## Empirical and Sustainability Aspects of Software Engineering Research in the Era of Large Language Models: A Reflection - Analysis Notebook

# Number of Papers at Each Stage

```
Total Papers: 692
Number of Papers matching AI Keywords: 304/692 (43.93%)
Number of Relevant Papers: 177/692 (25.58%)
Papers Per Year:
year
2023
     210
2024 236
2025 246
Name: title, dtype: int64
AI Keyword Papers Per Year:
year
       59
2023
2024
       99
2025 146
Name: title, dtype: int64
Relevant Papers Per Year:
2023
      32
2024 55
2025
       90
Name: title, dtype: int64
```

# RQ1 - Which LLMs are used in SE research and how are they benchmarked?

### Open vs. Closed (Commercial) Models

```
models_open_closed
open 71
both 65
closed 41
Name: count, dtype: int64
Open models in 136 out of 177 papers (76.8%)
Closed models in 106 out of 177 papers (59.9%)
```

#### 2023 Papers:

Only open models in 22 out of 32 papers (68.8%)
Only closed models in 4 out of 32 papers (12.5%)
Open models in 28 out of 32 papers (87.5%)
Closed models in 10 out of 32 papers (31.2%)
Both model types in 6 out of 32 papers (18.8%)

#### 2024 Papers:

Only open models in 25 out of 55 papers (45.5%)
Only closed models in 13 out of 55 papers (23.6%)
Open models in 42 out of 55 papers (76.4%)
Closed models in 30 out of 55 papers (54.5%)
Both model types in 17 out of 55 papers (30.9%)

#### 2025 Papers:

Only open models in 24 out of 90 papers (26.7%)
Only closed models in 24 out of 90 papers (26.7%)
Open models in 66 out of 90 papers (73.3%)
Closed models in 66 out of 90 papers (73.3%)
Both model types in 42 out of 90 papers (46.7%)

