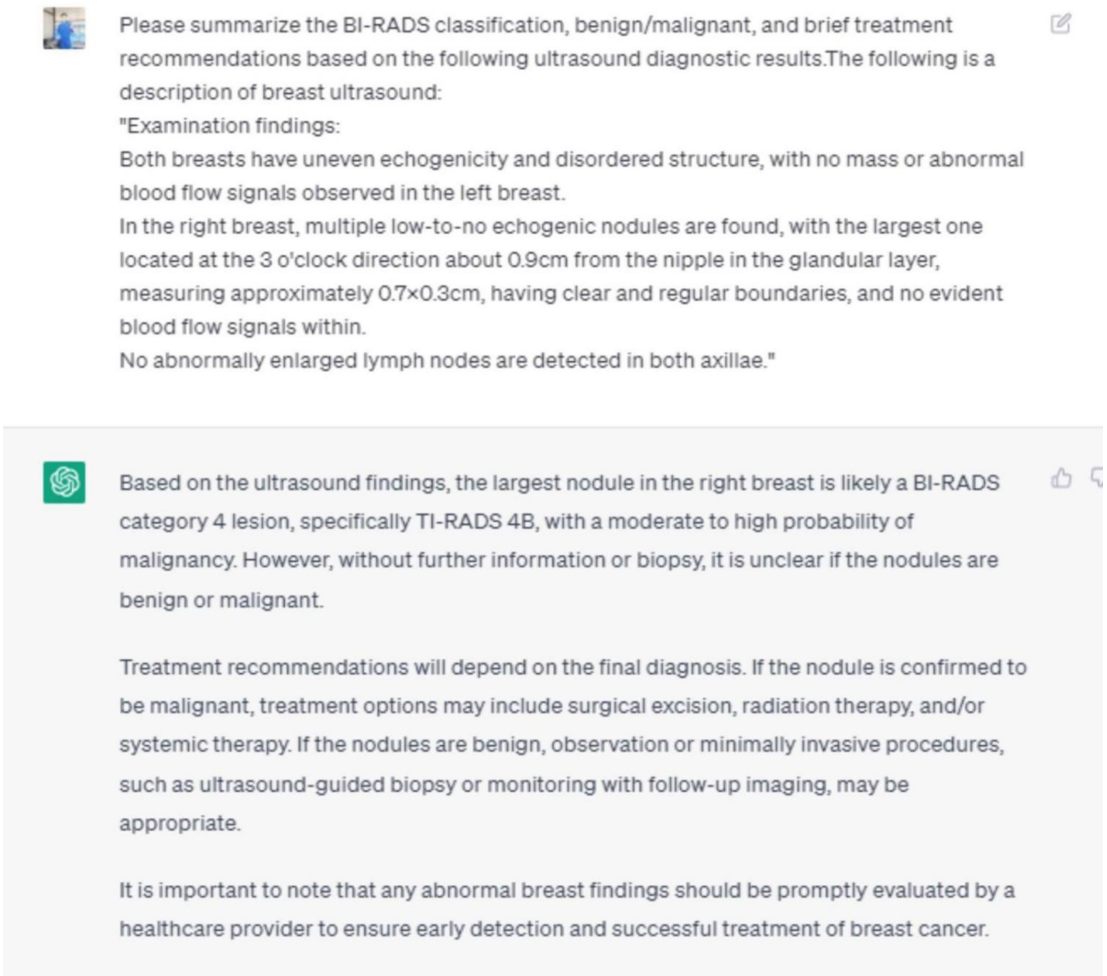

Figure S1 The procedure for posing questions to ChatGPT and gathering its responses. BI-RADS, Breast Imaging Reporting and Data System; TI-RADS, Thyroid Imaging Reporting and Data System.

https://dx.doi.org/10.21037/qims-24-141

Table S1 Evaluation criteria for the performance of ChatGPT in generating breast cancer diagnosis reports

Criterion	Range	Description
Structure and organization (S&O)	1 to 5	Rate the structure of the report
Professional terminology and expression (PTE)	1 to 5	Rate the use of professional terminology in the report
BI-RADS classification	1 to 5	Rate the diagnostic accuracy of the report
Malignancy diagnosis (MD)	1 to 5	Rate the accuracy of malignancy diagnosis
Treatment recommendations (TR)	1 to 5	Rate the appropriateness of treatment recommendations
Clarity and comprehensibility (C&C)	1 to 5	Rate the clarity of expression in the report
Likelihood of being written by a physician (LWBP)	1 to 5	Rate the likelihood that the report was written by a doctor
Ultrasound doctor AI acceptance (UDAIA)	1 to 5	Rate the acceptance of AI-generated report by ultrasound doctors
Overall evaluation (OE)	1 to 5	Provide an overall rating for the entire report

For each criterion, a score of 1 represents the lowest rating (i.e., very poor, completely incorrect), while a score of 5 represents the highest rating (i.e., excellent, completely correct).

Table S2 The score table for assessing the generated report.

Attribute	Value
ID	
Gender (male/female)	
Age (years)	
Reason for consultation	
Description	
Ultrasound conclusion	
Structure and organization	
Professional terminology and expression	
BI-RADS classification	
Malignancy diagnosis	
Treatment recommendations	
Clarity and comprehensibility	
Likelihood of being written by physician	
Ultrasound doctor AI acceptance	
Overall evaluation	
Structure and organization	

BI-RADS, Breast Imaging Reporting and Data System; AI, artificial intelligence.

https://dx.doi.org/10.21037/qims-24-141