



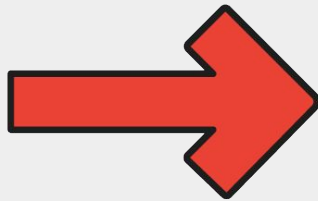
Google Developer Group  
Editable University Name



deepseek

# What is DeepSeek?

**GPT** 수준의 추론 능력을 어떻게 가지게 되었을까



# DeepSeek-R1이란?

- DeepSeek-R1은 AI가 스스로 생각하고 문제를 해결할 수 있도록 훈련된 모델
- 복잡한 문제(예: 수학 문제, 코딩 문제, 사실 기반 질문 등)에 대해 논리적이고 일관된 답변을 제공

## 목표

- GPT와 같은 최신 AI 모델과 유사한 추론 능력을 목표



deepseek

# 스스로 생각하는 과정이란?

ChatGPT o1 pro

공유하기

PRO

세부 정보

홈페이지 웹사이트를 코드를 작성해줘 매우 복잡하게

o1 pro mode에 요청  
코드 조각 작성 중...

세부 정보

구현할 코드 작성 중

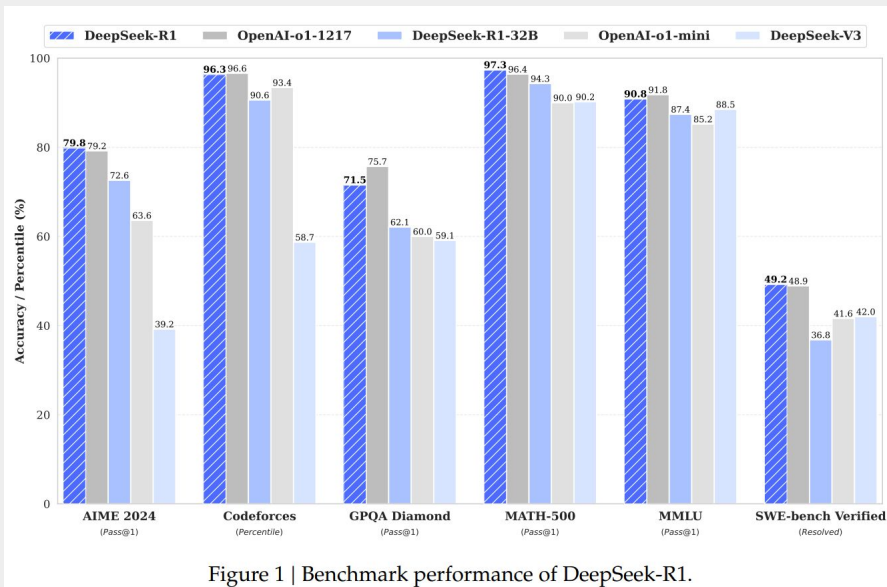
Korean 사용자가 요청한 "매우 복잡한 코드"에 대해, 시각적으로 다양한 섹션을 포함하고, 정적 및 동적 효과를 구현하고 있어.

코드 조각 작성 중

응답형 디자인을 통해 <html>, <head>, <style>, <script> 등을 포함한 코드를 제공하며, 고급 CSS 애니메이션, 파라렉스 스크롤링, 플로팅 웨이프 애니메이션, 폼 검증 등을 구현하고 있어.

“여러 가지 퍼즐 조각을 모아 하나의 그림을 완성하는 사람”처럼, 문제의 여러 단서를 모아 최종 답을 도출

# DeepSeek vs GPT-o1



American Invitational Mathematics Examination(AIME)  
고난도 수학 문제들을 모은 벤치마크

Codeforces

전 세계 프로그래머가 참가하는 알고리즘 대회 플랫폼

MATH-500

초중고급 수준의 수학 문제 500개

Massive Multitask Language Understanding

47개 과목문제로 구성된 지식 능력 평가 벤치마크

SWE-bench

소프트웨어 엔지니어링 관련 벤치마크

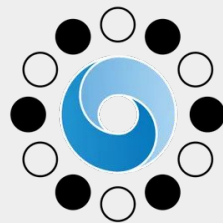
# 강화 학습(Reinforcement Learning)

강화 학습은 AI가 시행착오를 통해 스스로 배우는 방법

올바른 답을 내면 '보상'을, 틀리면 다시 시도하도록 하여 점차 실력 향상

예시

- 자전거를 처음 배우는 아이가 넘어지고 다시 일어나며 타는 법을 익히는 것과 비슷
- 알파고 : 이기면 보상을 제공하여 더 나은 바둑을 두도록 함