### **Model Families**

model_families_list         47           GPT-4         47           GPT-3.5         44           CodeBRT         34           CodeLlama         26           CodeT5         22           CodeGen         19           StarCoder         18           GraphCodeBERT         18           Llama         17           RoBERTa         15           ChatGPT         14           BERT         13           DeepSeekCoder         12           UniXcoder         11           Codex         10           InCoder         19           T5         8           Claude         7           Gemini         7           ChatGLM         6           DeepSeek         6           CodeQwen         6           GPT-3         6           UnixCoder         6           PLBART         6           CodeGPT         5           CodeParrot         4           Cojlot         4           WizardCoder         4           GPT-2         4           PolyCoder         4	Overall - Number of	Papers	per	Model	Family:
GPT-4 GPT-3.5 GPT-3.5 GPT-3.5 GPT-3.5 GPT-3.5 GPT-3.5 GPT-3.5 GODEBERT GODELT G					
GPT-3.5  CodeBERT  GOMEBERT  GOMEDTS  COGEGEN  STARCODER  STARCODE	_			47	
CodeBERT CodeLlama CodeLlama CodeCodeT5 CodeGen 19 StarCoder 18 GraphCodeBERT 18 Llama 17 ROBERTa 15 ChatGPT 14 BERT 13 DeepSeekCoder 12 UniXcoder 11 Codex 10 InCoder 19 T5 8 Claude 7 Gemini 7 ChatGLM 6 DeepSeek 6 CodeQwen 6 GPT-3 0 UniXcoder 6 COdeST 1 UniXcoder 7 Gemini 7 ChatGLM 6 DeepSeek 6 CodeQwen 6 GPT-3 16 UnixCoder 17 CodeParrot 18 CodeParrot 19 CodeParrot 4 Copilot 4 WizardCoder 4 GPT-2 4 PolyCoder 4 Mistral 3 Gemma 3 DistilBERT 3 LineVul 3 CodeT5+ 3 LineVul 3 CodeT5+ 3 LineVul 3 CodeGemma 1 TransCoder 2 CodeGeemma 1 TransCoder 2 CodeGeemma 1 TransCoder 2 CodeGeex 1 CodeGer 1 CodeGeex 1 CodeGer 2 CodeGeex 2 CodeGeex 2 CodeGeex 2 CodeGeex 3 CodeFT-J 3 CodeFT-J 4 CodeFT-J 4 CodeFT-J 5 CodeGeT-J 6 CodeCer 7 CodeFT-J 7 CodeFT-NEO 7 CODESTT 7 COTENT 7					
CodeLlama       26         CodeT5       22         CodeGen       19         StarCoder       18         GraphCodeBERT       18         Llama       17         ROBERTa       15         ChatGPT       14         BERT       13         DeepSeekCoder       12         UniXcoder       11         Codex       10         InCoder       9         T5       8         Claude       7         Gemini       7         Gemini       7         Gemini       7         ChatGLM       6         DeepSeek       6         CodeQwen       6         GPT-3       6         UnixCoder       6         PLBART       6         CodeGPT       5         CodeGPT       5         CodeParrot       4         Copilot       4         WizardCoder       4         GPT-2       4         PolyCoder       4         Mistral       3         Gemma       3         text-davinci       3					
CodeFTS       22         CodeGen       19         StarCoder       18         GraphCodeBERT       18         Llama       17         ROBERTA       15         ChatGPT       14         BERT       13         DeepSeekCoder       12         UniXcoder       10         InCodex       10         InCoder       9         T5       8         Claude       7         Gemini       7         ChatGLM       6         DeepSeek       6         CodeQwen       6         GPT-3       6         UnixCoder       6         GPT-3       6         UnixCoder       6         GPBART       6         CodeGPT       5         CodeGPT       5         CodePorto       4         Kopilot       4         WizardCoder       4         GPT-2       4         PolyCoder       4         Mistral       3         Gemma       3         text-davinci       3         Pythia       3				26	
CodeGen       19         StarCoder       18         GraphCodeBERT       18         Llama       17         ROBERTa       15         ChatGPT       14         BERT       13         DeepSeekCoder       12         UnixCoder       10         InCodex       10         InCoder       9         T5       8         Claude       7         Gemini       7         ChatGLM       6         DeepSeek       6         CodeQwen       6         GPT-3       6         UnixCoder       6         PLBART       6         CodeGPT       5         CodeGPT       5         CodeGPT       4         CodeJot       4         WizardCoder       4         GPT-2       4         PolyCoder       4         Mistral       3         Gemma       3         text-davinci       3         Pythia       3         Viuna       3         DistilBERT       3         LineVul       3 <t< td=""><td></td><td></td><td></td><td></td><td></td></t<>					
StarCoder       18         GraphCodeBERT       18         Llama       17         ROBERTa       15         ChatGPT       14         BERT       13         DeepSeekCoder       12         UniXcoder       11         Codex       10         InCoder       9         T5       8         Claude       7         Gemini       6         CodeGeWen       6         GPT-3       6         UnixCoder       6         CodePT-3       5         CodePT-2       4         POJyCoder       4         Mistral					
GraphCodeBERT 18 Llama 17 ROBERTA 15 ChatGPT 14 BERT 13 DeepSeekCoder 12 UniXcoder 11 Codex 10 InCoder 9 T5 8 Claude 7 Gemini 7 ChatGLM 6 DeepSeek 66 CodeQwen 66 GPT-3 6 UnixCoder 6 PLBART 6 CodeParrot 4 Copilot 4 WizardCoder 4 GPT-2 4 PolyCoder 4 Mistral 3 Gemma 3 text-davinci 3 Pythia 3 Vicuna 3 DistilBERT 3 LineVul 3 CodeT5+ 3 text-embedding 3 TransCoder 2 CodeGeex 2 CodeGer 3 Cod					
Llama 17 ROBERTA 15 ChatGPT 14 BERT 13 DeepSeekCoder 12 UniXcoder 11 Codex 10 InCoder 9 T5 8 Claude 7 Gemini 7 ChatGLM 6 DeepSeek 6 CodeQwen 6 GPT-3 6 UnixCoder 6 PLBART 6 CodeGPT 5 CodeParrot 4 Copilot 4 WizardCoder 4 GPT-2 4 PolyCoder 4 Mistral 3 Gemma 3 Gemma 3 Text-davinci 3 Pythia 3 Vicuna 3 DistilBERT 1 LineVul 3 CodeT5+ 1 text-embedding 3 TransCoder 2 CodeGeex 1 CodeGeex 1 CodeGeex 2 CodeGeex 1 CodeGeex 1 CodeGeex 2 CodeGeex 2 CodeGeex 2 CodeGeex 1 CodeGeex 2 CodeGeex 2 CodeGeex 1 CodeGeex 1 CodeGeex 2 CodeGeex 1 CodeGeex 2 CodeGeex 2 CodeGeex 1 CodeGeex 2 CodeGeex 2 CodeGeex 2 CodeGeex 1 CodeGeex 2 CodeGeex 3 CodeGeex 3 CodeGeex 3 CodeGeex 3 CodeGeex 4 CodeGeex 3 CodeGeex 4 Co					
ROBERTA ChatGPT 14 BERT 13 DeepSeekCoder 12 UniXcoder 11 Codex 10 InCoder 9 T5 8 Claude 7 Gemini 7 ChatGLM 6 DeepSeek 6 CodeQwen 6 GPT-3 10 UniXcoder 6 PLBART CodeSPT CodeParrot Copilot WizardCoder 4 WizardCoder 4 WizardCoder 4 Mistral 3 Gemma 3 text-davinci 9 Pythia Vicuna 3 DistilBERT 1 LineVul CodeTs+ 1 SineVul CodeSertal Qwen 2 GPT3.5 CodeGemma Incoder CodeGemma Incoder CodeGex VulberTa CubeRT CodeGex VulbeRTT CodeGex VulbeRTT CodeGex VulbeRTT CodeGex VulbeRTT CodeGex VulbeRTT CodeGex VulbeRTT CodeGex CodeGemma Incoder CodeGex VulbeRTT CodeGex VulbeRTT CodeGex CodeGemma CodeT- CodeGex CodeGemma CodeT- CodeGex CodeGex CodeGemma CodeT- CodeGex Co	•				
ChatGPT 14 BERT 13 DeepSeekCoder 12 UniXcoder 11 Codex 10 InCoder 9 T5 8 Claude 7 Gemini 7 ChatGLM 6 DeepSeek 6 CodeQwen 6 GPT-3 6 UniXcoder 6 PLBART 6 CodeGPT 5 CodeParrot 4 Copilot 4 WizardCoder 4 GPT-2 4 PolyCoder 4 Mistral 3 Gemma 3 text-davinci 3 Pythia 3 Vicuna 3 DistilBERT 3 LineVul 3 CodeT5+ 3 text-embedding 3 TransCoder 2 CodeSemma 1 Incoder 2 CodeGemma 2 CodeGemma 1 Incoder 2 CodeGemma 2 CodeGeeX 2 VulbERTa 2 CubERT 1 CubERT 2 CubERT 1 Cub					
