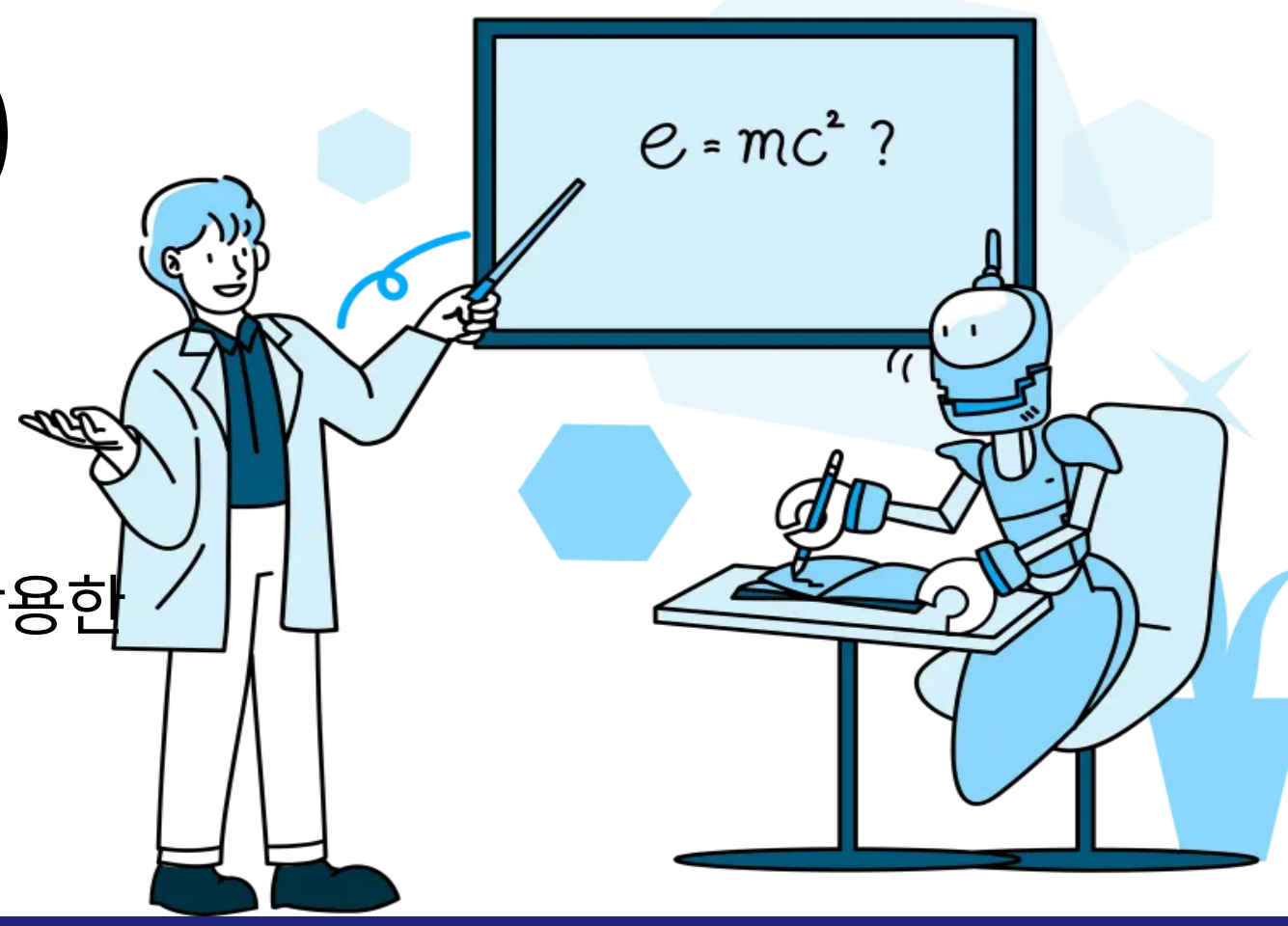


# Alpha Zero 바이트 오목

AlphaZero 스타일 자기대국(Self-Play)를 활용한  
11×11 오목 인공지능 학습 시스템



[https://github.com/minu2001/AlphaZero\\_Gomoku\\_Pygame](https://github.com/minu2001/AlphaZero_Gomoku_Pygame)



