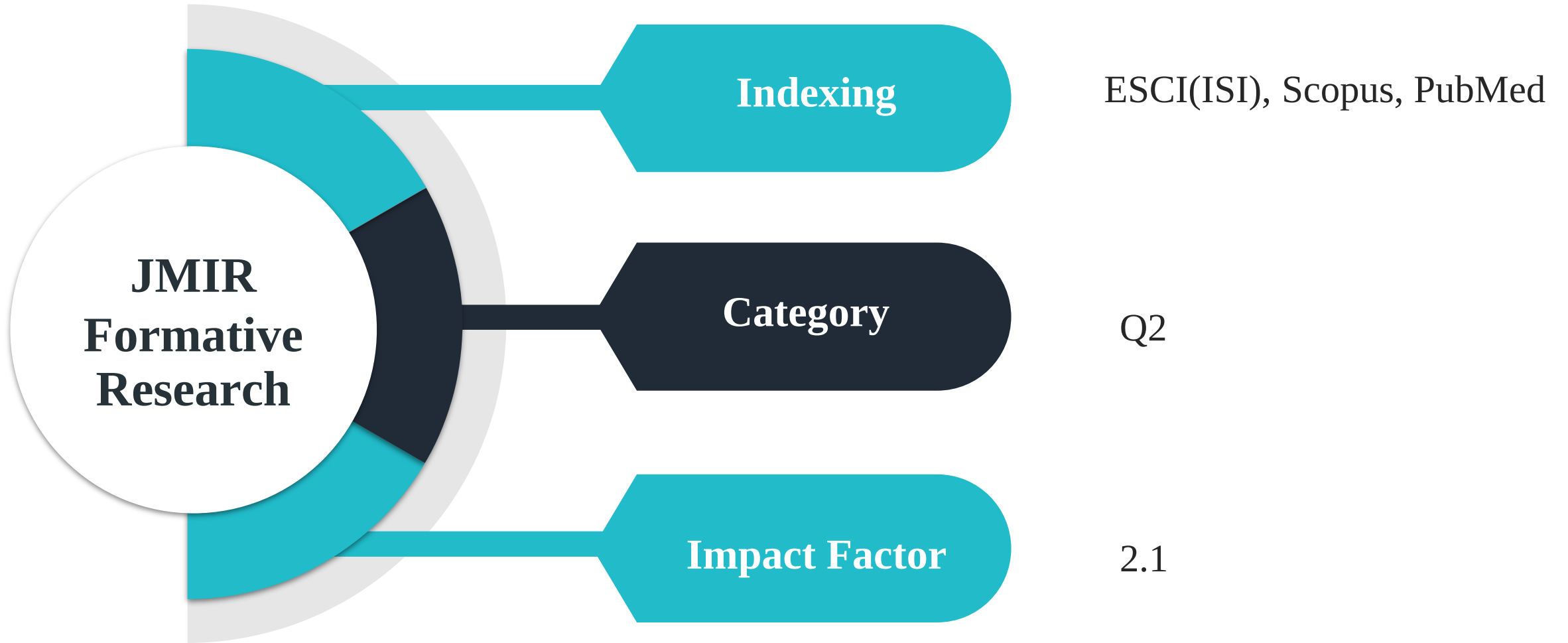




Developing an ICD-10 Coding Assistant: Pilot Study Using RoBERTa and GPT-4 for Term Extraction and Description-Based Code Selection

دانشجو: محمدرضا خیرالسادات
استاد راهنما: آقای دکتر مظاهری

تاریخ ارائه: ۱۲ آبان ۱۴۰۴



فهرست مطالب

مقدمه



۱

هدف پژوهش



۲

روش اجرا



۳

یافته‌ها



۴

نتیجه‌گیری



۵

پیشنهادهای



۶

مقدمه





ICD-10 code: K70.30

lead terms

Description:

K70.30: Alcoholic cirrhosis of liver without ascites

Textual Evidence (Example):

Among his personal history prior to the transplant, notable are an appendectomy in 1977, liver cirrhosis of probable alcoholic origin diagnosed in 1988, upper gastrointestinal bleeding secondary to grade II esophageal varices in 1995 along with portal hypertension and ascites.