AlphaGo

# DeepSeek-R1-Zero 스스로 배우다

DeepSeek-R1-Zero는 기존의 지도 학습 없이 오로지 시행착오(강화 학습)만으로 스스로 문제를 해결하는 법을 학습

이 과정에서 AI는 스스로 “어떻게 생각할지”를 학습

→ 인간이 직접 방향을 제시하는 것보다 오히려 효과적으로 학습

예시

- 처음 보는 요리 레시피 없이 여러 번 시도하며 자신만의 요리법을 찾는 셰프와 동일
- 수험생에게 문제푸는 방법을 알려주지 않고 직접 해결하도록 함

# DeepSeek-R1-Zero 보상 방법

## 평가 및 보상

- 한 문제에 대해 여러 가지 답을 만들어내고, 그 중 가장 좋은 답을 선택하도록 훈련
- 문제를 해결할 때 “정답인지”와 “어떻게 답변했는지”에 따라 보상을 받습니다.

→ 정확한 답변과 명확한 사고 과정을 모두 평가하여, 더 나은 답변을 하도록 유도

# DeepSeek-R1-Zero 오답 노트

문제 해결 전 과정을 기록하도록 훈련

→ 단순히 답을 내는 것이 아니라 그 과정을 개선함

‘Aha Moment’의 중요성

Question: If  $a > 1$ , then the sum of the real solutions of  $\sqrt{a - \sqrt{a+x}} = x$  is equal to

Response: <think>

To solve the equation  $\sqrt{a - \sqrt{a+x}} = x$ , let's start by squaring both ...

$$\left(\sqrt{a - \sqrt{a+x}}\right)^2 = x^2 \implies a - \sqrt{a+x} = x^2.$$

Rearrange to isolate the inner square root term:

$$(a - x^2)^2 = a + x \implies a^2 - 2ax^2 + (x^2)^2 = a + x \implies x^4 - 2ax^2 - x + (a^2 - a) = 0$$

...

**Wait, wait. Wait. That's an aha moment I can flag here.**

Let's reevaluate this step-by-step to identify if the correct sum can be ...

We started with the equation:

$$\sqrt{a - \sqrt{a+x}} = x$$

First, let's square both sides:

$$a - \sqrt{a+x} = x^2 \implies \sqrt{a+x} = a - x^2$$

Next, I could square both sides again, treating the equation: ...

...



# DeepSeek-R1-Zero 문제점

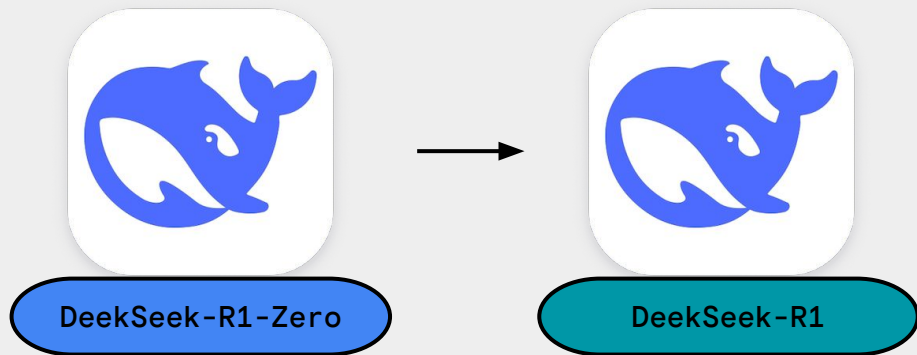
논문에서는 R1-Zero가 순수 강화 학습만으로 학습되어 강력한 추론 능력을 보임

결과물이 읽기 어렵고 여러 언어가 혼합되는 문제 등 실용성 측면에서 한계 발생

# DeepSeek-R1

소량의 고품질 콜드 스타트 데이터를 활용하고 다단계 학습 전략을 적용한 R1 모델을 도입

- 콜드 스타트 데이터는 AI에게 '어떻게 생각해야 하는지'에 대한 기본 틀을 제공
- 이는 AI가 스스로 생각을 정리하고, 문제 해결 방법을 개선



