Zero-Shot LLMs vs Encoder-Only Transformers: A Comparative Study on Spam Text Classification

UCI Machine Learning Repository

SMS Spam Collection

Jun 21, 2012 — The SMS Spam Collection is a public set of SMS labeled messages that have been collecte...

5,574 SMS messages labeled as either scam (spam) or legitimate (ham) with about 13.4% of the messages are spam.

Ham Word Cloud (Bigrams) friends of a new happy text 11 friends of a new happy pick phonon pick from the same of the same of



Real Online Threats from SMS Spam Dataset

- Lures victims into free claim reward/romance scams.
- Pretends to be Government Agencies to gain trust.
- Fake undelivered parcels to trick users.
- Fake Business Opportunities & Investment Scams



I got an SMS from my bank asking me to verify my identity via a link. Would a real bank do that?

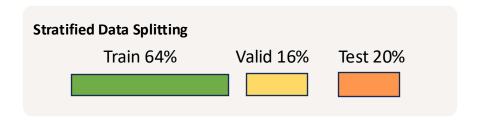


Do delivery services really ask you to confirm parcels through strange SMS links

Model Performance Comparison

Zero-Shot LLM with Prompt Engineering

BERT Fine-tuning



Candidate Model

- DistilBERT-base-uncased
- RoBERTa-base

smallest model, designed for speed and efficiency larger, higher F1 score, best for accuracy.

Performance evaluated on the same test set

Model	Precision	Recall	F1
GPT-4o mini (zero-shot)	0.8045	0.9664	0.8780
DistilBERT (fine-tuned)	0.9726	0.9530	0.9627
RoBERTa (fine-tuned)	0.9931	0.9597	0.9761

Key Evaluation Metric: F1 Score

Conclusion from Model Comparison

MODEL SUMMARY

RoBERTa achieved the highest F1 score → best overall spam detection balance. **DistilBERT** performed slightly lower but is more lightweight and efficient. **GPT-4o** (zero-shot) underperformed compared to fine-tuned models, but it required **no labeled training data**.

CONCLUSION

Fine-tuned transformer models remain superior for spam classification when labeled data is available. Zero-shot LLMs are useful when training data is limited.

Practical takeaway:

Use **RoBERTa** for accuracy-critical spam detection.
Use **DistilBERT** for lightweight, fast deployment.
Use **GPT-4o** for quick prototyping or when labeled data is scarce.

Appendix

Dataset

UCI Machine Learning Repository

SMS Spam Collection

Jun 21, 2012 — The SMS Spam Collection is a public set of SMS labeled messages that have been collecte...

Ham Word Cloud (Bigrams)

deres to the low of the low o



=== HAM : 5 sample ===

- Awww dat is sweet! We can think of something to do he he! Have a nice time tonight ill probably txt u later cos im lonely :(xxx.
- Have you been practising your curtsey?
- Sry da..jst nw only i came to home..
- Im good! I have been thinking about you...
- 7 wonders in My WORLD 7th You 6th Ur style 5th Ur smile 4th Ur Personality 3rd Ur Nature 2nd Ur SMS and 1st \Ur Lovely Friendship\"... good morning dear"

=== SPAM : 5 sample ===

- Reminder: You have not downloaded the content you have already paid for. Goto http://doit. mymoby. tv/ to collect your content.
- You are guaranteed the latest Nokia Phone, a 40GB iPod MP3 player or a \$f500 prize! Txt word: COLLECT to No: 83355! IBHltd LdnW15H 150p/Mtmsgrcvd18
- YOU ARE CHOSEN TO RECEIVE A &£350 AWARD! Pls call claim number 09066364311 to collect your award which you are selected to receive as a valued mobile customer.
- Get your garden ready for summer with a FREE selection of summer bulbs and seeds worth â£33:50 only with The Scotsman this Saturday. To stop go2 notxt.co.uk
 December only! Had your mobile 11mths+? You are entitled to update to the latest colour camera mobile for Free! Call The Mobile Update Co FREE on 08002986906

5,574 SMS messages labeled as either scam (spam) or legitimate (ham) with about 13.4% of the messages are spam.

