PHISHING

prompt = (  
 "### SYSTEM ###\n" +  
 system\_text +  
 "\nRespond with exactly ONE word 'phish' or 'legitimate', after ### ANSWER ###. no other tokens." + "\n\n"   
 "### URL ###\n\n" + url\_text +  
 "\n\n ### ANSWER ###\n\n"  
  
)

{

"prompt": "Decide whether the following URL is legitimate or phish.",

"counts": {

"TP": 113,

"FP": 23,

"TN": 102,

תמונה שמכילה טקסט, צילום מסך, תרשים, גופן

תוכן בינה מלאכותית גנרטיבית עשוי להיות שגוי. "FN": 12,

"unknown": 0

},

"precision": 0.8308823529411765,

"recall": 0.904,

"f1": 0.8659003831417624,

"fpr": 0.184,

"num\_samples": 250

},

High FN Low FP

"prompt": "Decide whether the following URL is legitimate or phish. If you detect at least two phishing red flags (e.g., punycode, IP address host, credential keywords), mark it as phish; otherwise treat it as legitimate.",

תמונה שמכילה טקסט, צילום מסך, תרשים, מלבן

תוכן בינה מלאכותית גנרטיבית עשוי להיות שגוי. "counts": {

"TP": 99,

"FP": 9,

"TN": 116,

"FN": 26,

"unknown": 0

},

"precision": 0.9166666666666666,

"recall": 0.792,

"f1": 0.8497854077253219,

"fpr": 0.072,

"num\_samples": 250

},

Low FN High FP

"prompt": "Analyse the URL in four stages\u2014host, path, query, then overall\u2014flagging IP hosts, high-risk TLDs, suspicious keywords, and obfuscated encodings before reaching a verdict of legitimate or phish.",

תמונה שמכילה טקסט, צילום מסך, תרשים, מלבן

תוכן בינה מלאכותית גנרטיבית עשוי להיות שגוי. "counts": {

"TP": 120,

"FP": 31,

"TN": 94,

"FN": 5,

"unknown": 0

},

"precision": 0.7947019867549668,

"recall": 0.96,

"f1": 0.8695652173913043,

"fpr": 0.248,

"num\_samples": 250

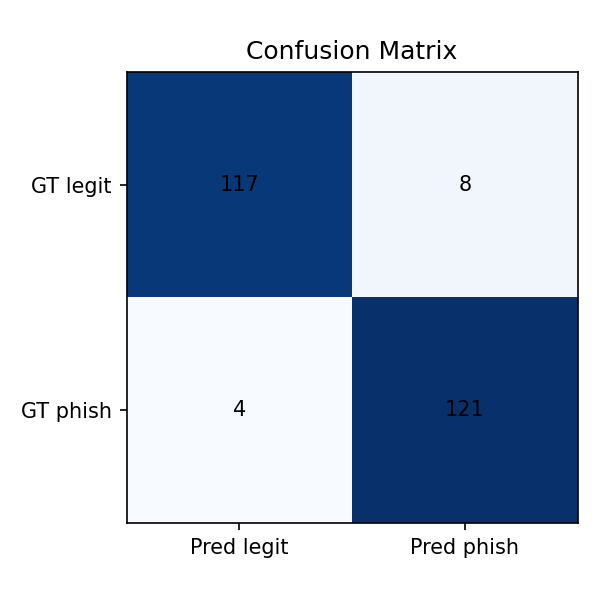
PHISHING - LORA

{

"prompt": "Decide whether the following URL is legitimate or phish.",

"counts": {

"TP": 121,

 "FP": 8,

"TN": 117,

"FN": 4,

"unknown": 0

},

"precision": 0.937984496124031,

"recall": 0.968,

"f1": 0.952755905511811,

"fpr": 0.064,

"num\_samples": 250

}

Future – if prompy engeniring may give approximation of fine tune?