Alphabetic Index:

Diagnosis index entries containing back-references to K70.30

Cirrhosis, cirrhotic (hepatic) (liver) K74.60

alcoholic K70.30

Laennec's K70.30

alcoholic K70.30

liver K74.60

alcoholic K70.30

macronodular K74.69

alcoholic K70.30

micronodular K74.69

alcoholic K70.30

nutritional K74.69

alcoholic K70.30

portal K74.69

alcoholic K70.30

postnecrotic K74.69

alcoholic K70.30

Disease, diseased - see also Syndrome

liver (chronic) (organic) K76.9

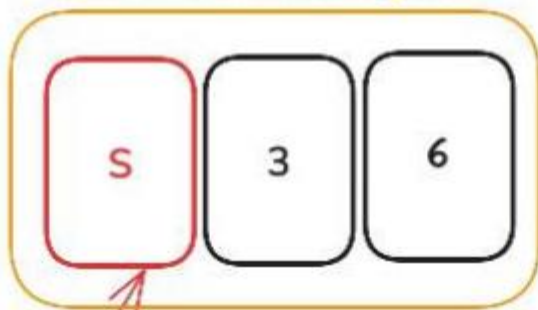
alcoholic (chronic) K70.9

cirrhosis K70.30

Laennec's cirrhosis K70.30



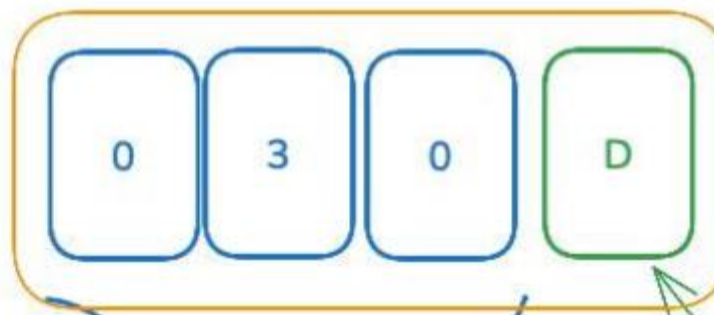
Category
"Injury of
intra-abdominal organs"



Chapter
S00-T98

"Injury, poisoning and certain other consequences
of external causes"

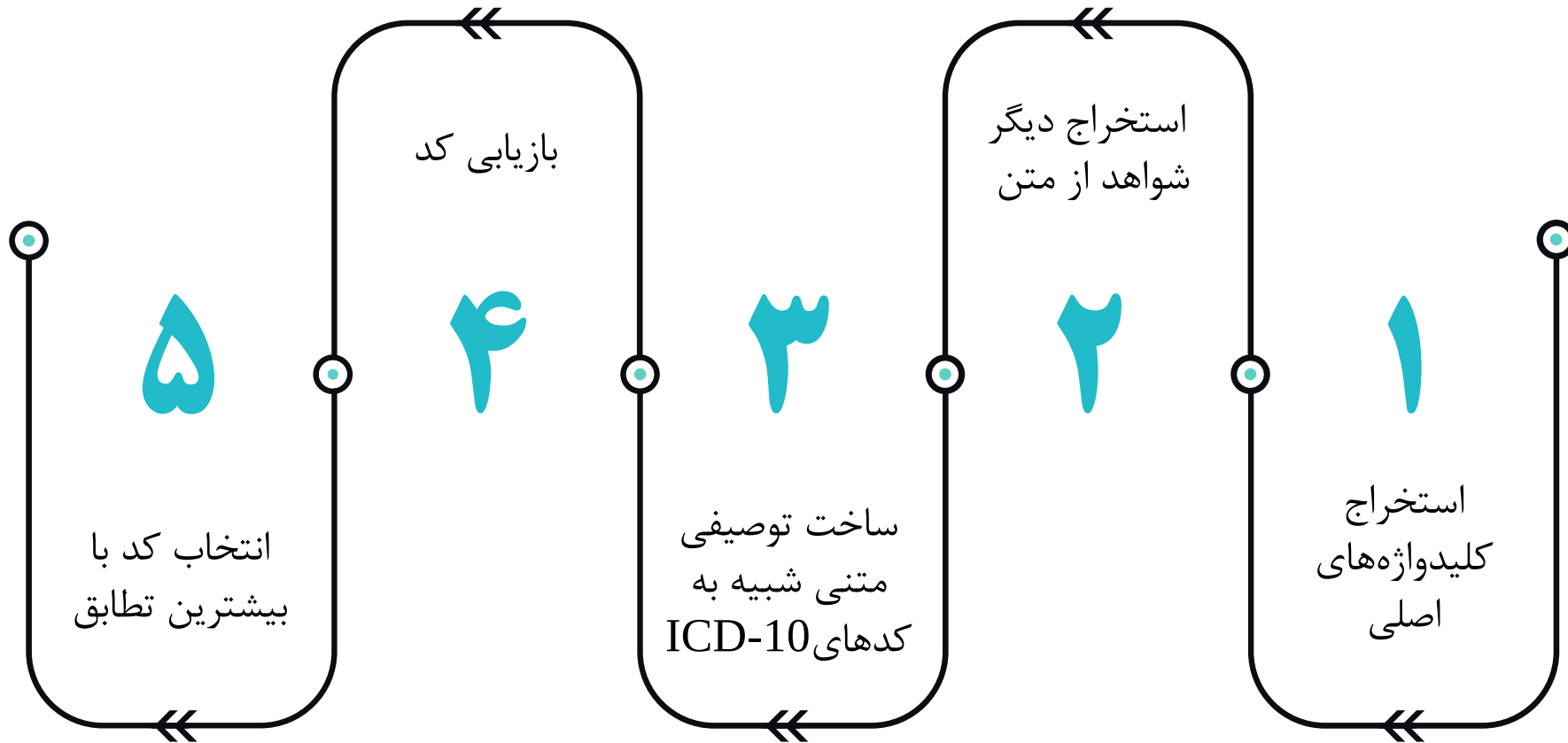
subclassification
"Superficial (capsular) laceration of spleen"



