

네이버가 도전하는 AI 서비스

한국어 초거대 언어 모델을 통한 AI 혁신의 길을 기대하다

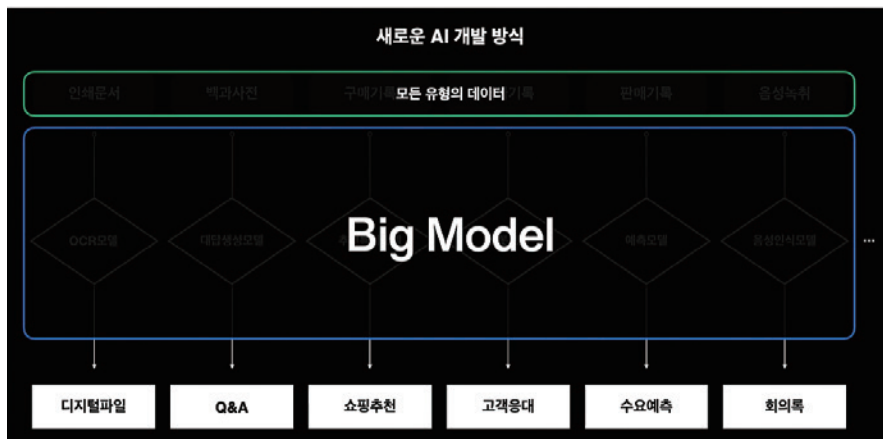
글. 최홍규 EBS 정책기획부 연구위원 / 미디어학 박사

네이버가 지향하는 AI 서비스는 어떤 모습일까? 이번 원고에서는 지난 5월 네이버가 발표한 내용을 통해, 네이버가 꿈꾸는 AI 서비스에 대해 논의해보고자 한다. 또한, AI 서비스를 기술적으로 잘 알지 못하는 이용자 관점에서 네이버 AI 서비스가 과연 혁신의 바람을 일으킬지 예측해보고자 한다.

국내 포털 1위 사업자의 AI 브랜드 : 하이퍼클로바(HyperCLOVA)

네이버는 지난 2021년 5월 25일 'NAVER AI NOW'라는 동영상상을 공개한다. 아직 우리나라의 웹/모바일 서비스 이용자들에게는 생소한 형식의 동영상인데, 네이버에서 새로 출시할 AI 서비스에 대한 설명 영상들이다. 동영상들은 온라인 세미나 형태로 구성되어 있고, 발표자들은 모두 이번 네이버에 소속된 임직원이자 AI 서비스 개발에 참여한 사람들이다. 동영상들을 자세히 살펴보니, 국내 포털 1위 사업자가 AI 서비스에 대한 화두를 던지는 것이 어서인지 혁신적으로 보이고 야심에 가득 찬 모습들도 보인다.

동영상들에 반복적으로 등장하는 단어 '하이퍼클로바(HyperCLOVA)'. NAVER AI NOW는 이 서비스명 하나를 위해 만들어진 동영상으로 보인다. 물론, 네이버의 하이퍼클로바는 단순히 작은 AI 서비스 하나를 뜻하는 것이 아니다. 하이퍼클로바는 네이버가 추구하는 AI 서비스의 새로운 이름이자, 혁신의 방향, 비전 등을 모두 담고 있는 개념인 것이다.



Big Model을 통한 새로운 AI 서비스 개발 방식¹⁾

1) 본 글의 모든 그림은 NAVER TV에 업로드된 <NAVER AI NOW>에서 발췌함. URL: <https://tv.naver.com/ainow>

하나의 AI 서비스를 만들어내기 위해서, 기존에는 서비스별로 각각의 데이터를 모으고 이를 서비스에 맞춰 엔지니어링 하는데 많은 시간이 들었다. 그러나 ‘빅모델(Big Model)’을 활용하면 이러한 시간을 획기적으로 줄일 수 있고 데이터를 공유하여 각각의 서비스에 활용하는 일도 가능하다고 한다. 네이버는 바로 이러한 빅모델을 적용하여 하이퍼클로바를 개발했다고 한다.

