

Figure : Distribution of countries by country category.

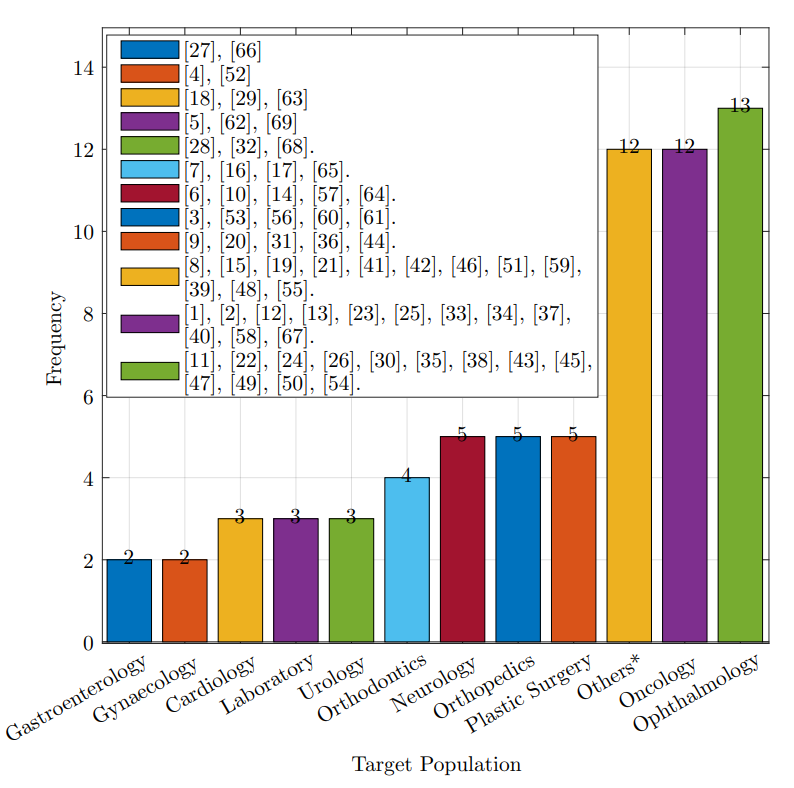


Figure : Distribution of the targeted medical topics. Others\*: BMT, Chronic Diseases, Clinical Imaging, Clinical Procedures, Dermatologic Diseases, Emergency Procedure, End-Of-Life Care, Hepatic Disorder, Pharmacogenetics, Renal Diseases, Respiratory Disorders, and Surgery.

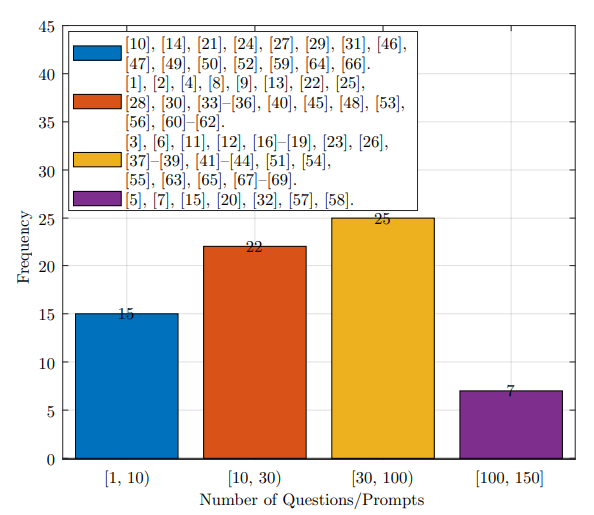


Figure : Distribution of the questions/Prompts count that LLMs in the studies were tested on.

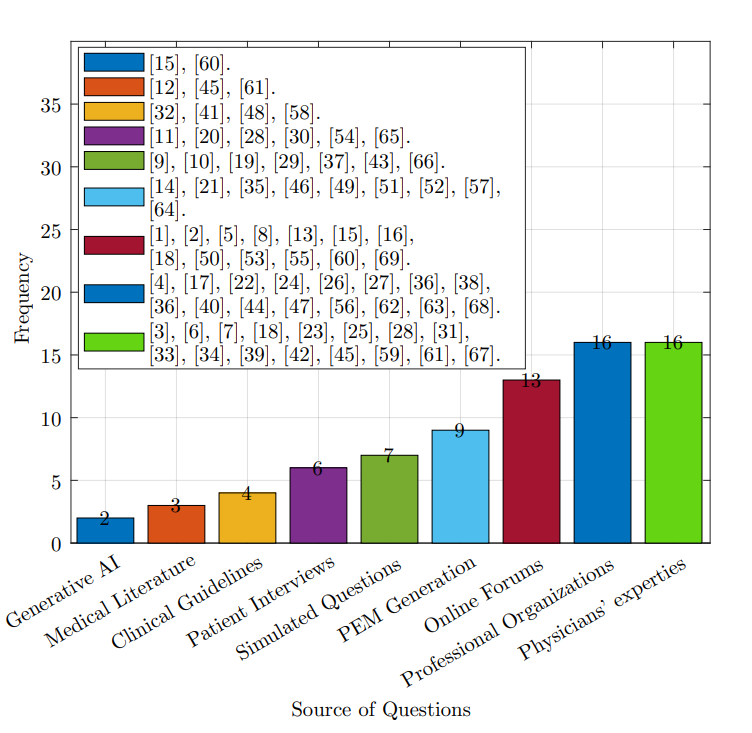


Figure : Distribution of the sources of questions that LLMs in the studies were prompted with.

A graph with numbers and numbers

AI-generated content may be incorrect.

Figure : Readability test distribution. Where GFI is Gunning Fog Index, CLI is Coleman–Liau Index, RO is reviewer Opinion, ARI is Automated Readability Index, LWI is Linsear Write Index, and Others\*: Chinese Readability Analysis Platform, Fry Score, General Standard Chinese Character Table, Raygor Estimate, Syllabus of Graded Words and Characters for Chinese Proficiency, Word Count. Each one listed as others was used only once by the studies [7], [54], [61], [67].

A graph with different colored bars

AI-generated content may be incorrect.

Figure : Distribution of the most accurate LLMs according to the studies.

A graph with colorful bars and text

AI-generated content may be incorrect.

Figure : Distribution of the most readable LLMs according to the studies.