

- Higher-Order Thinking & Skill Development

Personalization, Adaptation & Learning Support
- Instruction Following & Task Completion

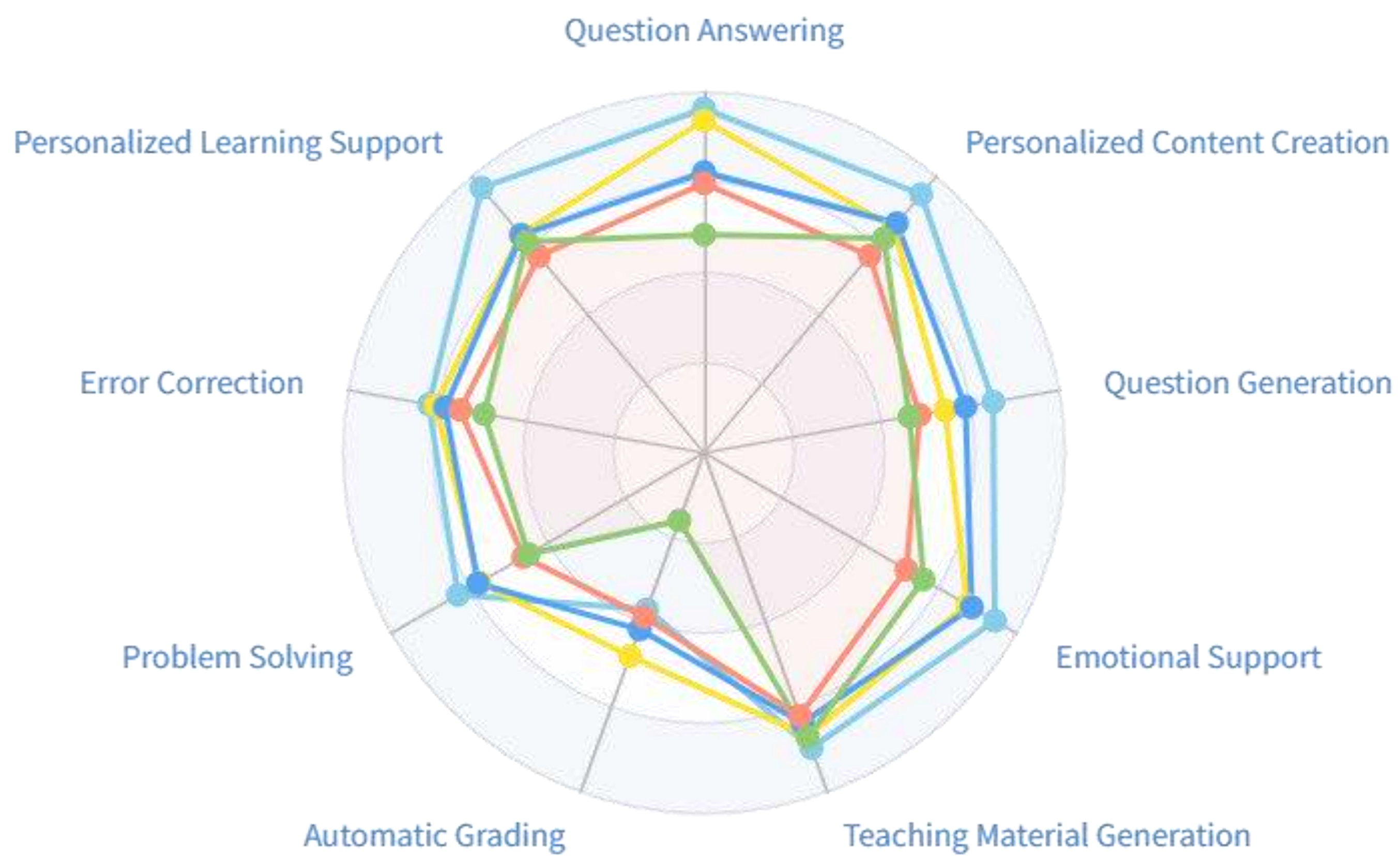
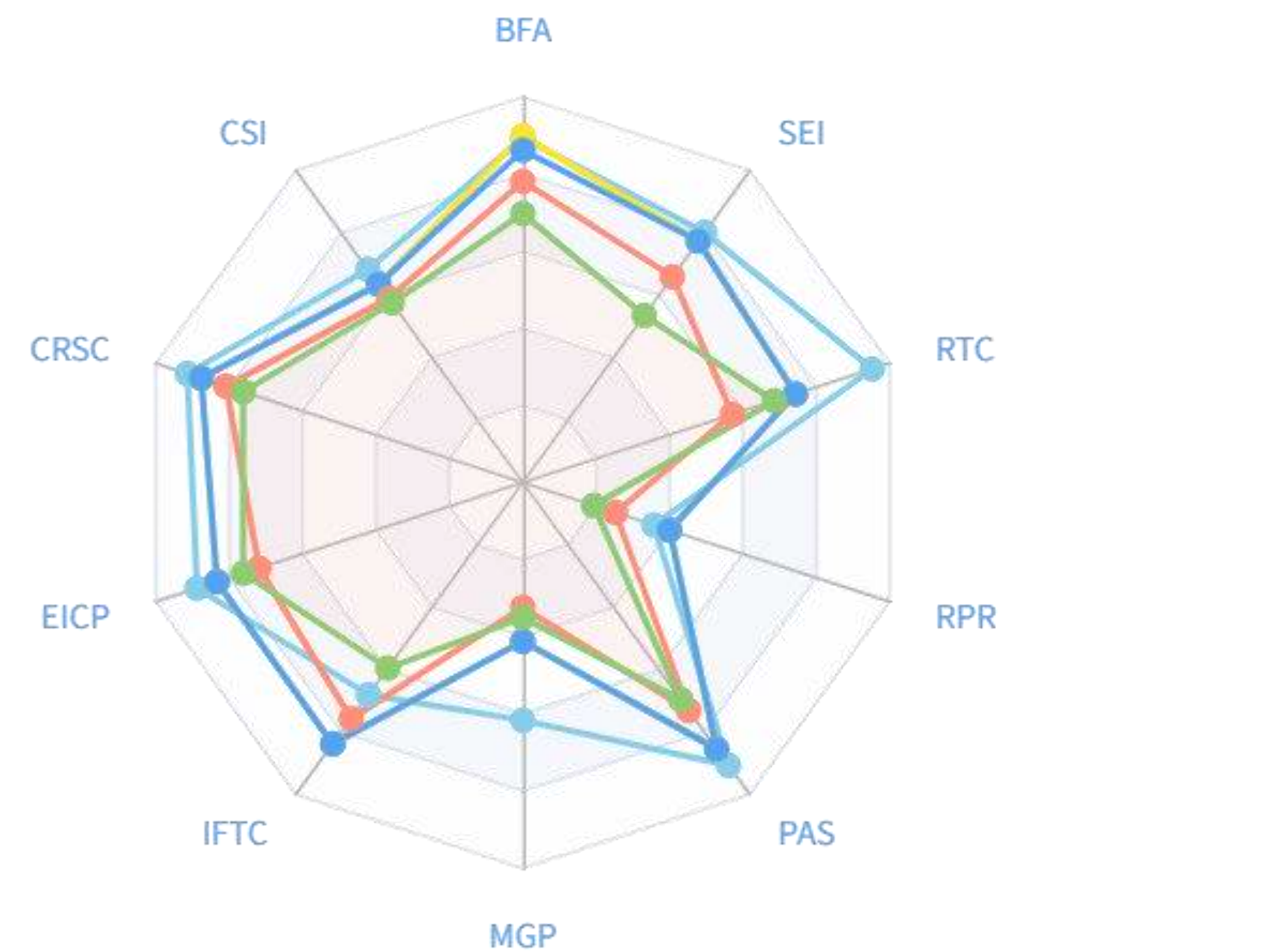
Role & Tone Consistency
- Basic Factual Accuracy

Domain Knowledge Accuracy
- Motivation, Guidance & Positive Feedback

Clarity, Simplicity & Inspiration
- Scenario Element Integration

Content Relevance & Scope Control
- Reasoning Process Rigor

Error Identification & Correction Precision



- DeepSeek R1

DeepSeek V3

Qwen Max
- Qwen2.5-14B-Instruct

Qwen2.5-7B-Instruct