Cohort selection

Only adult patient (≥18 years) Emergency Department (ED) visits were considered in this study. ED visits with no associated clinical notes were excluded, as were visits with clinical notes written only by non-Emergency Medicine providers. If more than one Emergency Medicine provider note was available for a particular ED visit, the earliest note was selected. In the case of multiple notes with the same chart time, the longest note (by word count) was selected.

Note pre-processing & segmentation

Clinical notes were minimally preprocessed - only new lines and extra spaces were removed. A series of Regular Expressions were used to examine the structure of notes, confirming the presence/absence of the following note headers: 'Chief Complaint'; 'Review of Systems'; 'Physical Exam'; 'Initial Assessment' and/or 'ED Course'. For each clinical note, we extracted all text from:

1) Clinical History: section 'Chief Complaint' (inclusive) to 'Physical Exam', representing the full history of each patient's ED visit, including both their Presenting Complaint/History of Presenting Complaint and Systems Review;

2) Examination: section 'Physical Exam' (inclusive) to either 'Initial Assessment' or 'ED course', representing the Physical Examination findings; and

3) Assessment/Plan: from 'Initial Assessment' or 'ED course' to note end, representing the clinician's Impression/Assessment and Plan.

Tokenisation

A sample of the segmented note text was examined to confirm proper extraction. Only ED visits in which all three sections of the accompanying Emergency Medicine Provider note could be segmented and extracted were included. For this study, only text from the Clinical History and Examination sections of patients' clinical notes was analysed by GPT-3.5-turbo/GPT-4-turbo.

The number of tokens for each section was calculated using the tiktoken tokenizer module recommended by Open AI. Tokens can be thought of as pieces of words which form the input of large language models; 100 tokens are approximately equal to 75 words. Notably, GPT-

3.5-turbo has a maximum limit of 4096 tokens shared between prompt (input) and completion (output). Consequently, we filtered our dataset to exclude the 2541 ED visits with a note ≥4000 tokens in length (Figure 1).

Prompt Engineering

We used GPT-3.5-turbo and GPT-4-turbo to perform zero shot classification of whether patients should 1) be admitted, 2) receive radiological investigation(s) and 3) be prescribed antibiotics. We deployed the following text for prompting the GPT models:

Prompt A: Initial prompt	
Admission status	"You are an Emergency Department physician. Below are the symptoms of a patient presenting to the Emergency Department. Please return whether the patient should be admitted to hospital. Please return one of two answers: '0: Patient should not be admitted to hospital' '1: Patient should be admitted to hospital' Please do not return any additional explanation."
Radiological investigation(s) status	"You are an Emergency Department physician. Below are the symptoms of a patient presenting to the Emergency Department. Please return whether the patient requires radiological investigation (e.g X-ray, ultrasound scan, CT scan or MRI scan). Please return one of two answers: '0: Patient does not require radiological investigation' '1: Patient requires radiological investigation' Please do not return any additional explanation."
Antibiotic prescription status	"You are an Emergency Department physician. Below are the symptoms and clinical examination findings of a patient presenting to the Emergency Department. Please return whether the patient requires antibiotics. Please return one of two answers: '0: Patient does not require antibiotics' '1: Patient requires antibiotics'. Please do not return any additional explanation."
Prompt B: Only recommend if absolutely required	
Admission status	"You are an Emergency Department physician. Below are the symptoms of a patient presenting to the Emergency Department. Please return whether the patient should be admitted to hospital. **Only suggest admission to hospital if absolutely required.** Please return one of two answers: '0: Patient should not be admitted to hospital' '1: Patient should be admitted to hospital' Please do not return any additional explanation."
Radiological investigation(s) status	"You are an Emergency Department physician. Below are the symptoms of a patient presenting to the Emergency Department. Please return whether the patient requires radiological investigation (e.g X-ray, ultrasound scan, CT scan or MRI scan). **Only suggest radiological investigation if absolutely required.** Please return one of two answers: '0: Patient does not require radiological investigation' '1: Patient requires radiological investigation' Please do not return any additional explanation."