하이퍼클로바를 가능하게 만드는 요소는 3개로 요약할 수 있는데, 슈퍼컴퓨터(Super Computer), 데이터(Data), AI 전문가(AI Specialist)라고 한다. AI 서비스를 개발하는 데는 고효율로 데이터 처리가 가능한 인프라나 대용량의 데이터가 필요하니 슈퍼컴퓨터나 데이터는 그 요소로서 이해가 간다. 그런데 네이버는 AI 전문가를 어떻게 육성할 수 있었을까? 바로 연구개발에 대규모 리소스를 투입해 연구성과를 발굴했다고 한다. 네이버가 AI 분야에서 이룬 학문적 성과는 국내에서는 가장 많은 성과로, 글로벌 학회에서 많은 논문을 발표하며 그 성과들을 축적했다. 2020년에 발표한 논문 수가 43편이고 2021년에는 더 많은 논문을 발표한다고 한다. 이러한 연구개발 성과는 전 세계에 연결된 네이버의 ‘글로벌 AI 연구 생태계’도 한몫했다. 대규모 리소스 투입, 그에 걸맞은 연구성과, 글로벌로 연결된 지식개발 체계. 이 모든 것이 어우러져 네이버 AI 서비스에 특화된 AI 전문가를 육성한 것이다.



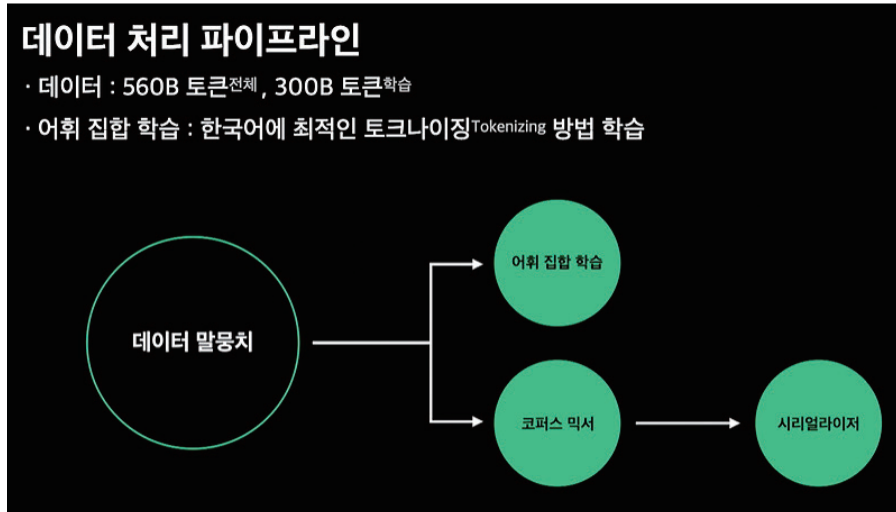
전 세계에 구축된 네이버의 AI 연구 생태계

한국어 초거대 언어 모델로 만드는 AI 서비스

그렇다면, 하이퍼클로바는 대체 어떠한 서비스인가? ‘한국어 초거대 언어 모델로 만든 AI 서비스’다. ‘한국어 초거대 언어 모델’이라... 용어가 다소 어려운데, 앞서 언급한 빅모델을 한국어라는 언어를 기반으로 만들고 이 모델을 기반으로 해서 다양한 AI 서비스를 만들어냈다고 이해하면 될 것이다.

네이버는 한국어 초거대 언어 모델 구축을 위해 2020년 12월 슈퍼컴퓨터를 도입한다. 이로써 세계 최대의 한국어 언어 모델은 기존보다 단축된 140분의 1 정도의 시간 안에 개발되기에 이르렀다. 네이버에서 구축한 기존 언어 모델보다 600배가 큰 언어 모델을 개발하면서 그 시간을 획기적으로 단축하는 데 성공한 것이다.

