

思考题

- 大模型自动评价还有哪些局限性？
- 如何缓解大模型自动评价中的局限性？
- 能否把典型的自动评价指标和大模型自动评价相结合？