**\*콜드 데이터 (cold data)** : 모델이 본격적 강화 학습을 시작하기 전에 초기 기초 교육용으로 사용하는 소량이지만 고품질 데이터

# DeepSeek-R1 Rejection Sampling

## Rejection Sampling

AI가 문제를 해결할 때 한 번에 여러 가지 답변에서 올바르지 않거나 품질이 낮은 답변은 과감히 제거하고, 좋은 답변만 선택하여 학습에 활용하는 과정

- AI는 한 문제에 대해 여러 답을 생성하는데, 이 과정은 여러 번의 시도와 실수를 포함
- 이 과정에서 '잘못된 답변'이나 '이해하기 어려운 답변'은 걸러내고, 올바른 답변들만 선택합니다.
- 결국 **AI가 “내가 왜 이 방법을 선택해야 하는지”에 대해 더 잘 배우게 하는 역할**을 합니다.



# DeepSeek-R1 SFT

Supervised Fine-Tuning, SFT

Rejection Sampling을 통해 선별된 '좋은' 답변들을 사용하여 AI 모델을 다시 한 번 더 세밀하게 훈련

- 선별된 좋은 답변을 가지고 AI 모델을 추가로 훈련시키면, AI는 그 답변을 모범 사례로 삼아 앞으로 유사한 문제를 풀 때 더 정확하고 명확한 답을 내도록 발전
- 학생이 모범 답안을 반복 학습함으로써 시험에서 더 좋은 성적을 내게 되는 과정과 유사

# Experiments

Benchmark (Metric)		Claude-3.5-Sonnet-1022	GPT-4o-0513	DeepSeek-V3	OpenAI-o1-mini	OpenAI-o1-1217	DeepSeek-R1
Architecture		-	-	MoE	-	-	MoE
# Activated Params		-	-	37B	-	-	37B
# Total Params		-	-	671B	-	-	671B
English	MMLU (Pass@1)	88.3	87.2	88.5	85.2	<b>91.8</b>	90.8
	MMLU-Redux (EM)	88.9	88.0	89.1	86.7	-	<b>92.9</b>
	MMLU-Pro (EM)	78.0	72.6	75.9	80.3	-	<b>84.0</b>
	DROP (3-shot F1)	88.3	83.7	91.6	83.9	90.2	<b>92.2</b>
	IF-Eval (Prompt Strict)	<b>86.5</b>	84.3	86.1	84.8	-	83.3
	GPQA Diamond (Pass@1)	65.0	49.9	59.1	60.0	<b>75.7</b>	71.5
	SimpleQA (Correct)	28.4	38.2	24.9	7.0	<b>47.0</b>	30.1
	FRAMES (Acc.)	72.5	80.5	73.3	76.9	-	<b>82.5</b>
	AlpacaEval2.0 (LC-winrate)	52.0	51.1	70.0	57.8	-	<b>87.6</b>
	ArenaHard (GPT-4-1106)	85.2	80.4	85.5	92.0	-	<b>92.3</b>
Code	LiveCodeBench (Pass@1-COT)	38.9	32.9	36.2	53.8	63.4	<b>65.9</b>
	Codeforces (Percentile)	20.3	23.6	58.7	93.4	<b>96.6</b>	96.3
	Codeforces (Rating)	717	759	1134	1820	<b>2061</b>	2029
	SWE Verified (Resolved)	<b>50.8</b>	38.8	42.0	41.6	48.9	49.2
	Aider-Polyglot (Acc.)	45.3	16.0	49.6	32.9	<b>61.7</b>	53.3
Math	AIME 2024 (Pass@1)	16.0	9.3	39.2	63.6	79.2	<b>79.8</b>
	MATH-500 (Pass@1)	78.3	74.6	90.2	90.0	96.4	<b>97.3</b>
	CNMO 2024 (Pass@1)	13.1	10.8	43.2	67.6	-	<b>78.8</b>
Chinese	CLUEWSC (EM)	85.4	87.9	90.9	89.9	-	<b>92.8</b>
	C-Eval (EM)	76.7	76.0	86.5	68.9	-	<b>91.8</b>
	C-SimpleQA (Correct)	55.4	58.7	<b>68.0</b>	40.3	-	63.7

Table 4 | Comparison between DeepSeek-R1 and other representative models.



# Conclusion

DeepSeek-R1은 강화 학습과 콜드 스타트 및 여러 학습 전략을 결합하여 AI가 스스로 사고하고 문제를 해결할 수 있도록 만든 모델

DeepSeek-R1은 GPT와 유사한 수준의 추론 능력