네이버의 AI 언어 모델에 활용된 한국어 데이터는 5,600억 토큰으로 구성되었는데, 이는 뉴스 50년 치나 네이버 블로그 9년 치에 해당하는 대용량의 데이터다. 이러한 데이터 말뭉치에 ‘어휘 집합 학습’의 과정을 적용하고, 전처리 과정을 효율화하는 ‘코퍼스 믹서(Corpus Mixer)’를 적용하며, 모델이 데이터를 2진수로 읽을 수 있도록 ‘시리얼라이저(Serializer)’ 과정을 거치도록 하여 한국어 기반의 언어 AI 서비스를 완성할 수 있었다고 한다.



한국어 초거대 언어 데이터 처리 파이프라인

이렇게, 한국어 초거대 언어 모델로 완성된 네이버 AI 서비스의 특징점은 아래와 같다.

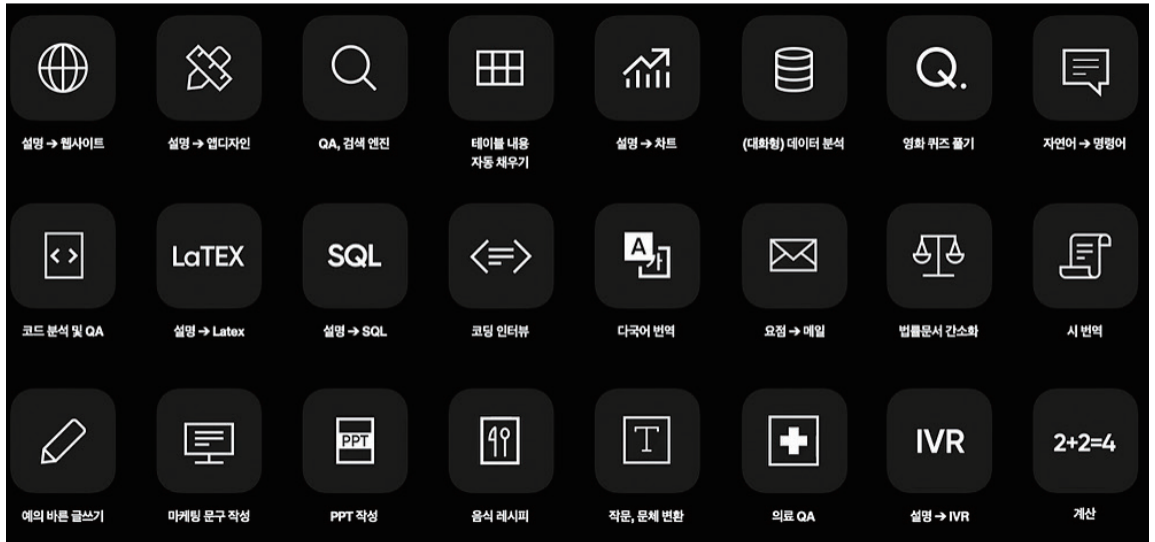
- 대용량 데이터를 활용한 초대형 한국어 언어 모델 학습
- 한국어 특성에 부합하는 어휘 집합 적용
- 언어 모델의 유창성을 평가하는 판별 정확도 지표 제안
- 더 크고 유창한 언어 모델을 구축하는 중

한국어 초거대 언어 모델을 활용한 네이버 AI 서비스의 특징점

요약하자면, 금번 네이버의 하이퍼클로바는 대용량의 한국어 데이터를 활용해 만들어진 서비스이고, 한국어 특성에 부합하는 어휘 집합을 적용하여 AI가 기존보다 한국어를 알아듣고 이해하고 말하는 것도 쉽도록 개발된 서비스다. 또한, 하이퍼클로바에 활용한 언어모델은 판별 정확도와 같은 지표를 통해 유창성을 평가할 수 있고 이러한 지표를 통해 더욱 유창한 서비스로 업그레이드할 수도 있다. 그런데 네이버는 현재도 더 크고 유창한 언어 모델을 지속하여 구축하고 있다. 하이퍼클로바는 현재진행형의 서비스인 것이다. 계속 진보하고 혁신적인 서비스로 거듭날 것이라고 한다.

