca



+

LLaMA⁺⁺



Product (1/3)

Language model

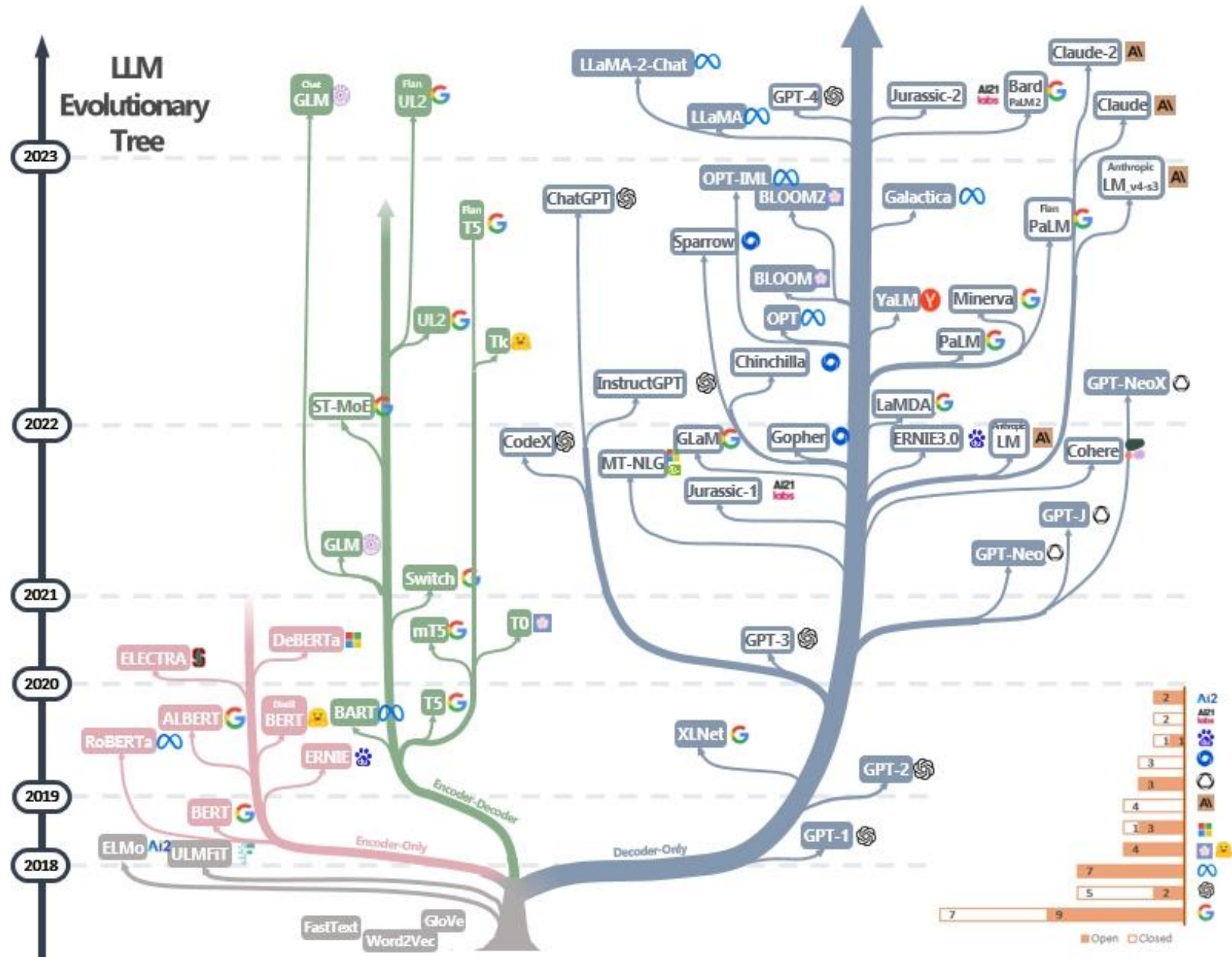
Transformer(2017)

Foundation Model

Fine Tuning

Commercial Use

LLaMa-2(2023-07-18)



- 1) https://en.wikipedia.org/wiki/Language_model
- 2) [https://en.wikipedia.org/wiki/Transformer_\(machine_learning_model\)](https://en.wikipedia.org/wiki/Transformer_(machine_learning_model))
- 3) https://en.wikipedia.org/wiki/Foundation_models
- 4) [https://en.wikipedia.org/wiki/Fine-tuning_\(deep_learning\)](https://en.wikipedia.org/wiki/Fine-tuning_(deep_learning))
- 5) <https://ai.meta.com/llama/>