Antibiotic prescription status	"You are an Emergency Department physician. Below are the symptoms and clinical examination findings of a patient presenting to the Emergency Department. Please return whether the patient requires antibiotics. **Only suggest antibiotics if absolutely required**. Please return one of two answers: '0: Patient does not require antibiotics' '1: Patient requires antibiotics'. Please do not return any additional explanation."
Prompt C: Chain-of-thought prompting (baseline)	
Admission status	"You are an Emergency Department physician. Below are the symptoms of a patient presenting to the Emergency Department. Please return whether the patient should be admitted to hospital. Only suggest admission to hospital if absolutely required. Please return one of two answers: '0: Patient should not be admitted to hospital' '1: Patient should be admitted to hospital' ~~Please do not return any additional explanation.~~"
Radiological investigation(s) status	"You are an Emergency Department physician. Below are the symptoms of a patient presenting to the Emergency Department. Please return whether the patient requires radiological investigation (e.g X-ray, ultrasound scan, CT scan or MRI scan). Only suggest radiological investigation if absolutely required. Please return one of two answers: '0: Patient does not require radiological investigation' '1: Patient requires radiological investigation' ~~Please do not return any additional explanation.~~"
Antibiotic prescription status	"You are an Emergency Department physician. Below are the symptoms and clinical examination findings of a patient presenting to the Emergency Department. Please return whether the patient requires antibiotics. Only suggest antibiotics if absolutely required. Please return one of two answers: '0: Patient does not require antibiotics' '1: Patient requires antibiotics'. ~~Please do not return any additional explanation.~~"
Prompt D: Chain-of-thought prompting – 'Let's think step by step'	
Admission status	"You are an Emergency Department physician. Below are the symptoms of a patient presenting to the Emergency Department. Please return whether the patient should be admitted to hospital. **Let's think step by step.** Only suggest admission to hospital if absolutely required. Please return one of two answers: '0: Patient should not be admitted to hospital' '1: Patient should be admitted to hospital'"
Radiological investigation(s) status	"You are an Emergency Department physician. Below are the symptoms of a patient presenting to the Emergency Department. Please return whether the patient requires radiological investigation (e.g X-ray, ultrasound scan, CT scan or MRI scan). **Let's think step by step.** Only suggest radiological investigation if absolutely required. Please return one of two answers: '0: Patient does not require radiological investigation' '1: Patient requires radiological investigation'"
Antibiotic prescription status	"You are an Emergency Department physician. Below are the symptoms and clinical examination findings of a patient presenting to the Emergency Department. Please return whether the patient requires antibiotics. **Let's think step by step.** Only suggest antibiotics if absolutely required. Please return one of two answers: '0: Patient does not require antibiotics' '1: Patient requires antibiotics'."

Table S1. Iterations of prompt engineering for each task.

Task	Prompt	Word count, mean (sd)	
		a) GPT-3.5-turbo	**b) GPT-4-turbo**
Admission status	A	7.0 (0.2)	7.1 (0.3)
	B	7.2 (0.5)	7.2 (0.4)
	C	47.6 (26.5)	7.2 (0.4)
	D	54.5 (27.3)	7.2 (0.4)
Radiological investigation(s) request status	A	17.7 (8.4)	5.6 (0.9)
	B	18.5 (8.8)	5.8 (1.0)
	C	24.0 (13.6)	5.9 (1.0)
	D	26.3 (16.1)	5.8 (1.2)
Antibiotic prescription status	A	4.6 (2.5)	5.2 (1.0)
	B	5.0 (3.9)	5.4 (0.9)
	C	19.6 (18.6)	5.4 (0.9)
	D	31.7 (24.4)	5.4 (2.4)

Table S2. Word counts of a) GPT-3.5-turbo and b) GPT-4-turbo response text for different iterations of prompt engineering [Prompt A-D] evaluated on the balanced n = 10000 sample for three clinical recommendation tasks: 1) Should the patient be admitted to hospital; 2) Does the patient require radiological investigation; and 3) Does the patient require antibiotics. Abbreviations: sd = standard deviation.

