



Supplementary Figure 1. Evaluation Process for State-of-the-Art Online Large Language Models in Responding to Steatotic Liver Disease-Related Queries. This figure illustrates the assessment workflow for evaluating the responses generated by LLMs to 30 steatotic liver disease (SLD)-related queries. Each LLM's responses are represented by a unique color. These responses were randomly ordered and stripped of revealing information to ensure a blind assessment. Subsequently, three physicians independently graded the responses as either "appropriate" or "inappropriate" over five days. The final grade for each response was determined using a majority consensus approach.