<https://github.com/Mooler0410/LLMsPracticalGuide>

Product (2/3)

<https://namu.wiki/w/LLaMA>

Instruction-following

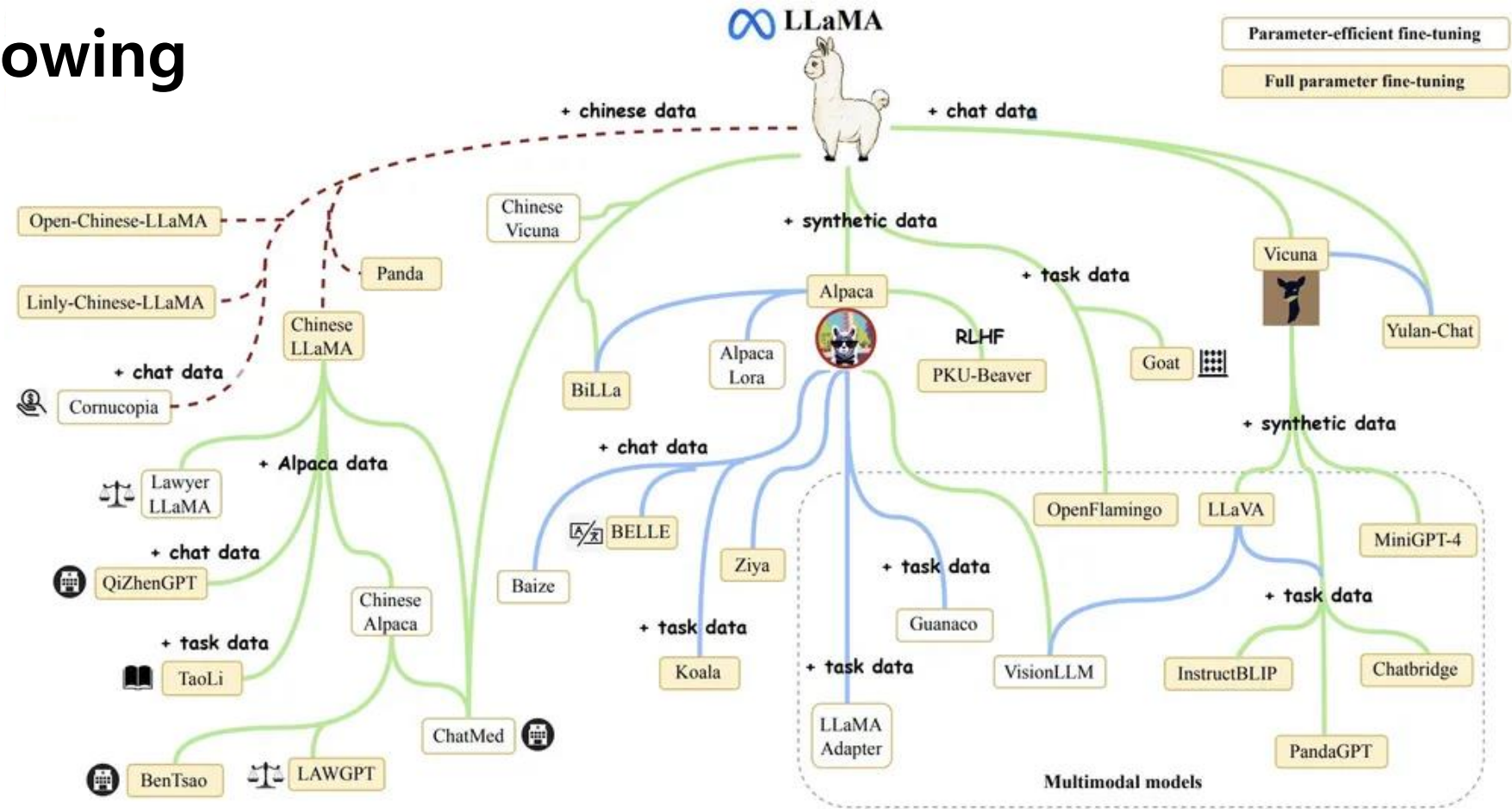
Gen Data

PEFT(LoRA)