BERT       13         DeepSeekCoder       12         UniXcoder       11         Codex       10         InCoder       9         T5       8         Claude       7         Gemini       7         ChatGLM       6         DeepSeek       6         CodeQwen       6         GPT-3       6         Unixcoder       6         PLBART       6         CodeGPT       5         CodeParrot       4         Copilot       4         WizardCoder       4         GPT-2       4         PolyCoder       4         Mistral       3         Gemma       3         text-davinci       3         Pythia       3         Vicuna       3         DistilBERT       3         LineVul       3         CodeT5+       3         text-embedding       3         TransCoder       2         CodeGemma       2         Incoder       2         CodeGeexX       2         VulBERTa       2					
DeepSeekCoder       12         UniXcoder       11         Codex       10         InCoder       9         T5       8         Claude       7         Gemini       7         ChatGLM       6         DeepSeek       6         CodeQwen       6         GPT-3       6         UnixCoder       6         PLBART       6         CodeGPT       5         CodeGePT of       5         CodeParrot       4         Copilot       4         WizardCoder       4         GPT-2       4         PolyCoder       4         Mistral       3         Gemma       3         text-davinci       3         Pythia       3         Vicuna       3         DistilBERT       3         LineVul       3         CodeT5+       3         text-embedding       3         TransCoder       2         CodeGemma       2         LineVal       2         CodeGeex       2         VulBERTa       2 </td <td></td> <td></td> <td></td> <td></td> <td></td>					
UniXcoder 11 Codex 100 InCoder 99 T5 88 Claude 77 Gemini 77 ChatGLM 66 DeepSeek 66 CodeQwen 66 GPT-3 66 UnixCoder 66 PLBART 66 CodeGPT 55 CodeParrot 44 Copilot 44 WizardCoder 46 GPT-2 44 PolyCoder 44 Mistral 33 Gemma 33 text-davinci 33 Pythia 33 Vicuna 33 DistilBERT 33 LineVul 33 CodeT5+ 33 text-embedding 33 TransCoder 22 Codestral 22 Qwen 22 GPT3.5 22 CodeGeex 22 VulBERT 22 UniLog 22 BART 22 GPT-U 22 GPT-Neo 22 GPT-Neo 22 GPT-Neo 25 GP					
Codex       10         InCoder       9         T5       8         Claude       7         Gemini       7         ChatGLM       6         DeepSeek       6         CodeQwen       6         GPT-3       6         UnixCoder       6         PLBART       6         CodeGPT       5         CodeParrot       4         Copilot       4         WizardCoder       4         GPT-2       4         PolyCoder       4         Mistral       3         Gemma       3         text-davinci       3         Pythia       3         Vicuna       3         DistilBERT       3         LineVul       3         CodeT5+       3         text-embedding       3         TransCoder       2         CodeSetral       2         Qwen       2         GPT3.5       2         CodeGeeX       2         VulBERTa       2         UniLog       2         GPT-Neo       2					
InCoder       9         T5       8         Claude       7         Gemini       7         ChatGLM       6         DeepSeek       6         CodeQwen       6         GPT-3       6         UnixCoder       6         PLBART       6         CodeGPT       5         CodeParrot       4         Copilot       4         WizardCoder       4         GPT-2       4         PolyCoder       4         Mistral       3         Gemma       3         text-davinci       3         Pythia       3         Vicuna       3         DistilBERT       3         LineVul       3         CodeT5+       3         text-embedding       3         TransCoder       2         CodeSemma       2         Incoder       2         CodeGeemma       2         Incoder       2         CodeGeeX       2         VulBERT       2         UniLog       2         BART       2					
T5       8         Claude       7         Gemini       7         ChatGLM       6         DeepSeek       6         CodeQwen       6         GPT-3       6         UnixCoder       6         PLBART       6         CodeGPT       5         CodeParrot       4         Copilot       4         WizardCoder       4         GPT-2       4         PolyCoder       4         Mistral       3         Gemma       3         text-davinci       3         Pythia       3         Vicuna       3         DistilBERT       3         LineVul       3         CodeT5+       3         text-embedding       3         TransCoder       2         Codestral       2         Qwen       2         GPT3.5       2         CodeGeemma       2         Incoder       2         CodeGeeX       2         VulBERTa       2         UniLog       2         BART       2					
Claude 7 Gemini 7 ChatGLM 6 DeepSeek 6 CodeQwen 6 GPT-3 6 UnixCoder 6 PLBART 6 CodeGPT 5 CodeParrot 4 Copilot 4 WizardCoder 4 GPT-2 4 PolyCoder 4 Mistral 3 Gemma 3 text-davinci 3 Pythia 3 Vicuna 3 DistilBERT 3 LineVul 3 CodeT5+ 3 text-embedding 3 TransCoder 2 Codestral 2 Qwen 2 GPT3.5 2 CodeGemma 2 Incoder 2 CodeGeeX 2 VulBERTa 2 CubERT 1 UniLog BART 2 GPT-J 2 GPT-NEO GPT-C 5 SynCoBERT 2 CoTexT 2 SynCoBERT 2 CoTexT 2 CodeRIC CODEST CODEST COTEXT 2 CODEST COD					
Gemini       7         ChatGLM       6         DeepSeek       6         CodeQwen       6         GPT-3       6         UnixCoder       6         PLBART       6         CodeGPT       5         CodeParrot       4         Copilot       4         WizardCoder       4         GPT-2       4         PolyCoder       4         Mistral       3         Gemma       3         text-davinci       3         Pythia       3         Vicuna       3         DistilBERT       3         LineVul       3         CodeT5+       3         text-embedding       3         TransCoder       2         Codestral       2         Qwen       2         GPT3.5       2         CodeGemma       2         Incoder       2         CodeGeeX       2         VulBERTa       2         UniLog       2         BART       2         GPT-Neo       2         GPT-Neo       2					
ChatGLM       6         DeepSeek       6         CodeQwen       6         GPT-3       6         UnixCoder       6         PLBART       6         CodeGPT       5         CodeParrot       4         Copilot       4         WizardCoder       4         GPT-2       4         PolyCoder       4         Mistral       3         Gemma       3         text-davinci       3         Pythia       3         Vicuna       3         DistilBERT       3         LineVul       3         CodeT5+       3         text-embedding       3         TransCoder       2         Codestral       2         Qwen       2         GPT3.5       2         CodeGeemma       2         Incoder       2         CodeGeeX       2         VulBERTa       2         UniLog       2         BART       2         GPT-Neo       2         GPT-Neo       2         GPT-Neo       2 <tr< td=""><td></td><td></td><td></td><td></td><td></td></tr<>					
DeepSeek       6         CodeQwen       6         GPT-3       6         UnixCoder       6         PLBART       6         CodeGPT       5         CodeParrot       4         Copilot       4         WizardCoder       4         GPT-2       4         PolyCoder       4         Mistral       3         Gemma       3         text-davinci       3         Pythia       3         Vicuna       3         DistilBERT       3         LineVul       3         CodeT5+       3         text-embedding       3         TransCoder       2         Codestral       2         Qwen       2         GPT3.5       2         CodeGeemma       2         Incoder       2         CodeGeeX       2         VulBERTa       2         UniLog       2         BART       2         GPT-Neo       2         GPT-Neo       2         GPT-NEO       2         GPT-C       2					