[https://www.youtube.com/watch?v=ge36JB\\_28YQ](https://www.youtube.com/watch?v=ge36JB_28YQ)

## 프로젝트 개요

오목(Gomoku)은 규칙은 단순하지만  
장기 전략과 국소 패턴 인식이 동시에 요구되는 보드게임이다.  
본 연구는 인간 기보 없이,  
자기대국(Self-Play)만으로 오목 AI 학습이 가능한지 검증한다.



## 방법론

AlphaZero 스타일 학습 구조

Self-Play



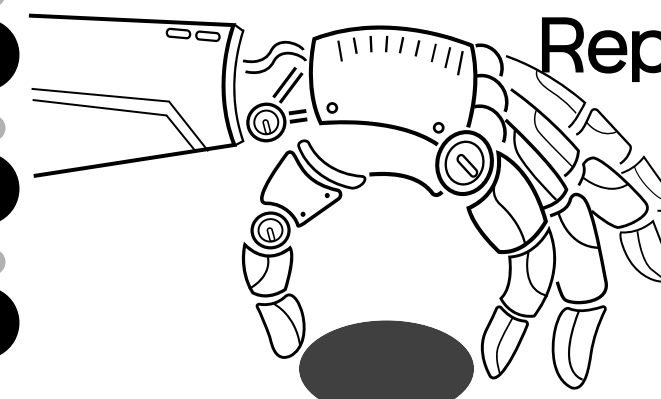
Monte Carlo Tree Search (400 simulations)



정책-가치 신경망(Policy-Value Network)



Replay Buffer 기반 학습



## 학습 규모 및 설정

항목	내용
보드 크기	11 × 11
입력	5채널 상태 표현
신경망	Residual CNN (10 blocks)
출력	정책(121차원) + 가치(-1~1)

항목	값
MCTS 시뮬레이션	400 / 수
자기대국	100판 / 반복
리플레이 버퍼	30,000
옵티마이저	Adam (lr=0.0002)
총 학습 시간	13일 이상

