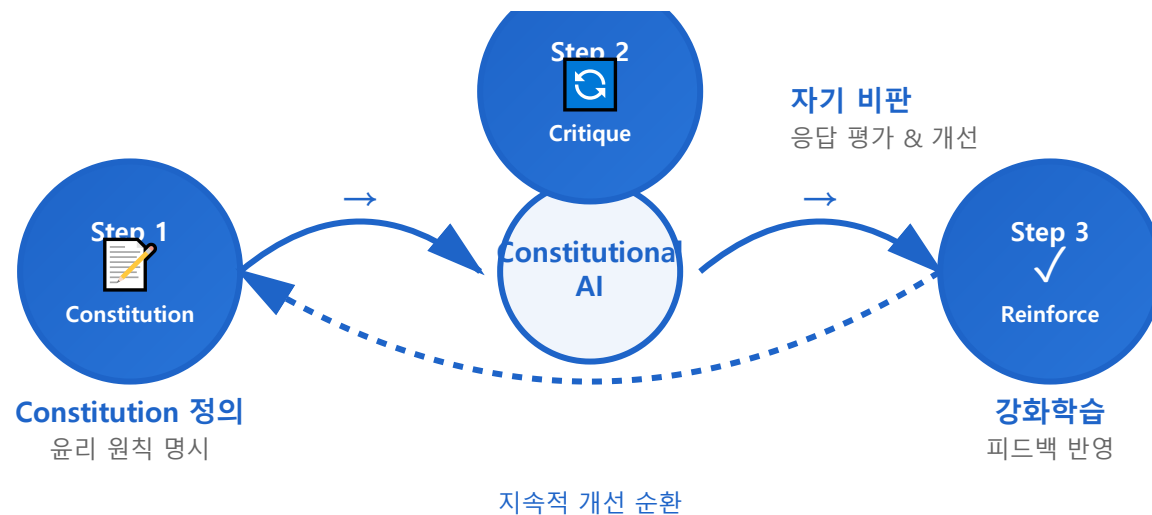


Constitutional AI (CAI) Overview

Constitutional AI는 명시적인 윤리 원칙(constitution)을 기반으로 **자기 개선(self-improvement)**하는 AI 시스템



자기 개선

스스로 응답을 평가하고 개선



윤리 정렬

명시적 원칙에 맞춰 정렬



피드백 루프

지속적인 학습과 개선