CodeQwen       6         GPT-3       6         UnixCoder       6         PLBART       6         CodeGPT       5         CodeParrot       4         Copilot       4         WizardCoder       4         GPT-2       4         PolyCoder       4         Mistral       3         Gemma       3         text-davinci       3         Pythia       3         Vicuna       3         DistilBERT       3         LineVul       3         CodeT5+       3         text-embedding       3         TransCoder       2         Codestral       2         Qwen       2         GPT3.5       2         CodeGeemma       2         Incoder       2         CodeGeeX       2         VulBERTa       2         UniLog       2         BART       2         GPT-Neo       2         GPT-NEO       2         GPT-NEO       2         GPT-C       2         SynCoBERT       2					
GPT-3       6         UnixCoder       6         PLBART       6         CodeGPT       5         CodeParrot       4         Copilot       4         WizardCoder       4         GPT-2       4         PolyCoder       4         Mistral       3         Gemma       3         text-davinci       3         Pythia       3         Vicuna       3         DistilBERT       3         LineVul       3         CodeT5+       3         text-embedding       3         TransCoder       2         Codestral       2         Qwen       2         GPT3.5       2         CodeGeemma       2         Incoder       2         CodeGeeX       2         VulBERTa       2         UniLog       2         BART       2         GPT-NEO       2         GPT-NEO       2         GPT-NEO       2         GPT-C       2         SynCoBERT       2         CoTexT       2					
UnixCoder PLBART 6 PLBART 6 CodeGPT 5 CodeParrot 4 Copilot 4 WizardCoder 4 GPT-2 4 PolyCoder 4 Mistral 3 Gemma 3 text-davinci 3 Pythia Vicuna 3 DistilBERT LineVul 3 CodeT5+ 3 text-embedding TransCoder 2 Codestral 2 Qwen 2 GPT3.5 2 CodeGemma 2 Incoder 2 CodeGeeX VulBERTa 2 UniLog BART CUBERT UniLog BART GPT-J GPT-NEO GPT-NEO GPT-C SynCoBERT COTEXT 2 COTEXT 2 CODEST  6 6 6  PLBART A 4 CODEST A 5 CODEST	-				
PLBART       6         CodeGPT       5         CodeParrot       4         Copilot       4         WizardCoder       4         GPT-2       4         PolyCoder       4         Mistral       3         Gemma       3         text-davinci       3         Pythia       3         Vicuna       3         DistilBERT       3         LineVul       3         CodeT5+       3         text-embedding       3         TransCoder       2         Codestral       2         Qwen       2         GPT3.5       2         CodeGeemma       2         Incoder       2         CodeGeeX       2         VulBERTa       2         UniLog       2         BART       2         GPT-Neo       2         GPT-Neo       2         GPT-NEO       2         GPT-C       2         SynCoBERT       2         CoTexT       2					
CodeGPT       5         CodeParrot       4         Copilot       4         WizardCoder       4         GPT-2       4         PolyCoder       4         Mistral       3         Gemma       3         text-davinci       3         Pythia       3         Vicuna       3         DistilBERT       3         LineVul       3         CodeT5+       3         text-embedding       3         TransCoder       2         Codestral       2         Qwen       2         GPT3.5       2         CodeGeemma       2         Incoder       2         CodeGeeX       2         VulBERTa       2         UniLog       2         BART       2         GPT-Neo       2         GPT-Neo       2         GPT-NEO       2         GPT-C       2         SynCoBERT       2         CoTexT       2					
CodeParrot       4         Copilot       4         WizardCoder       4         GPT-2       4         PolyCoder       4         Mistral       3         Gemma       3         text-davinci       3         Pythia       3         Vicuna       3         DistilBERT       3         LineVul       3         CodeT5+       3         text-embedding       3         TransCoder       2         Codestral       2         Qwen       2         GPT3.5       2         CodeGeemma       2         Incoder       2         CodeGeeX       2         VulBERTa       2         UniLog       2         BART       2         GPT-J       2         GPT-NEO       2         GPT-NEO       2         GPT-NEO       2         GPT-C       2         SynCoBERT       2         CoTexT       2					
Copilot       4         WizardCoder       4         GPT-2       4         PolyCoder       4         Mistral       3         Gemma       3         text-davinci       3         Pythia       3         Vicuna       3         DistilBERT       3         LineVul       3         CodeT5+       3         text-embedding       3         TransCoder       2         Codestral       2         Qwen       2         GPT3.5       2         CodeGemma       2         Incoder       2         CodeGeeX       2         VulBERTa       2         UniLog       2         BART       2         GPT-J       2         GPT-Neo       2         GPT-Neo       2         GPT-NEO       2         GPT-C       2         SynCoBERT       2         CoTexT       2					
WizardCoder       4         GPT-2       4         PolyCoder       4         Mistral       3         Gemma       3         text-davinci       3         Pythia       3         Vicuna       3         DistilBERT       3         LineVul       3         CodeT5+       3         text-embedding       3         TransCoder       2         Codestral       2         Qwen       2         GPT3.5       2         CodeGemma       2         Incoder       2         CodeGeeX       2         VulBERTa       2         UniLog       2         BART       2         GPT-J       2         GPT-Neo       2         GPT-NEO       2         GPT-NEO       2         GPT-C       2         SynCoBERT       2         CoTexT       2					
GPT-2       4         PolyCoder       4         Mistral       3         Gemma       3         text-davinci       3         Pythia       3         Vicuna       3         DistilBERT       3         LineVul       3         CodeT5+       3         text-embedding       3         TransCoder       2         Codestral       2         Qwen       2         GPT3.5       2         CodeGeemma       2         Incoder       2         CodeGeeX       2         VulBERTa       2         UniLog       2         BART       2         GPT-J       2         GPT-NEO       2         GPT-NEO       2         GPT-NEO       2         GPT-C       2         SynCoBERT       2         CoTexT       2					
PolyCoder Mistral 3 Gemma 3 text-davinci 3 Pythia 3 Vicuna 3 DistilBERT 3 LineVul 3 CodeT5+ 3 text-embedding 3 TransCoder 2 Codestral 2 Qwen 2 GPT3.5 2 CodeGeemma 2 Incoder 2 CodeGeeX 2 VulBERT 2 UniLog BART 2 GPT-J 2 GPT-NEO 2 GPT-NEO 2 GPT-C 2 SynCoBERT 2 CoTexT 2					
