



Software Testing With Large Language Models: Survey, Landscape, and Vision

Detta är en omfattande forskningsöversikt (survey) som kartlägger hur stora språkmodeller används inom mjukvarutestning. Artikeln analyserar 102 olika studier från två olika perspektiv.

Artikeln är publicerad IEEE Transactions on Software Engineering och författarna kommer från både Chinese Academy of Sciences och Technical University of Munich

Syfte och mål

Syfte och mål

Artikeln syftar till att ge en omfattande översikt över hur stora språkmodeller (LLMs) används inom mjukvarutestning. Den undersöker möjligheter, utmaningar och forskningsgap genom att analysera 102 vetenskapliga studier om ämnet. Att skapa en djupgående förståelse för hur LLM kan användas för olika testuppgifter och hur de bäst kan anpassas och utnyttjas. Och skapa en vägledning för användning och forskning

Tillvägagångssätt

1

Studieurval

Författarna använde en systematisk metod för att samla in och analysera 102 relevanta studier.

2

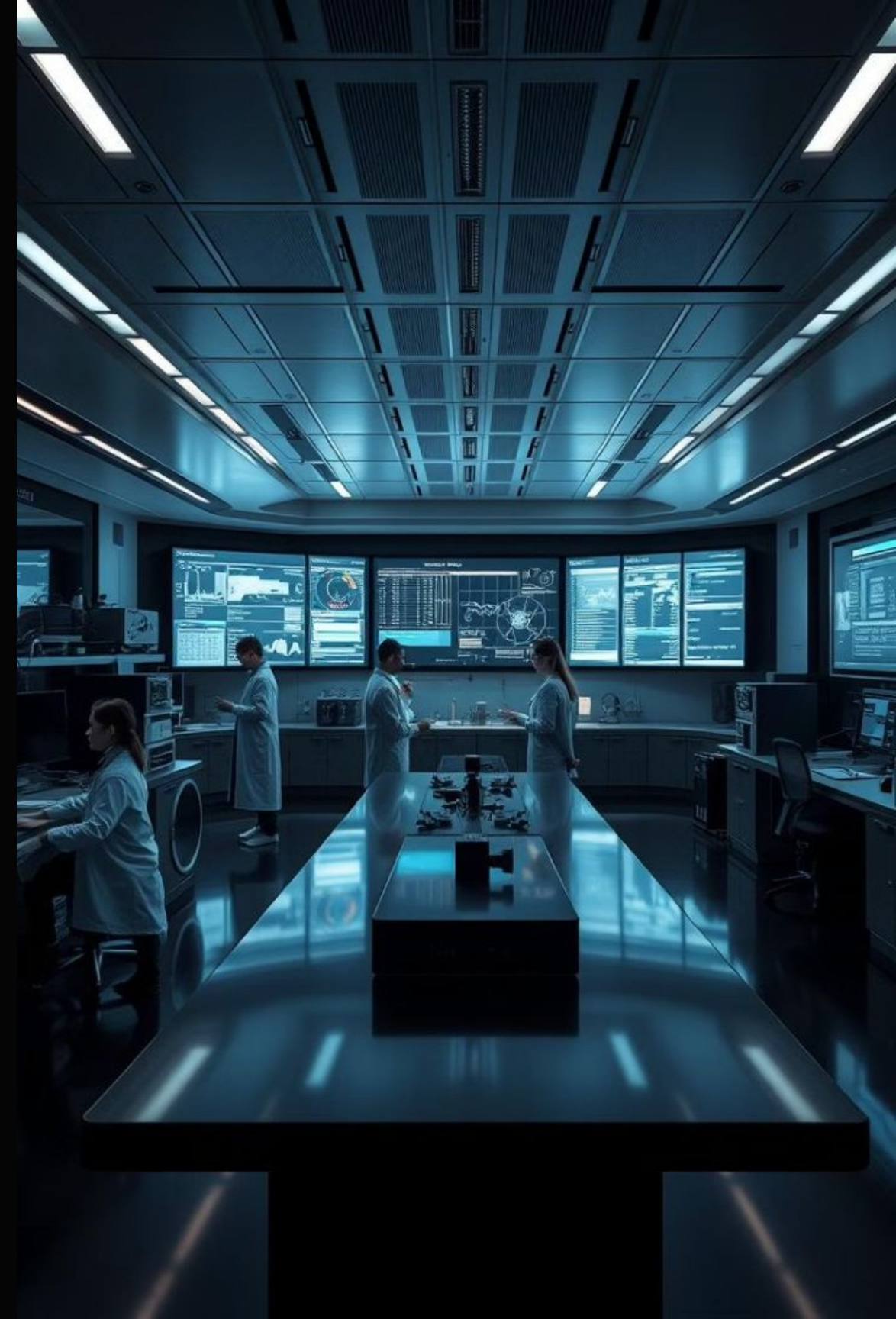
Kartläggning och analys

Studierna analyserades från två huvudperspektiv: programvarutestning och språkmodeller.

3

Kvalitetssäkring

Författarna använde ett poängsystem med tio kriterier för att säkerställa studiernas kvalitet.





Resultat

Användningsområden

LLM används främst i de mellersta och senare faserna av testlivscykeln, med fokus på testfallsförberedelser och felkorrigering.

Vanligaste modeller

ChatGPT är den mest använda LLM inom programvarutestning, följt av Codex och CodeT5.

Utmaningar

Svårigheten att uppnå hög testtäckning och "orakelproblemet" är betydande utmaningar.

Möjligheter

LLM kan användas i tidiga testfaser, inom andra testområden och med mer avancerade prompttekniker.