## 모델 구성

입력: 11×11×5 상태 텐서

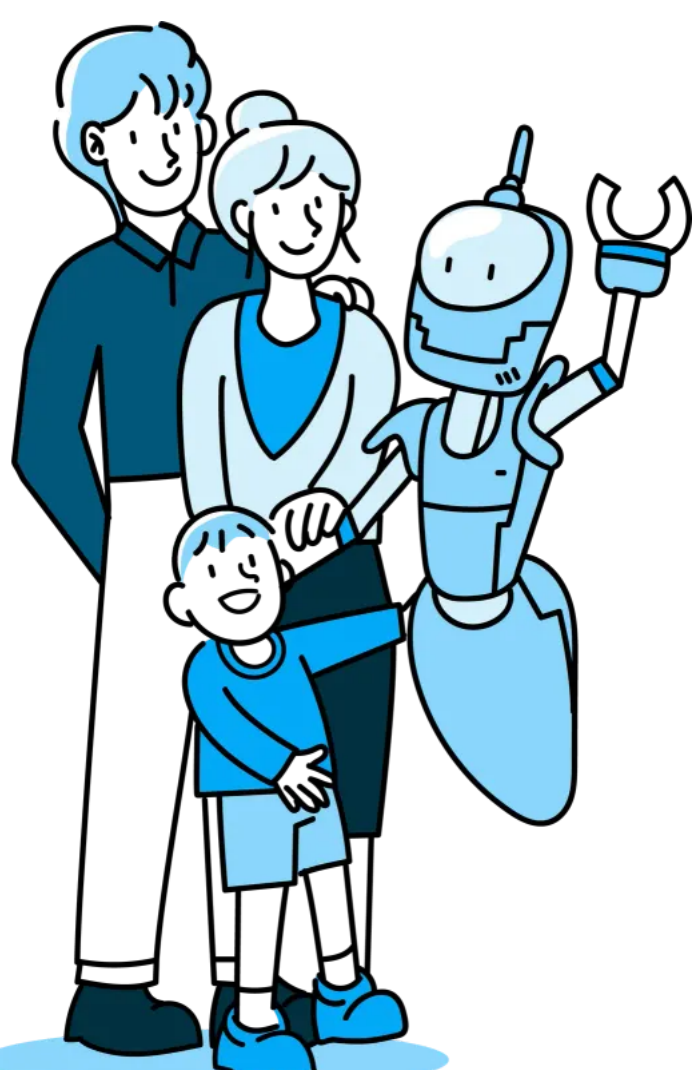
구조: Residual Block 기반 CNN

출력: Policy(수 분포) + Value(승률 예측)

손실함수: Value MSE + Policy Cross Entropy

Self-Play → MCTS (400 sims) → Policy-Value Net → Training

## 의의 및 기대 효과

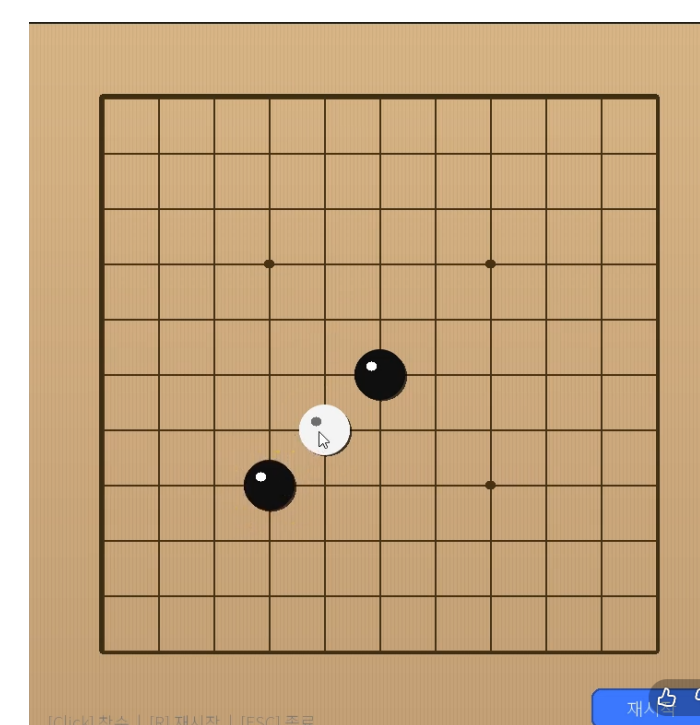


강화학습을 통해 인간 수준 전략  
습득 가능성 실험

AI의 의사결정 과정을 시각적으로  
이해 가능

교육, 전시, 연구 등 확장성 높은  
프로젝트

## 시연



```
Policy:
[[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. ]
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. ]
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. ]
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. ]
[0. 0. 0. 0.01 0.42 0.01 0.13 0. 0. 0. 0. ]
[0. 0. 0. 0.03 0.02 0. 0.01 0. 0. 0. 0. ]
[0. 0. 0.01 0.05 0. 0.03 0.2 0.01 0. 0. 0. ]
[0. 0. 0. 0. 0.01 0.03 0. 0. 0. 0. 0. ]
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. ]
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. ]
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. ]]

White's win!: 65.54%
-----
AI가 수를 결정했습니다.
```

오목이 확률이 가장 높은 곳을 계산하고 그곳에  
수를 놓는 모습을 볼 수 있다.