상태 -> (레시피, 제외재료1, 제외재료2…)

행동 -> 제외재료를 대체

정책 -> 1.재료간의 food pairing 점수

2.사용자의 미각 파라미터

3.재료의 선호도(사용빈도)

가치함수 -> 벨만 방정식 사용

보상 -> 유저의 직접 피드백(미각 파라미터)와 간접 피드백(재료 사용 빈도) 사용

이후 정책에 반영