Etiology, anatomic site,
severity, laterality

Extension
"subsequent encounter"

مراحل انجام پژوهش



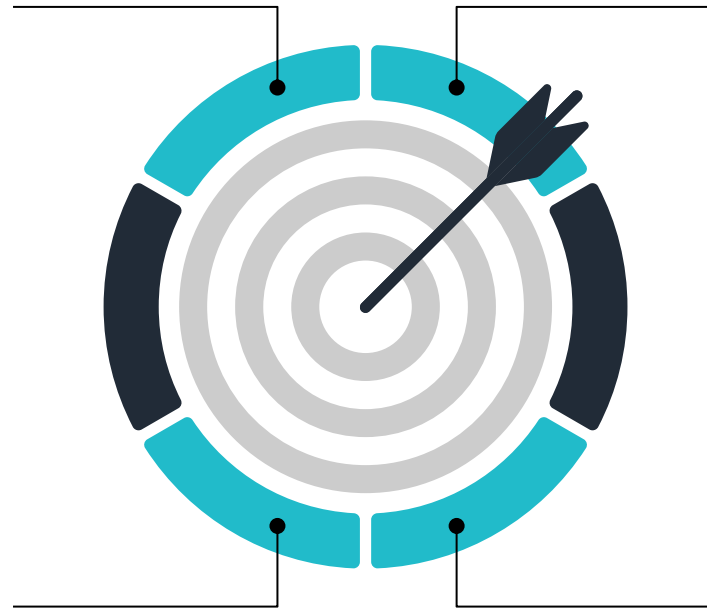
هدف پژوهش



اهداف پژوهش

استخراج خودکار اصطلاحات اصلی از متون
بالینی با Fine-Tune کردن مدل زبانی
RoBERTa

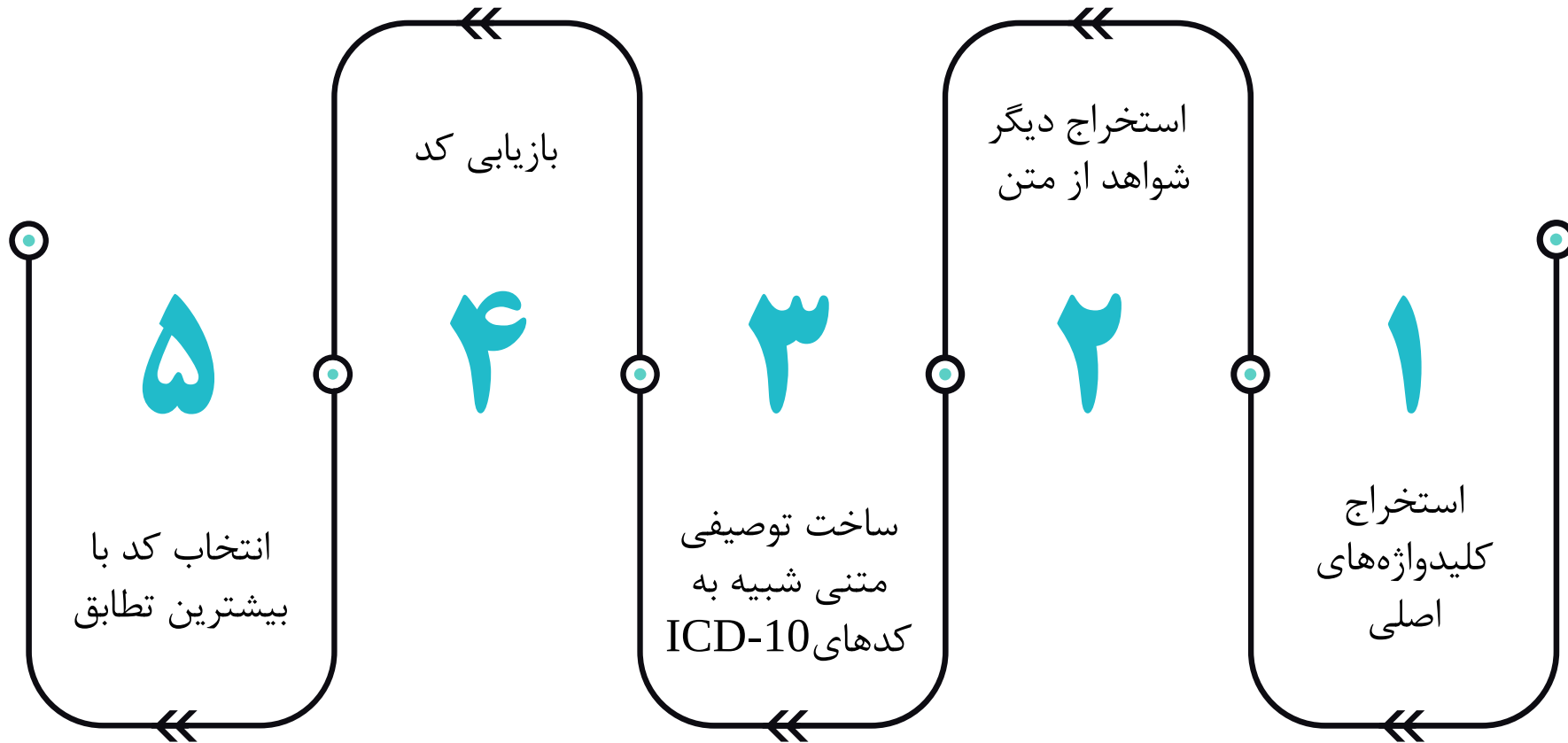
توسعه یک دستیار کدگذاری بر
اساس ICD-10 با تقلید از رفتار
کدگذار انسانی



