

# Evolution of Alignment Methods (2022-2025)

**RLHF**  
(InstructGPT)

2022

**Constitutional AI**  
(Anthropic)

**RLAIF**  
(AI Feedback)

2023

**DPO**  
(Stanford)

**IPO, KTO**  
(Variants)

2024

**Process Supervision**  
(Step-level)

**GRPO**  
(DeepSeek)

2025

*Trend: Simpler, More Scalable, Less Human Labeling*