Mistral       3         Gemma       3         text-davinci       3         Pythia       3         Vicuna       3         DistilBERT       3         LineVul       3         CodeT5+       3         text-embedding       3         TransCoder       2         Codestral       2         Qwen       2         GPT3.5       2         CodeGemma       2         Incoder       2         CodeGeeX       2         VulBERTa       2         CubERT       2         UniLog       2         BART       2         GPT-J       2         GPT-Neo       2         GPT-NEO       2         GPT-C       2         SynCoBERT       2         CoTexT       2					
Gemma       3         text-davinci       3         Pythia       3         Vicuna       3         DistilBERT       3         LineVul       3         CodeT5+       3         text-embedding       3         TransCoder       2         Codestral       2         Qwen       2         GPT3.5       2         CodeGemma       2         Incoder       2         CodeGeeX       2         VulBERTa       2         CuBERT       2         UniLog       2         BART       2         GPT-J       2         GPT-Neo       2         GPT-NEO       2         GPT-C       2         SynCoBERT       2         CoTexT       2					
text-davinci 3 Pythia 3 Vicuna 3 DistilBERT 3 LineVul 3 CodeT5+ 3 text-embedding 3 TransCoder 2 Codestral 2 Qwen 2 GPT3.5 2 CodeGemma 2 Incoder 2 CodeGeeX 2 VulBERTa 2 CUBERT 2 UniLog 2 BART 2 GPT-J 2 GPT-Neo 2 GPT-NEO 2 GPT-C 5 SynCoBERT 2 COTEXT 2					
Pythia 3 Vicuna 3 DistilBERT 3 LineVul 3 CodeT5+ 3 text-embedding 3 TransCoder 2 Codestral 2 Qwen 2 GPT3.5 2 CodeGemma 2 Incoder 2 CodeGeeX 2 VulBERTa 2 CUBERT 2 UniLog 2 BART 2 GPT-J 2 GPT-Neo 2 GPT-Neo 2 GPT-NEO 2 GPT-C 5 SynCoBERT 2 COTEXT 2					
Vicuna       3         DistilBERT       3         LineVul       3         CodeT5+       3         text-embedding       3         TransCoder       2         Codestral       2         Qwen       2         GPT3.5       2         CodeGemma       2         Incoder       2         CodeGeeX       2         VulBERTa       2         CuBERT       2         UniLog       2         BART       2         GPT-J       2         GPT-Neo       2         GPT-NEO       2         GPT-C       2         SynCoBERT       2         CoTexT       2					
DistilBERT       3         LineVul       3         CodeT5+       3         text-embedding       3         TransCoder       2         Codestral       2         Qwen       2         GPT3.5       2         CodeGemma       2         Incoder       2         CodeGeeX       2         VulBERTa       2         CuBERT       2         UniLog       2         BART       2         GPT-J       2         GPT-NEO       2         GPT-NEO       2         GPT-C       2         SynCoBERT       2         CoTexT       2	•				
LineVul       3         CodeT5+       3         text-embedding       3         TransCoder       2         Codestral       2         Qwen       2         GPT3.5       2         CodeGemma       2         Incoder       2         CodeGeeX       2         VulBERTa       2         CuBERT       2         UniLog       2         BART       2         GPT-J       2         GPT-NEO       2         GPT-NEO       2         GPT-C       2         SynCoBERT       2         CoTexT       2					
CodeT5+       3         text-embedding       3         TransCoder       2         Codestral       2         Qwen       2         GPT3.5       2         CodeGemma       2         Incoder       2         CodeGeeX       2         VulBERTa       2         CuBERT       2         UniLog       2         BART       2         GPT-J       2         GPT-Neo       2         GPT-NEO       2         GPT-C       2         SynCoBERT       2         CoTexT       2					
text-embedding       3         TransCoder       2         Codestral       2         Qwen       2         GPT3.5       2         CodeGemma       2         Incoder       2         CodeGeeX       2         VulBERTa       2         CuBERT       2         Unilog       2         BART       2         GPT-J       2         GPT-Neo       2         GPT-NEO       2         GPT-C       2         SynCoBERT       2         CoTexT       2					
TransCoder       2         Codestral       2         Qwen       2         GPT3.5       2         CodeGemma       2         Incoder       2         CodeGeeX       2         VulBERTa       2         CuBERT       2         UniLog       2         BART       2         GPT-J       2         GPT-Neo       2         GPT-NEO       2         GPT-C       2         SynCoBERT       2         CoTexT       2					
Codestral       2         Qwen       2         GPT3.5       2         CodeGemma       2         Incoder       2         CodeGeeX       2         VulBERTa       2         CuBERT       2         UniLog       2         BART       2         GPT-J       2         GPT-Neo       2         GPT-NEO       2         GPT-C       2         SynCoBERT       2         CoTexT       2					
Qwen       2         GPT3.5       2         CodeGemma       2         Incoder       2         CodeGeeX       2         VulBERTa       2         CuBERT       2         UniLog       2         BART       2         GPT-J       2         GPT-Neo       2         GPT-NEO       2         GPT-C       2         SynCoBERT       2         CoTexT       2					
GPT3.5 2 CodeGemma 2 Incoder 2 CodeGeeX 2 VulBERTa 2 CUBERT 2 UniLog 2 BART 2 GPT-J 2 GPT-Neo 2 GPT-Neo 2 GPT-NEO 2 GPT-C 2 SynCoBERT 2 COTexT 2					
CodeGemma       2         Incoder       2         CodeGeeX       2         VulBERTa       2         CuBERT       2         UniLog       2         BART       2         GPT-J       2         GPT-Neo       2         GPT-NEO       2         GPT-C       2         SynCoBERT       2         CoTexT       2					
Incoder       2         CodeGeeX       2         VulBERTa       2         CuBERT       2         UniLog       2         BART       2         GPT-J       2         GPT-Neo       2         GPT-NEO       2         GPT-C       2         SynCoBERT       2         CoTexT       2					
CodeGeeX       2         VulBERTa       2         CuBERT       2         UniLog       2         BART       2         GPT-J       2         GPT-Neo       2         GPT-NEO       2         GPT-C       2         SynCoBERT       2         CoTexT       2					
VulBERTa       2         CuBERT       2         UniLog       2         BART       2         GPT-J       2         GPT-Neo       2         GPT-NEO       2         GPT-C       2         SynCoBERT       2         CoTexT       2					
CuBERT       2         UniLog       2         BART       2         GPT-J       2         GPT-Neo       2         GPT-NEO       2         GPT-C       2         SynCoBERT       2         CoTexT       2					
UniLog       2         BART       2         GPT-J       2         GPT-Neo       2         GPT-NEO       2         GPT-C       2         SynCoBERT       2         CoTexT       2					
BART 2 GPT-J 2 GPT-Neo 2 GPT-NEO 2 GPT-C 2 SynCoBERT 2 COTexT 2					
GPT-J       2         GPT-Neo       2         GPT-NEO       2         GPT-C       2         SynCoBERT       2         CoTexT       2					
GPT-Neo       2         GPT-NEO       2         GPT-C       2         SynCoBERT       2         CoTexT       2					
GPT-NEO 2 GPT-C 2 SynCoBERT 2 CoTexT 2					
GPT-C 2 SynCoBERT 2 CoTexT 2					
SynCoBERT 2 CoTexT 2					
CoTexT 2					
	-				
Ph1 2					
	Phi			2	

