Supplementary Appendix

1- All ChatGPT prompts used.

2- Three example cases and GPT responses.

3- Cohen's kappa matrices

1. All ChatGPT Prompts Used

All prompts were standardized. The primary prompt was:

"Based on the following clinical summary, recommend an initial treatment plan for a breast cancer patient: [summary]. Recommendations should include primary treatment (surgery, neoadjuvant chemotherapy, or adjuvant therapy) and any additional recommendations (e.g., clip marking, genetic counselling)."

Additional Prompt Examples:

• **Prompt 1 (Comorbid Patient):** "70-year-old female, Luminal A, stage 2b, Ki-67 10%, postmenopausal, comorbidities: dementia, hypertension. Recommend an initial treatment plan."

• **Prompt 2 (Early Stage):** "45-year-old female, HER2-positive, stage 1, Ki-67 25%, premenopausal, no comorbidities. Recommend an initial treatment plan."

• **Prompt 3 (Advanced Stage):** "60-year-old female, triple-negative, stage 4, Ki-67 40%, postmenopausal, lung metastasis, comorbidity: heart failure. Recommend an initial treatment plan."

Each of the 100 patient summaries included age, gender, histopathological subtype, stage, Ki-67, menopausal status, and comorbidities.

2. Three Example Cases and GPT Responses

Cases reflect cohort diversity (early stage, advanced stage, comorbid). Neither GPT-3.5 nor GPT-4 recommended clip marking, per request.

Case 1: Early Stage, No Comorbidities

Summary: 50-year-old female, Luminal A, stage 2a, Ki-67 15%, premenopausal, no comorbidities.

MDT Decision: Primary surgery (breast-conserving), adjuvant hormonal therapy, clip marking.

GPT-4 Response: Primary surgery (breast-conserving), adjuvant hormonal therapy, genetic counselling.

GPT-3.5 Response: Primary surgery (mastectomy), adjuvant hormonal therapy, genetic counselling.

Comment: GPT-4 agreed on primary treatment but omitted clip marking, suggesting genetic counselling. GPT-3.5's mastectomy was clinically acceptable but more invasive.

Case 2: Advanced Stage, Comorbid

Summary: 65-year-old female, triple-negative, stage 4, Ki-67 35%, postmenopausal, liver metastasis, heart failure.

MDT Decision: Neoadjuvant chemotherapy, genetic counselling.

GPT-4 Response: Neoadjuvant chemotherapy, genetic counselling.

GPT-3.5 Response: Palliative chemotherapy, genetic counselling.

Comment: GPT-4 fully agreed. GPT-3.5's palliative chemotherapy was inappropriate, ignoring heart failure constraints.

Case 3: Comorbid, Moderate Risk

Summary: 72-year-old female, Luminal B, stage 3a, Ki-67 20%, postmenopausal, dementia.

MDT Decision: Neoadjuvant hormonal therapy, genetic counselling.

GPT-4 Response: Neoadjuvant hormonal therapy, genetic counselling.

GPT-3.5 Response: Primary surgery, genetic counselling.

Comment: GPT-4 fully agreed. GPT-3.5's surgical recommendation was unsuitable due to dementia.

3. Cohen's Kappa Matrices and Heatmap

GPT-4 Primary Treatment Matrix (Table 2)			
MDT\GPT-4	**Adjuvant therapy**	**Surgery**	**Neoadjuvant chemotherapy**
Adjuvant therapy	18	1	1
Surgery	1	48	1
Neoadjuvant chemotherapy	1	3	26
Kappa: $\kappa = 0.712$ (95% CI 0·62–0·80, $p<0.001$)			

GPT-3.5 Primary Treatment Matrix (Table 2)			
MDT\GPT-3.5	**Adjuvant therapy**	**Surgery**	**Neoadjuvant chemotherapy**
Adjuvant therapy	17	2	1
Surgery	2	45	3
Neoadjuvant chemotherapy	2	2	26
Kappa: $\kappa = 0.600$ (95% CI 0.50–0.70, $p<0.001$)			
Inter-Rater Reliability			
Agreement between two researchers evaluating ChatGPT responses: $\kappa = 0.92$ (95% CI 0.88–0.96).			