01

Our own work

Do the formal objects match what we meant?

02

Landmark papers

Can the method handle important shared benchmarks?

03

Open problems

What does our community still want to prove?

How far can public LLMs go?