TF:	2
TFix	2
ALBERT	2
seBERT	2
Code-davinci	2
OpenDevin	2
SantaCoder	2
Longformer	1
CoditT5	1
Unixcoder	1
CodeBert	1
GPT4	1
	1
SVulD	
Poro	1
ChatDev	1
Self-collaboration	1
MetaGPT	1
AutoGPT	1
Multi-Turn Program Synthesis	1
AgentCoder	1
DetectGPT	1
GPT-2 Output Detector	1
GPTZero	1
GPTSniffer	1
Starcoder	1
LLM-Parser	1
LILAC	1
Repilot	1
RAP-Gen	1
ChatRepair	1
•	
FitRepair	1
AlphaRe-pair	1
Mixtral	1
VGX	1
ReVeal	1
Devign	1
0	_
VULGEN	1
COME	1
CCT5	1
NNGen	1
ALL-MINILM-L6-V210	1
UniTrans	1
Shipwright	1
MagiCoder	1
Magicoder	1
Parfum	1
TOGA	1
AthenTest	1
SEQ Graph& HYBRID	1
AppMap Naive	1
OpenCodeInterpreter	1
AutoCodeRover	1
CodeShell	1
	1
LLama	
CoCoSoDa	1
CodeRetriever	1
HedgeCode	1
SYNCOBERT	1
Vercel	1
	1
GPT-4, CodeBERT	_
Moatless Tools	1

