1. **DQN (Deep Q-Network)**: 회귀 문제(Regression)
   * 동작 원리: DQN은 Q-learning을 심층 신경망에 적용한 알고리즘. 가치 신경망을 사용하여 주어진 상태에서의 각 행동에 대한 가치(Q-value)를 예측하여 최적의 정책을 학습함. 각 행동마다 기대 손익을 수치적으로 예측함
   * 유리한 데이터: DQN은 각 상태와 행동에 대한 가치를 추정하는데 사용되기 때문에, 연속된 상태 및 행동 공간에서 잘 동작한다. 예를 들어, 게임이나 로봇 제어와 같은 고차원의 상태 공간에서 유용
2. **PG (Policy Gradient)**: 분류 문제(classification)
   * 동작 원리: 정책 신경망을 사용하여 각 상태에서 행동을 선택할 확률을 학습함, 정책을 파라미터화하고, 이를 통해 확률적으로 행동을 선택하고, 그 행동에 대한 보상을 기반으로 정책을 업데이트. 기대 손익을 예측하는 것이 아닌 단순히 현 상황에서 어떤 행동이 더 좋은지를 판단
   * 유리한 데이터: PG는 대규모의 이산 혹은 연속적인 행동 공간에서 유용함. 특히, 연속적인 행동 공간에서 작동하며, 대규모의 데이터셋에서 효과적으로 학습
3. **AC (Actor-Critic)**: 하이브리드 강화학습 모델
   * 동작 원리: 가치 신경망과 정책 신경망을 둘 다 사용함. Actor는 정책 경사 모델을 사용해 수행할 행동을 선택하고, Critic은 Q-learning을 통해 행동의 가치를 평가. 이 두 부분은 서로를 통해 학습하면서 보상을 최대화하는 정책을 찾는다
   * 유리한 데이터: AC는 대부분의 강화학습 환경에서 효과적입니다. 상대적으로 작은 데이터셋에서도 잘 동작하며, 상대적으로 고차원의 상태 및 행동 공간에서도 잘 작동
4. **A2C (Advantage Actor-Critic)**:
   * 동작 원리: A2C는 AC 알고리즘의 변형으로 더 안정적으로 학습하도록 설계. 가치 함수를 이용하여 정책 신경망을 예측함. A2C는 Advantage(상태-행동 가치에서 상태 가치를 뺀 값)를 사용하여 Critic의 예측 오류를 줄이고, Actor의 정책 업데이트를 안정화
   * 유리한 데이터: A2C는 AC와 유사하지만, 보다 안정적으로 학습됨. 이는 대규모의 데이터셋에서 더 빠르고 안정적으로 수렴 가능함
5. **A3C (Asynchronous Advantage Actor-Critic)**: 비동기적 개념 추가
   * 동작 원리: A3C는 여러 에이전트가 비동기적으로 환경과 상호작용하여 학습을 진행. 이는 학습 속도를 높이고 데이터의 재사용을 통해 보다 효율적인 학습을 가능하게 함.
   * 유리한 데이터: A3C는 대규모의 환경에서 특히 유용함. 분산 학습을 통해 병렬로 학습하므로, 대규모의 데이터셋에서 효과적으로 학습됨.
6. **PPO (Proximal Policy Optimization)**: 정책 신경망을 업데이트할 때 정책 변화의 크기를 제한하여 학습 안정성을 높인다.
7. **Monkey**: PPO 알고리즘의 변형으로, 분산 환경에서 더 효율적으로 학습하도록 설계됨