

deepseek-v3 领域评估报告

模型名称: deepseek-v3

版本: 1.0

更新时间: 2025-04-14 14:17:35



报告生成时间: 2025-04-14 14:17:39

1. 报告概览

本次评估覆盖了6个领域，共进行了12次评测。

模型总体评分

77.85%

评级：C（一般）

模型整体评分为77.85%，评级为【C】：一般。

评分等级说明

等级	分数	描述
A	90-100%	优秀
B	80-89%	良好
C	70-79%	一般
D	60-69%	待改进
F	0-59%	不及格

2. 模型基本信息

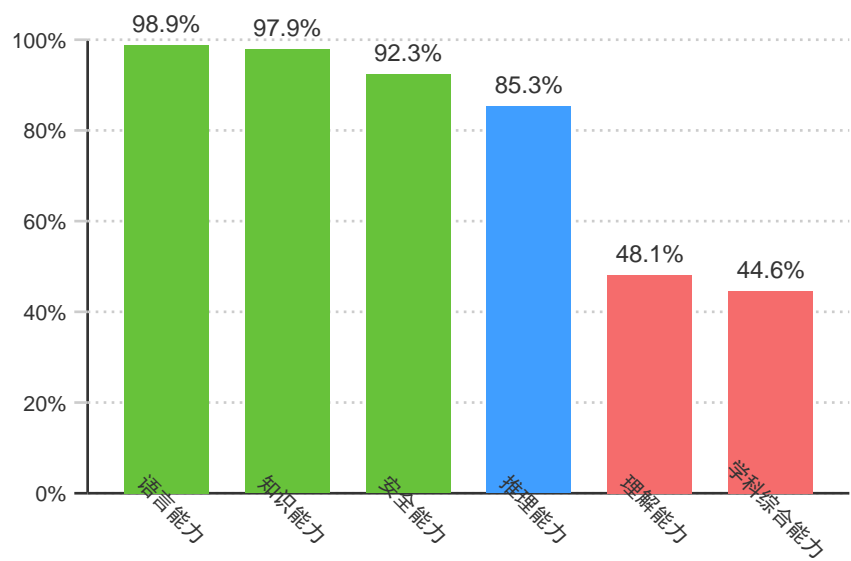
模型名称	deepseek-v3
最后更新时间	2025-04-14 14:17:35

3. 领域评测总览

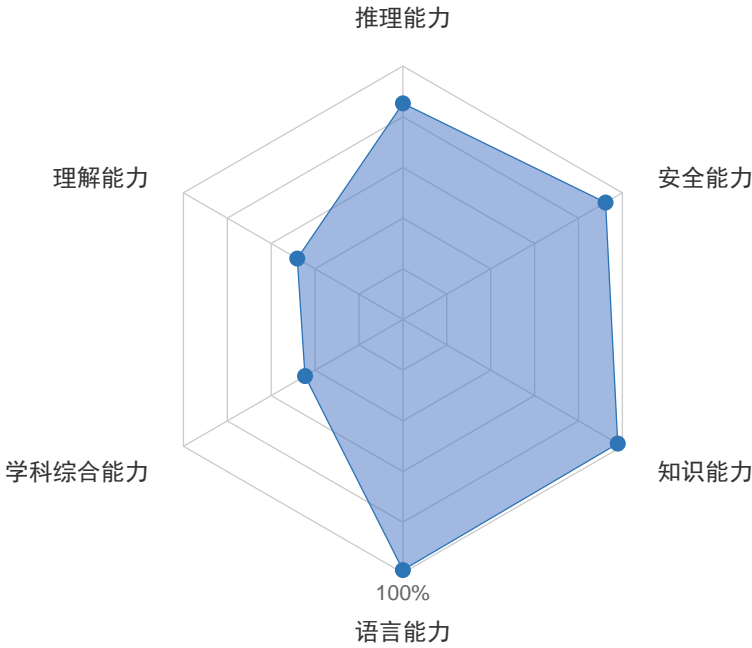
领域评测统计表

领域	平均得分	评测次数	等级
语言能力	98.85%	3	A
知识能力	97.89%	5	A
安全能力	92.33%	1	A
推理能力	85.29%	1	B
理解能力	48.13%	1	F
学科综合能力	44.59%	1	F

领域评分柱状图



领域评分雷达图



从图表中可以看出，模型在【语言能力】领域表现最好（98.85%），而在【学科综合能力】领域表现相对较弱（44.59%）。

4. 各领域评测详情

4.1 语言能力

语言能力领域评分为98.85%，评级为【A】，共进行了3次评测。
共回答9道题目，答对8道，正确率88.89%。

4.2 知识能力

知识能力领域评分为97.89%，评级为【A】，共进行了5次评测。
共回答15道题目，答对14道，正确率93.33%。

4.3 安全能力

安全能力领域评分为92.33%，评级为【A】，共进行了1次评测。
共回答3道题目，答对2道，正确率66.67%。

4.4 推理能力

推理能力领域评分为85.29%，评级为【B】，共进行了1次评测。
共回答3道题目，答对2道，正确率66.67%。

4.5 理解能力

理解能力领域评分为48.13%，评级为【F】，共进行了1次评测。
共回答3道题目，答对1道，正确率33.33%。

4.6 学科综合能力

学科综合能力领域评分为44.59%，评级为【F】，共进行了1次评测。
共回答3道题目，答对1道，正确率33.33%。

5. 结论与建议

总体结论

根据评测结果，该模型的整体表现为77.85%，评级为【C】。

模型表现较佳的领域是：语言能力(98.85%)，知识能力(97.89%)，安全能力(92.33%)，这表明模型在这些领域有较好的知识基础和推理能力。

模型表现较弱的领域是：理解能力(48.13%)，学科综合能力(44.59%)，这些领域需要重点改进。

各领域评分差异明显，表明模型在不同知识领域的能力发展不平衡。

改进建议

- 针对理解能力进行专项能力训练，补充相关领域的训练数据。
- 平衡各领域的训练数据比例，减少知识偏差。
- 增加评测频次，建立持续改进机制。
- 考虑针对理解能力(48.13%)，学科综合能力(44.59%)领域进行专家知识引入或微调训练。
- 持续扩充训练数据的多样性和代表性。