Agentless	1
FuzzGPT	1
TitanFuzz	1
Aider	1
SWE-Agent	1
DeBERTa	1
OPT	1
GPT3-5	1
Tulu	1
Guanaco	1
PaLM	1
StarChat	1
CAT-LM	1
Exlong	1
GLTR	1
ContraBERT	1
ChatUniTest	1
TestGen-LLM	1
RustAssistant	1
Sonnet	1
Stable-Code	1
BigBird	1
Sapling	1
	1
KeyBERT	
DISCO	1
PDBERT	1
GPTBigCode	1
Sentence-BERT	1
Airboros	1
CodeBERTa	1
Flan	1
code-davinci	1
UnifiedQA	1
VRepair	1
CodeReviewer	1
GrammarT5	1
Transformer	1
LSTM	1
GIN	1
	1
TypeFix	_
PyTER	1
CoCoNuT	1
AlphaRepair	1
LANCE	1
XLNet	1
RepresentThemAll	1
Curie	1
Davinci	1
ELECTRA	1
MiniLM	1
DOBF	1
VulRepair	1
VulMaster	1
PanguCoder	1
flan-alpaca	1
SPT-Code	1
ProphetNet-Code	1
T5-learning	1
JavaBERT	1
	1
DeepDebug	1

C-BERT	1
CugLM	1
TreeBERT	1
PLBart	1
ATLAS	1
CoCoNut	1
Hoppity	1
Sequencer	1
CEDAR	1
Transformers	1
GPT-NeoX	1
FAIR	1
OSCAR	1
Transcoder*	1
IRGen	1
Deep-SE	1
GPT2SP	1
sentenceBERT	1
SDA-Trans	1
StableCode	1
CodeGen-NL	1
CodeGen-Mono	1
CodeGen-Multi	1
CodeGen2	1
PyCodeGPT	1
GPT-Code-Clippy	1
BERTOverflow	1