بررسی قابلیت GPT-4 در تولید
کدهای ICD

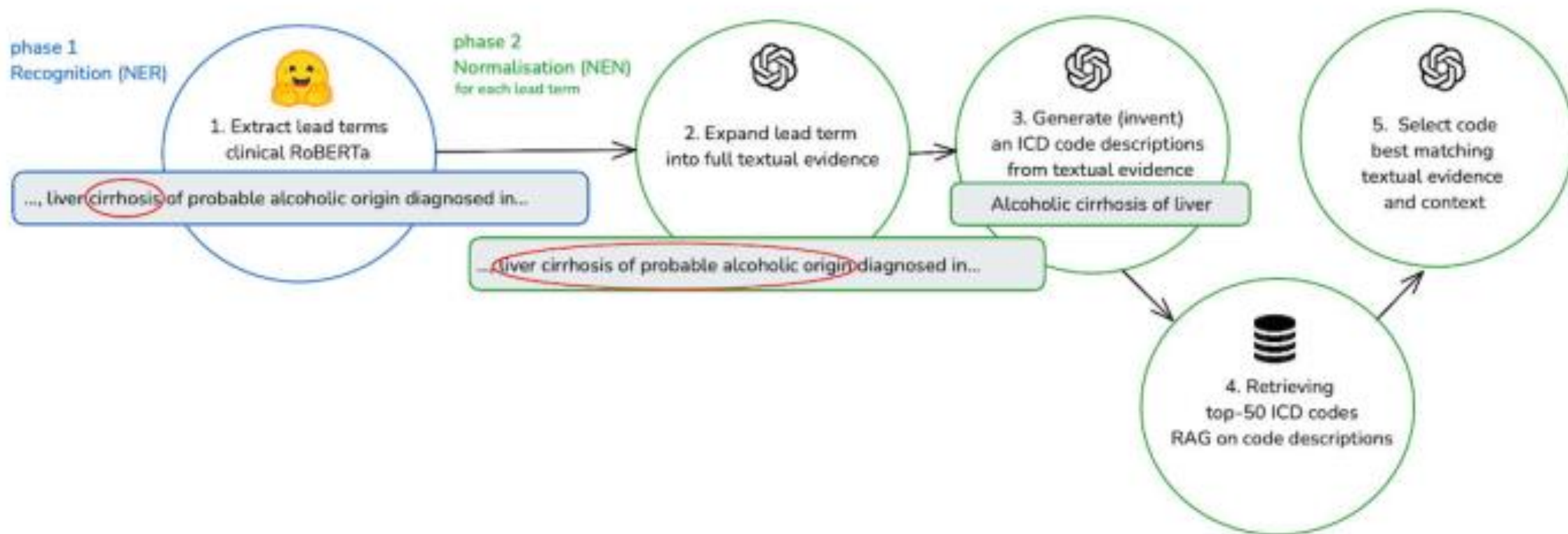
بررسی قابلیت GPT-4 در تولید
توصیف‌های مبتنی بر ICD

مراحل انجام پژوهش



روش اجرا





✓ مرحله ۱: استخراج کلیدواژه‌های اصلی

✓ مدل: RoBERTa Biomedical-clinical برای اسپانیایی

✓ کار: Named Entity Recognition (NER)

مرحله ۲: توسعه شواهد متنی ✓

مدل: GPT 4 ✓

کار: استخراج دیگر شواهد و علائم ✓

مرحله ۳: تولید توصیف کد ✓

مدل: GPT 4 ✓

کار: ایجاد متنی با ساختار کدهای ICD-10 ✓

مرحله ۴: بازیابی کد با RAG ✓

✓ پایگاه داده: Chroma vector

✓ مدل embedding: text-embedding-ada-002

✓ بازیابی ۵۰ کد برتر بر اساس شباهت

مرحله ۵: انتخاب نهایی کد ✓

ورودی: ۵۰ کد بازیابی شده + شواهد متنی ✓

مدل: GPT-4 ✓

کد نهایی انتخاب شده ✓

یافته‌ها





Table 1. Named entity recognition (NER) performance on extracting lead terms and textual evidence for diagnoses and or procedures on CodiEsp-X and CodiEsp-X-lead.

NER	F_1 -score
Textual evidence diagnoses and procedures combined (ours-baseline) [29]	0.65
Textual evidence diagnoses (SOTA ^a) [5]	0.71
Textual evidence procedures (SOTA) [5]	0.61
Main-terms diagnoses (ours) [29]	0.82
Main-terms diagnoses and procedures combined (ours) [29]	0.80

^aSOTA: state-of-the-art.