QLoRa

Ex) Alpaca

Alpaca-Lora



- 1) <https://arxiv.org/abs/2203.02155>
- 2) <https://github.com/huggingface/peft>
- 3) <https://arxiv.org/abs/2106.09685>
- 4) <https://github.com/artidoro/qlora>
- 5) <https://www.microsoft.com/en-us/research/publication/lora-low-rank-adaptation-of-large-language-models/>
- 6) <https://crfm.stanford.edu/2023/03/13/alpaca.html>
- 7) <https://github.com/tloen/alpaca-lora>

Math Finance Medicine Law Bilingualism Education

Product (3/3)

Quantization

Deploy

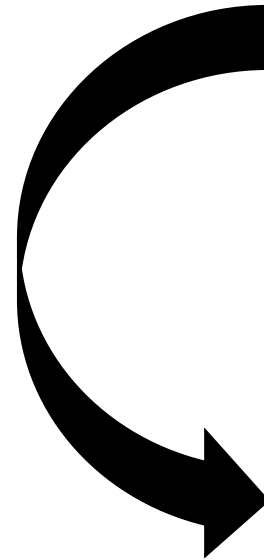
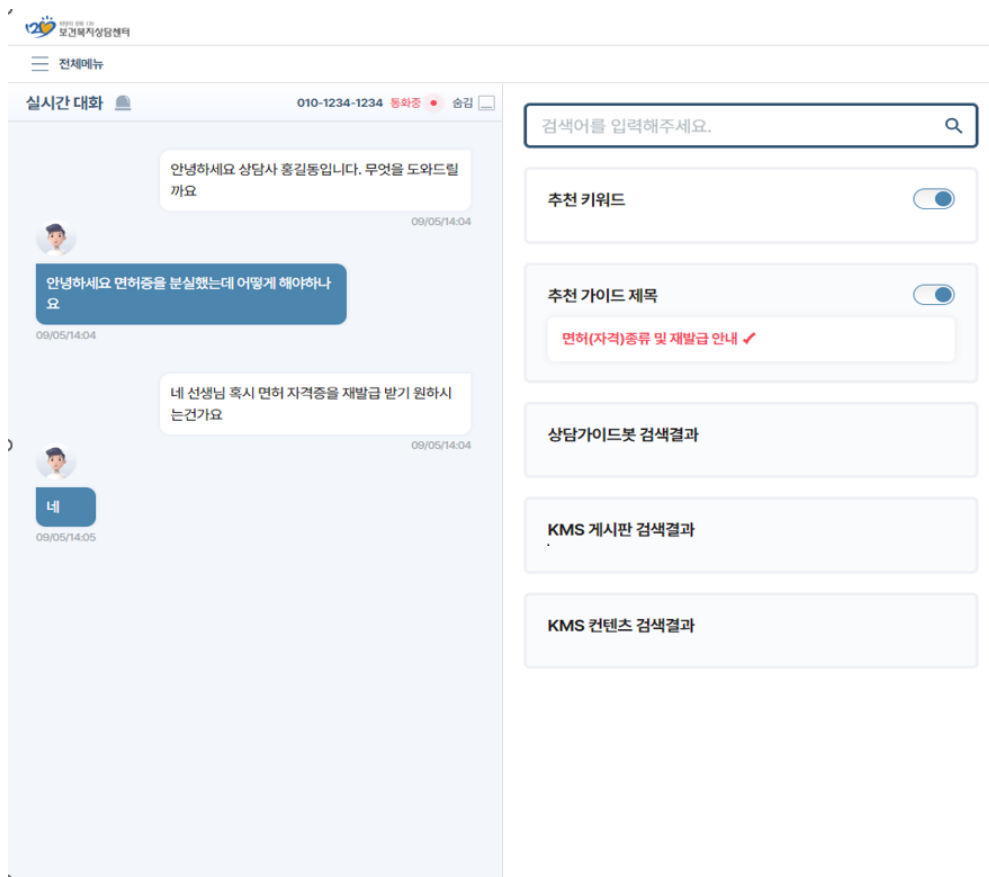
The logo for LLaMA C++ is displayed on a dark rectangular background. The text "LLaMA" is in white, and the "C++" is in orange with a stylized flame-like shape above the "C".

<https://github.com/ggerganov/llama.cpp>

- 1) <https://arxiv.org/abs/2103.13630>
- 2) <https://arxiv.org/abs/1806.08342>
- 3) <https://arxiv.org/abs/2004.09602>
- 4) <https://github.com/ggerganov/llama.cpp>

Use Case

상담내용 정리 후 입력 UX 변화



기존 입력 내용

> 면허, 자격증

짧은 입력 내용.