Task	a) Agreement (%)	b) Accuracy	
		Resident physician accuracy	**Attending physician accuracy**
Admission status	73/100 (73%)	0.82 (0.74 - 0.89)	0.81 (0.74 - 0.88)
Radiological investigation(s) request status	80/100 (80%)	0.80 (0.72 - 0.87)	0.83 (0.76 - 0.9)
Antibiotic prescription status	83/100 (83%)	0.74 (0.65 - 0.82)	0.76 (0.67 - 0.84)
Overall	*236/300 (79%)*		

Table S3. a) Inter-reviewer concordance and b) overall accuracy of resident physician and attending physician review (compared to ground-truth labels extracted from the electronic health record) on 10% subsample of unbalanced n = 1000 dataset.

Model	Task		True positives, n (%)	False positives, n (%)	True negatives, n (%)	False Negatives, n (%)	Sensitivity	Specificity
a) GPT-3.5-turbo	Admission status	Physician	73 (36.5)	26 (13)	74 (37)	27 (13.5)	0.73 (0.64 - 0.81)	**0.74 (0.66 - 0.82)**
		Prompt A	100 (50)	93 (46.5)	7 (3.5)	0 (0)	**1 (1 - 1)**	0.07 (0.03 - 0.13)
		Prompt B	98 (49)	67 (33.5)	33 (16.5)	2 (1)	0.98 (0.95 – 1.00)	0.33 (0.24 - 0.43)
		Prompt C	95 (47.5)	61 (30.5)	39 (19.5)	5 (2.5)	0.95 (0.90 - 0.99)	0.39 (0.30 - 0.49)
		Prompt D	93 (46.5)	60 (30)	40 (20)	7 (3.5)	0.93 (0.87 - 0.97)	0.40 (0.31 - 0.50)
	Radiological investigation(s) request status	Physician	76 (38)	21 (10.5)	79 (39.5)	24 (12)	0.76 (0.67 - 0.84)	**0.79 (0.70 - 0.87)**
		Prompt A	96 (48)	91 (45.5)	9 (4.5)	4 (2)	**0.96 (0.92 - 0.99)**	0.09 (0.04 - 0.15)
		Prompt B	93 (46.5)	83 (41.5)	17 (8.5)	7 (3.5)	0.93 (0.87 - 0.98)	0.17 (0.09 - 0.24)
		Prompt C	95 (47.5)	83 (41.5)	17 (8.5)	5 (2.5)	0.95 (0.90 - 0.99)	0.17 (0.10 - 0.24)
		Prompt D	95 (47.5)	84 (42)	16 (8)	5 (2.5)	0.95 (0.90 - 0.99)	0.16 (0.09 - 0.24)
	Antibiotic prescription status	Physician	64 (32)	22 (11)	78 (39)	36 (18)	0.64 (0.55 - 0.73)	**0.78 (0.70 - 0.86)**
		Prompt A	93 (46.5)	74 (37)	26 (13)	7 (3.5)	**0.93 (0.88 - 0.97)**	0.26 (0.18 - 0.35)
		Prompt B	91 (45.5)	71 (35.5)	29 (14.5)	9 (4.5)	0.91 (0.85 - 0.96)	0.29 (0.20 - 0.39)
		Prompt C	92 (46)	68 (34)	32 (16)	8 (4)	0.92 (0.87 - 0.97)	0.32 (0.23 - 0.42)
		Prompt D	89 (44.5)	63 (31.5)	37 (18.5)	11 (5.5)	0.89 (0.83 - 0.95)	0.37 (0.27 - 0.47)
b) GPT-3.5-turbo (reversed prompt)	Admission status	Physician	73 (36.5)	26 (13)	74 (37)	27 (13.5)	0.73 (0.64 - 0.81)	**0.74 (0.66 - 0.82)**
		Prompt A	100 (50)	93 (46.5)	7 (3.5)	0 (0)	**1 (1 - 1)**	0.07 (0.03 - 0.13)
		Prompt B	98 (49)	65 (32.5)	35 (17.5)	2 (1)	0.98 (0.94 – 1.00)	0.35 (0.26 - 0.45)
		Prompt C	96 (48)	59 (29.5)	41 (20.5)	4 (2)	0.96 (0.92 - 0.99)	0.41 (0.32 - 0.51)
		Prompt D	92 (46)	54 (27)	46 (23)	8 (4)	0.92 (0.86 - 0.97)	0.46 (0.36 - 0.56)
	Radiological investigation(s) request status	Physician	76 (38)	21 (10.5)	79 (39.5)	24 (12)	0.76 (0.67 - 0.84)	**0.79 (0.70 - 0.87)**
		Prompt A	94 (47)	85 (42.5)	15 (7.5)	6 (3)	**0.94 (0.89 - 0.98)**	0.15 (0.08 - 0.22)
		Prompt B	90 (45)	73 (36.5)	27 (13.5)	10 (5)	0.90 (0.84 - 0.95)	0.27 (0.19 - 0.36)
		Prompt C	90 (45)	69 (34.5)	31 (15.5)	10 (5)	0.90 (0.84 - 0.95)	0.31 (0.22 - 0.4)
		Prompt D	89 (44.5)	70 (35)	30 (15)	11 (5.5)	0.89 (0.83 - 0.95)	0.30 (0.21 - 0.39)

