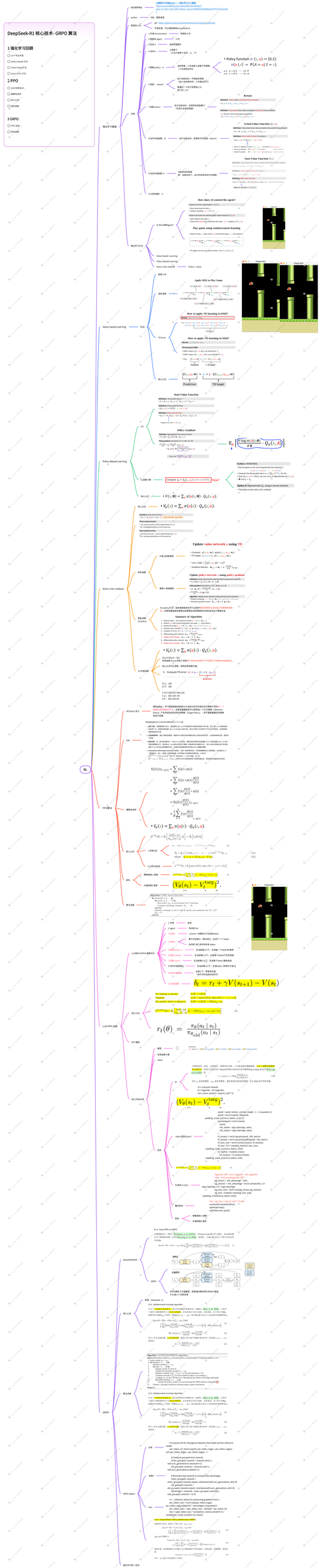
DeepSeek-R1 核心技术- GRPO 算法

1 强化学习回顾

- 1.0 个专业术语
- policy-based 方法
- value-based方法
- actor-critic 方法

2 PPO

- GAE(优势估计)
- 重要性采样
- 核心公式
- 优势函数

3 GRPO

- PPO 改进
- 优势函数