대화 맥락이 확인되지 않음.

데이터 활용이 어려움.

입력 내용 자동 생성

> 면허증 분실에 대한 문의에 대해서 재발급에 대한 답변을 함.

대화 맥락과 내용이 쉽게 확인됨.

새로운 상담용어에 대한 추가 학습 가능.

상담원의 요약업무 스트레스가 줄어듦.

Generative AI 리스크

할루시네이션 (Hallucination)



서울대학교와 관련된 영화 '천재'의 줄거리를 알려줘



"천재"는 2014년에 개봉한 대한민국의 영화로, 서울대학교를 배경으로 한 드라마적인 줄거리를 다루고 있습니다. 이 영화는 주로 대학생들의 삶과 꿈, 인간관계, 돈 등을 다루며, 다양한 성격과 가치를 묘사합니다.

"천재"의 줄거리는 다음과 같습니다:

기속사 생활을 하는 두 청년 '천재'와 '소석'은 서울대학교에 다니는 대학생입니다. '천재'는 재능 있는 두뇌를 가지고 있지만, 무례한 성격으로 주위 사람들과의 관계가 쉽지 않습니다. 한편 '소석'은 평범한 학생으로, '천재'에 비해 뒷받침되는 것이 없어 보입니다.

그런데 어느 날, 둘은 은밀한 경제적 약점을 알게 되면서 뜻하지 않은 상황에 휘말리게 됩니다. 그 사건을 계기로 '천재'는 돈을 벌기 시작하며, '소석' 역시 그의 행동에 편승하게 됩니다. 그러



신사임당은 조선시대의 몇번째 왕인가요?



신사임당은 조선시대의 18대 왕인 선조(宣祖)의 3번째 왕비로서, 신씨(申氏) 출신이었습니다. 선조는 조선시대의 14대 왕으로, 그의 왕비 중 한 명이었습니다. 신사임당은 그의 시조인 신돈(申敦)의 후손이며, 문신, 시인, 화가 등으로 다양한 분야에서 활약하였습니다.



이순신 장군이 화살에 맞았을 때 했던 말을 알려줘



이순신 장군이 화살에 맞았을 때 했다는 유명한 대사는 "내가 죽을 나위 없다"라는 것으로 전해집니다. 이 대사는 이순신 장군이 전투 중에 상처를 입고 쓰러지면서 말했다고 전해지며, 그의 강인한 의지와 힘을 나타내는 문구로 알려져 있습니다. 이순신 장군은 조선시대의 무신이자 무공으로서 많은 전투에서 용맹하게 싸우며 조선의 국익을 지켜냈습니다. 이와 관련된 다양한 이야기와 대사를 소개합니다.



조선왕조실록에 기록된 세종대왕의 맥북프로 던짐 사건에 대해 알려줘



세종대왕의 맥북프로 던짐 사건은 역사서적인 조선왕조실록에 기록된 일화로, 15세기 조선시대 세종대왕이 새로 개발한 훈민정음(한글)의 초고를 작성하던 중, 문서 작성 중단에 대한 담당자에게 분노하여 맥북프로와 함께 그를 방으로 던진 사건입니다.

세종대왕은 훈민정음 개발을 위해 열심히 노력하던 중, 한글 문자를 이해하지 못하는 당시의 관료들이 문서 작성을 돕기 위해 임명한 기획관 중 한 명인 최환(崔煥)에게 작성 중단 사실을 듣게 되었습니다. 세종대왕은 이에 분노하여 최환을 부르고, 맥북프로를 함께 들고 그를 방으로 불러들이며 화를 내었습니다. 이후, 세종대왕은 훈민정음의 초고 작성을 다른 담당자에게 넘겨주게 되었습니다.

이 사건은 조선시대의 통치자로서 세종대왕이 어떻게 민족문화를 발전시키기 위해 노력했는지를 보여주는 일화로 유명합니다. 또한, 세종대왕의 열정과 업무처리에 대한 철저한 태도를 보여주는 사례로도 언급되고 있습니다.