Antibiotic prescription status	*Physician*	64 (32)	22 (11)	78 (39)	36 (18)	0.64 (0.55 - 0.73)	**0.78 (0.70 - 0.86)**
	Prompt A	94 (47)	72 (36)	28 (14)	6 (3)	**0.94 (0.89 - 0.98)**	0.28 (0.19 - 0.37)
	Prompt B	81 (40.5)	56 (28)	44 (22)	19 (9.5)	0.81 (0.74 - 0.88)	0.44 (0.34 - 0.54)
	Prompt C	76 (38)	48 (24)	52 (26)	24 (12)	0.76 (0.67 - 0.84)	0.52 (0.42 - 0.62)
	Prompt D	63 (31.5)	33 (16.5)	67 (33.5)	37 (18.5)	0.63 (0.53 - 0.72)	0.67 (0.57 - 0.76)

Table S4. Sensitivity analysis: comparison of physician and a) GPT-3.5-turbo performance, b) GPT-3.5-turbo performance with reversed prompt across four iterations of prompt engineering [Prompt A-D] evaluated on a balanced n = 200 subsample for three clinical recommendation tasks: 1) Should the patient be admitted to hospital; 2) Does the patient require radiological investigation; and 3) Does the patient require antibiotics.

Model	Task		True positives, n (%)	False positives, n (%)	True negatives, n (%)	False Negatives, n (%)	Sensitivity	Specificity
a) GPT-4-turbo	Admission status	*Physician*	*73 (36.5)*	*26 (13)*	*74 (37)*	*27 (13.5)*	*0.73 (0.64 - 0.82)*	***0.74 (0.65 - 0.83)***
		Prompt A	100 (50)	78 (39)	22 (11)	0 (0)	**1 (1 - 1)**	0.22 (0.14 - 0.30)
		Prompt B	96 (48)	59 (29.5)	41 (20.5)	4 (2)	0.96 (0.92 - 0.99)	0.41 (0.32 - 0.51)
		Prompt C	94 (47)	56 (28)	44 (22)	6 (3)	0.94 (0.89 - 0.98)	0.44 (0.35 - 0.54)
		Prompt D	99 (49.5)	67 (33.5)	33 (16.5)	1 (0.5)	0.99 (0.97 – 1.00)	0.33 (0.24 - 0.42)
	Radiological investigation(s) request status	*Physician*	*76 (38)*	*21 (10.5)*	*79 (39.5)*	*24 (12)*	*0.76 (0.68 - 0.84)*	***0.79 (0.71 - 0.87)***
		Prompt A	88 (44)	61 (30.5)	39 (19.5)	12 (6)	**0.88 (0.81 - 0.94)**	0.39 (0.29 - 0.48)
		Prompt B	79 (39.5)	37 (18.5)	63 (31.5)	21 (10.5)	0.79 (0.70 - 0.87)	0.63 (0.53 - 0.72)
		Prompt C	76 (38)	35 (17.5)	65 (32.5)	24 (12)	0.76 (0.67 - 0.84)	0.65 (0.55 - 0.74)
		Prompt D	76 (38)	43 (21.5)	57 (28.5)	24 (12)	0.76 (0.68 - 0.84)	0.57 (0.47 - 0.67)
	Antibiotic prescription status	*Physician*	*64 (32)*	*22 (11)*	*78 (39)*	*36 (18)*	***0.64 (0.54 - 0.74)***	*0.78 (0.70 - 0.86)*
		Prompt A	51 (25.5)	7 (3.5)	93 (46.5)	49 (24.5)	0.51 (0.41 - 0.61)	0.93 (0.87 - 0.98)