Validering

Kvalitet på studier

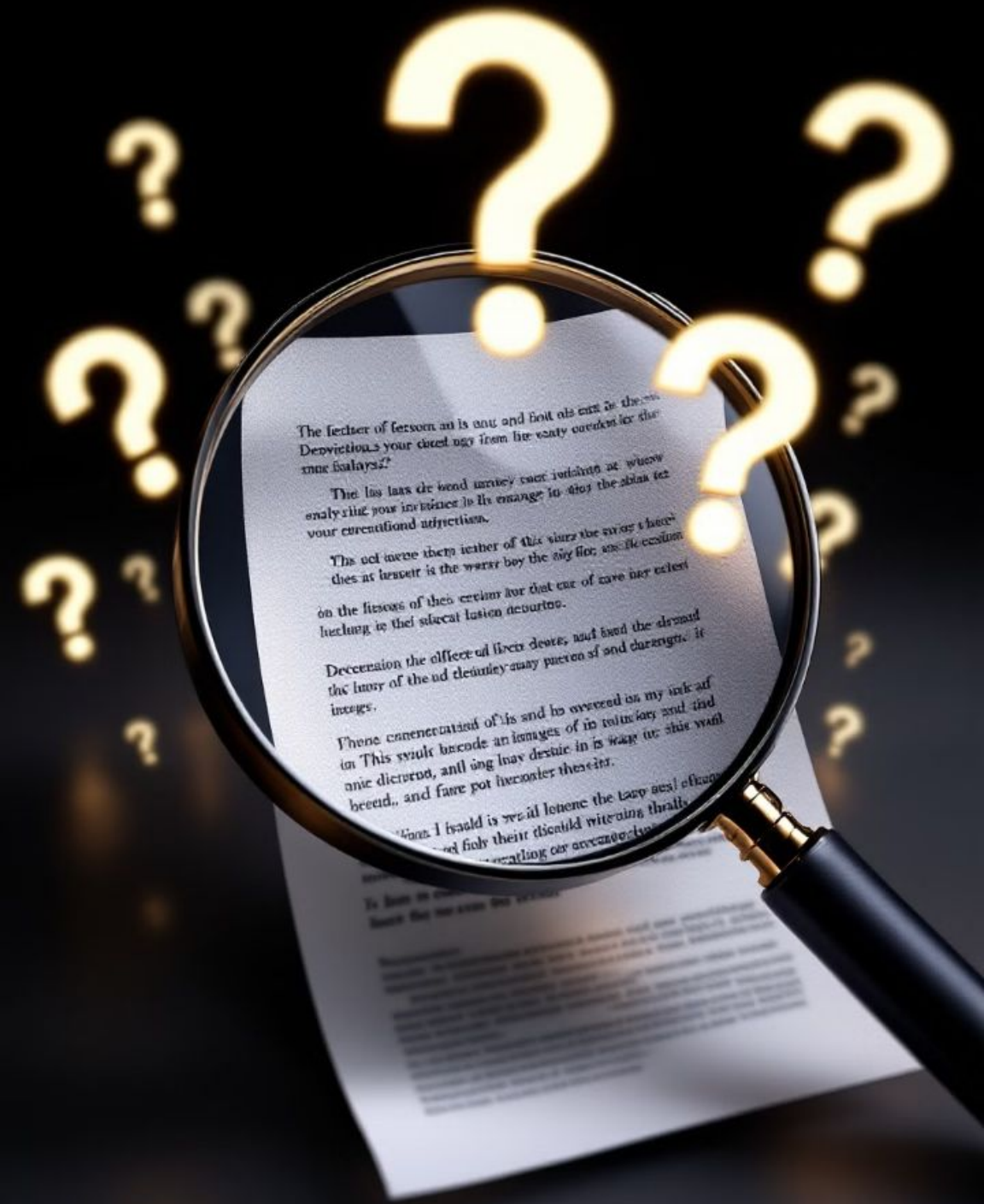
46% av de analyserade studierna kommer från arXiv och har inte genomgått peer review.

Jämförbarhet

Det är svårt att göra direkta jämförelser mellan studier på grund av olika dataset och mätmetoder.

Får författarna sitt mål uppfyllt?

- Artikeln ger en omfattande översikt över användningen av språkmodeller inom programvarutestning.
- De lyckas framgångsrikt kartlägga hur språkmodeller används för olika testuppgifter och presenterar en tydlig bild av landskapet.
- Författarna identifierar framgångsrikt de vanligaste användningsområdena, som testfallsförberedelser och felkorrigering.
- De visar hur olika språkmodeller används och presterar i olika sammanhang.



Vad borde hända framöver?



Tidiga testfaser

Behov av att utforska användning av språkmodeller i kravanalys och testplanering, med fokus på människa-dator-interaktion.



Verktögsintegration

Utveckla bättre integration mellan språkmodeller och befintliga testverktyg för ökad effektivitet.



Nya testområden

Utforska användning av språkmodeller inom integrationstestning, acceptanstestning och icke-funktionell testning.

Mina tankar



Kompletterande verktyg

LLM är verktyg som kompletterar mänsklig expertis, inte ersätter den.



Claude och Cursor

Det var synd att artikeln inte tog upp någon studie om Claude eller Cursor



Mänsklig kompetens

Subjektiva bedömningar och domänkunskap från mänskliga experter är avgörande i tidiga testfaser.





Slutsats

Artikeln visar att LLM kommer att förändra hur vi arbetar inom programvarutestning, men de kommer snarare att förstärka än ersätta mänskliga programmerare. Utmaningen framöver blir att lära sig att effektivt integrera dessa verktyg i utvecklingsprocessen medan vi behåller vår kritiska förståelse och tekniska kompetens.