이러한 하이퍼클로바 서비스가 제공하는, 실질적으로 이용자가 체감할만한 서비스는 무궁무진하다. 특정한 개념의 설명을 해주고 차트를 만들고, 질문에 답하고, 코드를 분석하고, 언어를 번역하고, 글을 요약하고, PT를 만들고, 다양한 포맷의 글을 쓰는 등. 대용량의 한국어 데이터를 활용해 지능화한 서비스가 제공해주는 일들은 계속해서 생겨난다.

ICT 서비스 이용자에게 체감되는 AI 서비스로 대표적인 것이 'AI 스피커'인데, 하이퍼클로바를 활용한 대화 서비스는 20회 이상의 문장 맥락을 이해하고 그 답변이 가능한 서비스라고 하니, AI 스피커의 기능이나 역할도 획기적으로 바뀔 것으로 예상된다. 즉 이용자에게 체감되는 AI 서비스의 모습이 지금과는 전혀 판판으로 바뀐다는 얘기다.



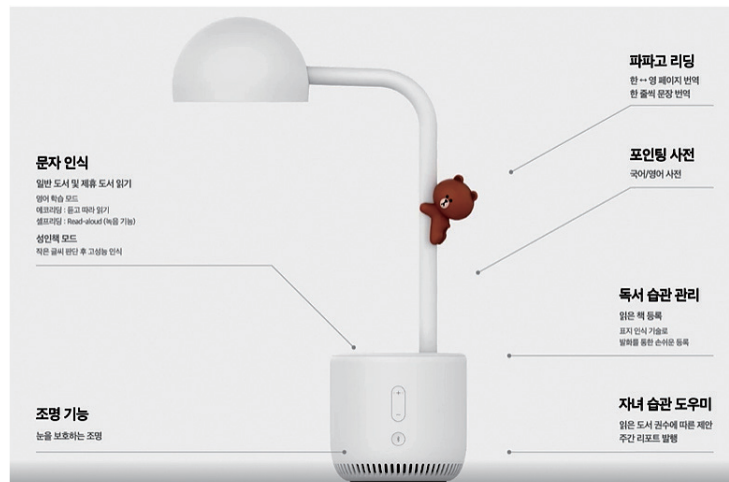
네이버의 언어 데이터 기반 'BIG AI'가 제공할 수 있는 서비스 사례

‘AI 한류’라는 새로운 붐이 시작될 수 있을까?

ICT 분야에서 아무리 좋은 서비스가 나온다고 해도, 이용자의 머리에 그려지는 이미지가 없으면 성공할 수 없다. 이제, 네이버가 발표한 하이퍼클로바가 한국어 데이터를 기반으로 기존과는 다른 AI 서비스를 만든다는 것은 이해할 수 있다. 그럼 혁신이라는 것은 어떠한 방식으로 더욱 체감될 수 있을까?

네이버가 출시한 클로바 램프의 다양한 기능을 보면, 네이버가 추구하는 AI 서비스를 빠르게 체감할 수 있다. 흔히 책상에 놓인 램프라고 하면, 공부나 글을 읽은 과정에서 책이나 노트를 밝혀주며 눈의 피로도를 낮추는 기능을 해왔다. 그런데, 네이버의 클로바 램프는 문자를 인식하여 책을 읽어주고 녹음도 해주며, 번역도 해주고, 사전