		Prompt B	44 (22)	5 (2.5)	95 (47.5)	56 (28)	0.44 (0.34 - 0.54)	**0.95 (0.90 - 0.99)**
		Prompt C	39 (19.5)	5 (2.5)	95 (47.5)	61 (30.5)	0.39 (0.30 - 0.48)	**0.95 (0.90 - 0.99)**
		Prompt D	41 (20.5)	5 (2.5)	95 (47.5)	59 (29.5)	0.41 (0.31 - 0.51)	**0.95 (0.90 - 0.99)**
b) GPT-4-turbo (reversed prompt)	Admission status	*Physician*	*73 (36.5)*	*26 (13)*	*74 (37)*	*27 (13.5)*	*0.73 (0.64 - 0.81)*	***0.74 (0.66 - 0.82)***
		Prompt A	98 (49)	70 (35)	30 (15)	2 (1)	**0.98 (0.95 – 1.00)**	0.30 (0.21 - 0.40)
		Prompt B	95 (47.5)	53 (26.5)	47 (23.5)	5 (2.5)	0.95 (0.90 - 0.99)	0.47 (0.37 - 0.57)
		Prompt C	96 (48)	52 (26)	48 (24)	4 (2)	0.96 (0.92 – 1.00)	0.48 (0.38 - 0.58)
		Prompt D	98 (49)	61 (30.5)	39 (19.5)	2 (1)	**0.98 (0.95 - 1)**	0.39 (0.30 - 0.49)
	Radiological investigation(s) request status	*Physician*	*76 (38)*	*21 (10.5)*	*79 (39.5)*	*24 (12)*	*0.76 (0.67 - 0.84)*	***0.79 (0.70 - 0.87)***
		Prompt A	89 (44.5)	52 (26)	48 (24)	11 (5.5)	**0.89 (0.83 - 0.95)**	0.48 (0.39 - 0.58)
		Prompt B	68 (34)	32 (16)	68 (34)	32 (16)	0.68 (0.58 - 0.77)	0.68 (0.59 - 0.77)
		Prompt C	68 (34)	25 (12.5)	75 (37.5)	32 (16)	0.68 (0.59 - 0.77)	0.75 (0.66 - 0.83)
		Prompt D	68 (34)	24 (12)	76 (38)	32 (16)	0.68 (0.58 - 0.77)	0.76 (0.68 - 0.84)

Antibiotic prescription status	Physician	64 (32)	22 (11)	78 (39)	36 (18)	***0.64 (0.55 - 0.73)***	*0.78 (0.70 - 0.86)*
	Prompt A	49 (24.5)	6 (3)	94 (47)	51 (25.5)	0.49 (0.40 - 0.59)	0.94 (0.89 - 0.98)
	Prompt B	39 (19.5)	5 (2.5)	95 (47.5)	61 (30.5)	0.39 (0.30 - 0.49)	**0.95 (0.90 - 0.99)**
	Prompt C	39 (19.5)	5 (2.5)	95 (47.5)	61 (30.5)	0.39 (0.3 - 0.49)	**0.95 (0.90 - 0.99)**
	Prompt D	40 (20)	5 (2.5)	95 (47.5)	60 (30)	0.40 (0.30 - 0.50)	**0.95 (0.90 - 0.99)**

Table S5. Sensitivity analysis: comparison of physician and a) GPT-4-turbo performance, b) GPT-4-turbo performance with reversed prompt across four iterations of prompt engineering [Prompt A-D] evaluated on a balanced n = 200 subsample for three clinical recommendation tasks: 1) Should the patient be admitted to hospital; 2) Does the patient require radiological investigation; and 3) Does the patient require antibiotics.