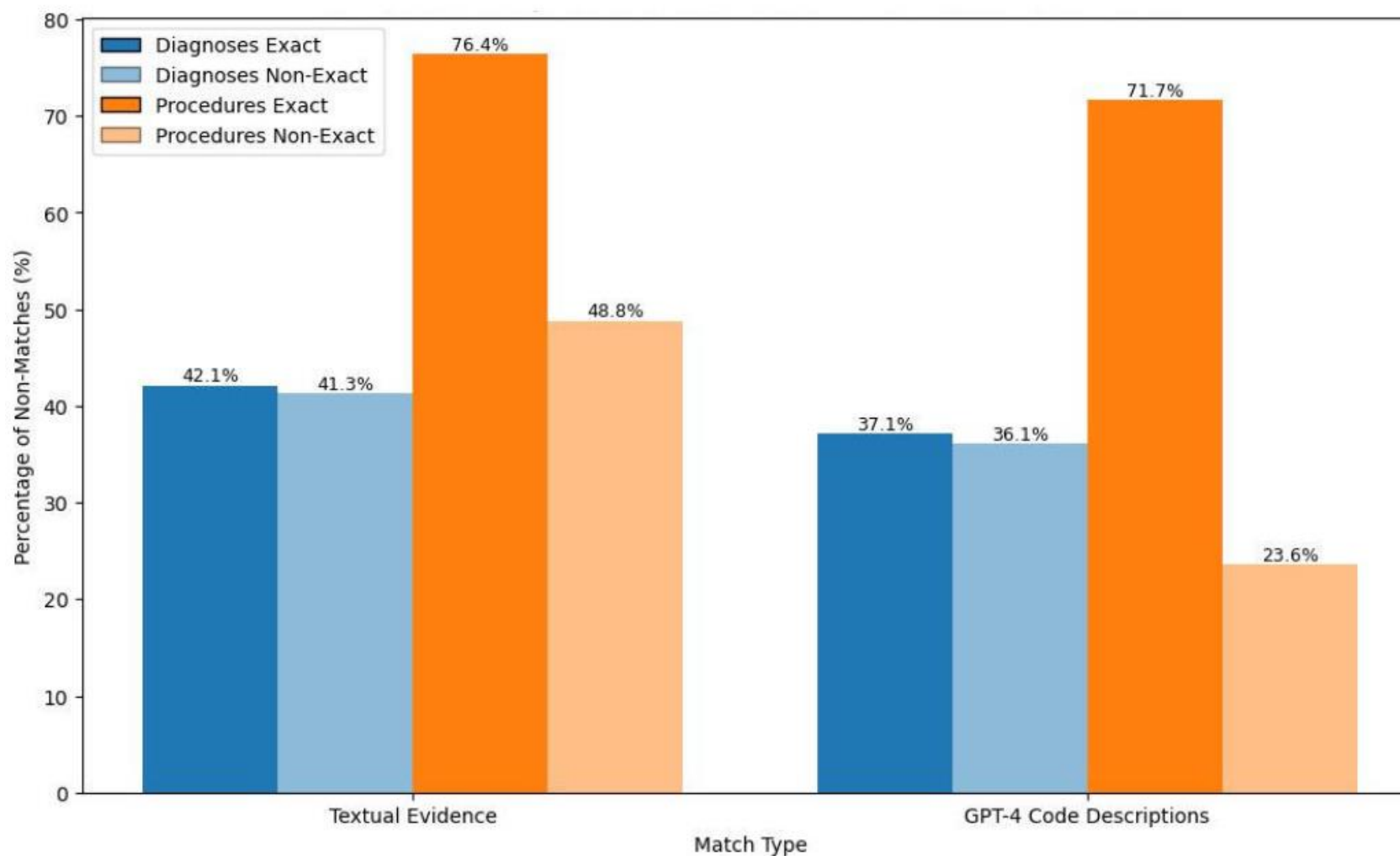
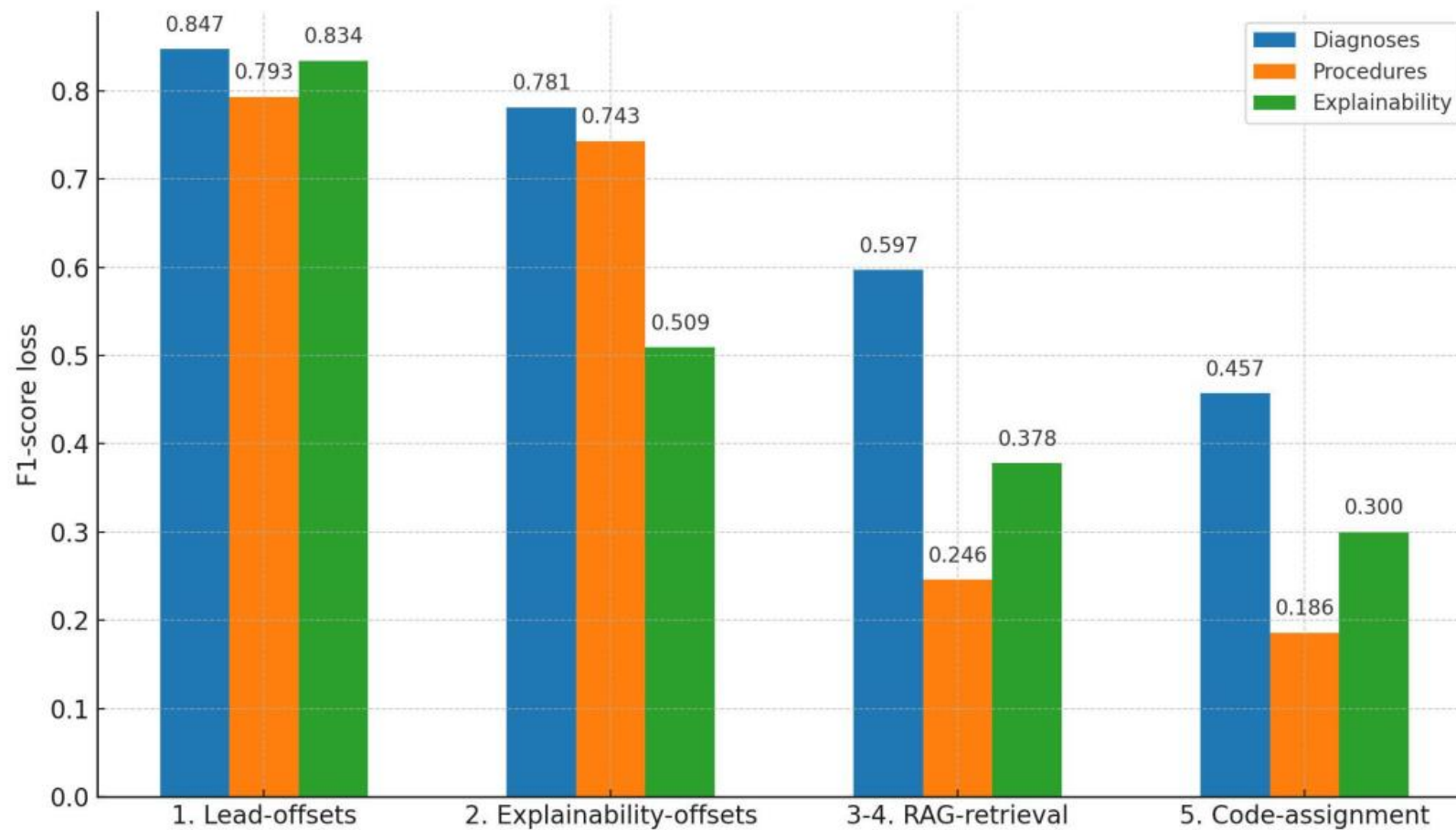




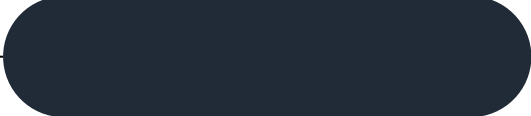
Figure 4. F1-Score performance loss for each individual step in the approach.



نتیجه گیری





- عدم عملکرد مناسب مدل  
- استخراج موفق اصطلاحات اصلی  
- تقلید قابل توجه کار کدگذاران انسانی  
- نیاز به شواهد متنی دقیق قبل از جستجو  

پیشنهادات



توسعه مدل های Neural Network

بررسی عملکرد مدل کلاسیک

بررسی توانایی مدل ها در زبان های بین المللی

سنجش عملکرد مدل ها با الگوریتم عملکردی متفاوت

با سپاس از استادان و حضار گرامی