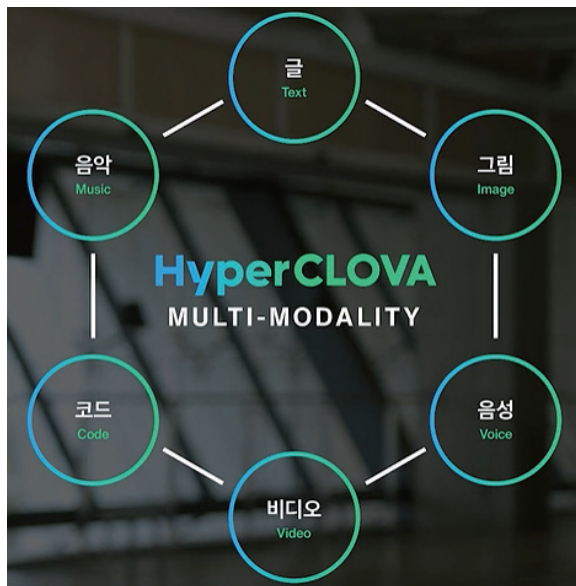
을 제공하는가 하면, 읽은 책을 등록하여 독서 습관을 관리해주시기도 한다. 형성된 독서 습관에 관한 리포트도 제공해준다고 한다.



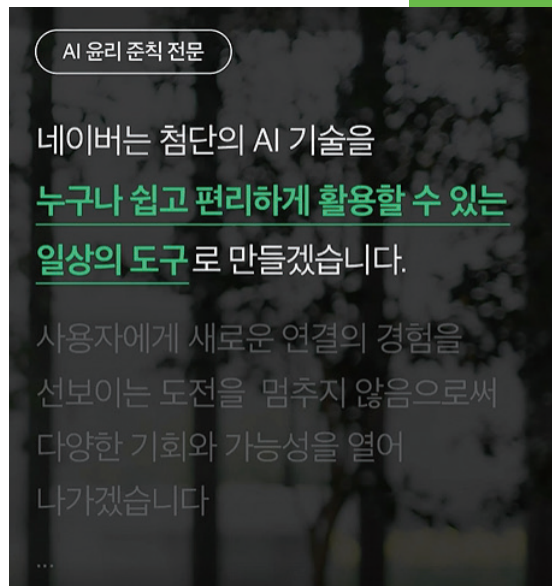
네이버가 출시한 클로바 램프의 다양한 기능

한국어 데이터를 활용한 하이퍼클로바의 서비스가 무한한 가능성을 지니는 이유는 한국어를 활용하는 일상에서 어떠한 서비스에도 활용할 수 있기 때문이다. 즉 글, 그림, 음악, 음성, 코드, 비디오를 매개로 한 어떠한 한국어 서비스에도 적용할 수 있다.

하이퍼클로바에 거는 기대감은, 국내 포털로서는 대표선수 격인 네이버가 출시한 서비스라서 더 크다. 오늘날 우리나라는 드라마, 영화, 음악, 게임 등 콘텐츠 측면에서 한류의 위상을 쌓아가고 있다. 방탄소년단이 태어난 나라가 우리나라이니 말이다.



하이퍼클로바(HyperCLOVA)로 확장 가능한 서비스 형태



네이버의 AI 윤리 준칙 내용

그런데, 이러한 한류의 분위기가 ICT 서비스 분야에서도 한 번쯤 나오면 좋겠다는 생각이다. 네이버가 이번에 야심 차게 준비한 하이퍼클로바는 단순한 AI 서비스가 아니고 포털 서비스를 통해 축적된 데이터와 노하우가 결합한 서비스이다. 대용량의 한국어 데이터를 기반으로 했기 때문에 다양한 한류 콘텐츠들과 결합하여 더 큰 시너지를 낼 수도 있다고 본다. 한류에 영향을 받은 전 세계의 서비스 이용자들은 한국어에 관심이 있는 이용자이기에 이들의 구미를 당길만한 서비스라고도 할 수 있다. 다만, 네이버의 AI 윤리 준칙의 문구처럼 이번에 발표된 하이퍼클로바가 '누구나 쉽고 편리하게 활용할 수 있는' 기술을 담고 있다면, 그리고 그것이 전 세계 이용자들에게 체감될 수 있다면, 'AI 한류'도 어렵지 않은 일이 될 것이다.

필자는 기대한다. 이번에 네이버가 하이퍼클로바로 AI 서비스 혁신을 제대로 이룬다면, 그것이 우리나라의 콘텐츠나 미디어 시장에 큰 활력을 줄 수 있을 것으로 말이다. 📺