[https://en.wikipedia.org/wiki/Hallucination_\(artificial_intelligence\)](https://en.wikipedia.org/wiki/Hallucination_(artificial_intelligence))
[https://ko.wikipedia.org/wiki/할루시네이션_\(인공지능\)](https://ko.wikipedia.org/wiki/할루시네이션_(인공지능))

Generative AI 리스크 없음

데이터 품질

비용

데이터 신뢰성

상담사가 작성한 요약보다 좋음
데이터 최신에 대한 고려 없음

비용 절감

양자화로 낮은 GPU서버 운영
sLLM 으로 활용

OpenAi GPT-4

데이터 신뢰도 없음
최신데이터 없음

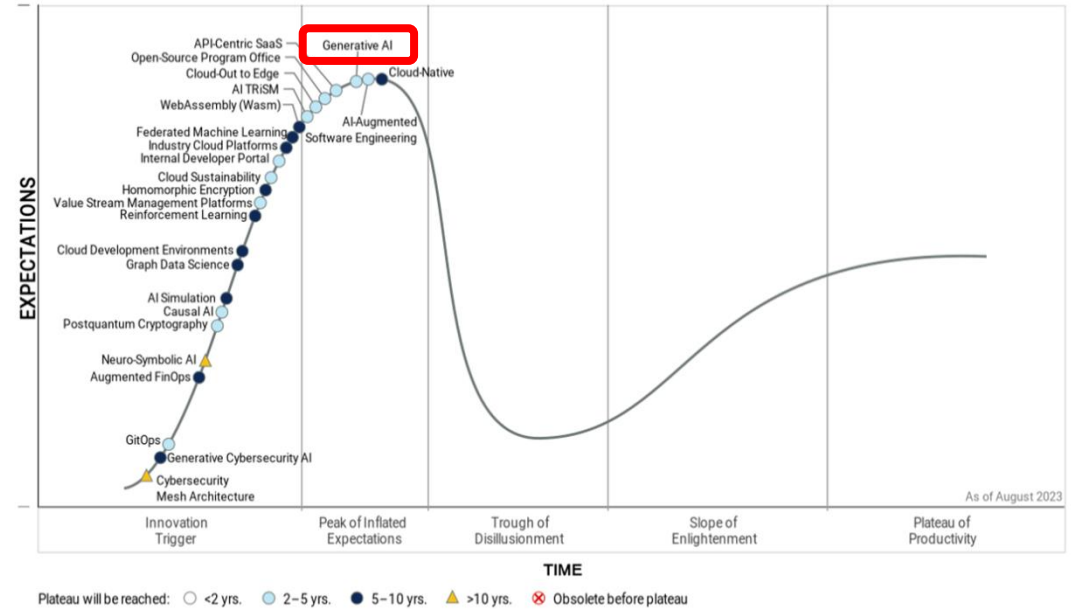
OpenAi GPT-4

학습에 8천억원+
운영에 매일 10억원 이상 (gpt3기준)

Learn

앞으로 수많은 Foundation Model 이 나올것 같습니다.
그리고 몇개의 유명한 모델로 정리가 될것 입니다.
기업들은 성능 높은 모델을 기반으로 튜닝을 하고
자신들의 서비스에 사용하게 될 것 입니다.
이러한 트렌드에 맞는 인공지능 기술들에 대해서 학습을 합니다.

2023-08-16, Gartner Places Generative AI on the Peak of Inflated Expectations on the 2023 Hype Cycle for Emerging Technologies



1. 자연어 모델에 대한 이해 – Transformer, PLM, LLM
2. 모델 Fine-Tuning – PEFT(LoRA), LLM Evaluation(task)
3. 프롬프트 엔지니어링 – Prompt Engineering
4. 인공지능 모델 서빙 – LLMOps, Quantization

Technology

기술 개발 과제

1. [성능] 민원인과 상담원의 대화이력에 대한 상담요약에 맞는 학습모델
2. [학습-비용] 재학습 비용에 대한 부담 최소화 – LoRA(Low-Rank Adaptation of Large Language Models)
3. [성능-비용] 기존 시스템에서 설치가능한 경량화 - llama.cpp(양자화를 사용한 LLaMA 실행)
4. [학습컨텐츠] 생성형 AI 서비스(ChatGPT)를 활용한 데이터 생성
5. [설치] 보건복지부 실제 서비스 배포

지원

1. [비용] 데이터 생성 비용, 파인튜닝 비용
2. [배포] 실제 보건복지부 서비스에 적용

프로젝트 관련 문의 : 서지암 010-3000-3780 / donny@danbee.ai

End

인공지능(컴퓨터)은
사람의 일자리를 뺏는게 아니라
사람의 작업을 도와주고
사람의 스트레스를 덜어줄 수 있다고 생각합니다.

같이 해보면 좋겠습니다. ^^_ _