Name: count, dtype: int64

2023 - Number of Pap	pers	per	Model	Family:
model_families_list		F		, .
CodeBERT	11			
RoBERTa	8			
BERT	7			
CodeT5	7			
Codex	5			
T5	5			
GraphCodeBERT	4			
DistilBERT	3			
PLBART	2			
GPT-J	2			
CodeGen	2			
BART	2			
InCoder GPT-Neo	2 2			
XLNet	1			
MiniLM	1			
ELECTRA	1			
ALBERT	1			
RepresentThemAll	1			
seBERT	1			
Code-davinci	1			
Curie	1			
Davinci	1			
T5-learning	1			
CodeGPT	1			
JavaBERT	1			
DOBF	1			
CuBERT	1			
ProphetNet-Code	1			
SPT-Code	1			
CoTexT	1			
C-BERT	1			
GPT-C CugLM	1 1			
TreeBERT	1			
GPT-2	1			
SynCoBERT	1			
DeepDebug	1			
UniXcoder	1			
GPT-NeoX	1			
PLBart	1			
GPT-3	1			
CodeParrot	1			
Copilot	1			
ATLAS	1			
CoCoNut	1			
Hoppity	1			
Sequencer	1			
TFix	1			
CEDAR	1			
Transformers	1			
TransCoder	1			
Transcoder*	1			
SDA-Trans	1			
BERTOverflow	1			
Name: count, dtype:	int	54		

2024 - Number of P	apers	per	Model	Family:
model_families_lis	t			
CodeBERT	11			
GPT-3.5	9			
GPT-4	8			
CodeT5	8			
ChatGPT	8			
CodeGen	8			
GraphCodeBERT	7			
UniXcoder	5			
BERT	5			
Codex	4			
RoBERTa	4			
InCoder	4			
StarCoder	4			
Llama	3			
ChatGLM	3			
UnixCoder	3			
PLBART	3			
T5	3			
GPT-3	3			
GPT-2	3			
text-davinci	3			
PolyCoder	3			
CodeGeeX	2			
CodeParrot	2			
	2			
UniLog	2			
GPT-NEO				
Copilot	2			
Vicuna	2			
Pythia	2			
CodeGPT	2			
Airboros	1			
code-davinci	1			
SantaCoder	1			
WizardCoder	1			
Sentence-BERT	1			
GPTBigCode	1			
seBERT	1			
VulBERTa	1			
PDBERT	1			
ALBERT	1			
KeyBERT	1			
LSTM	1			
TFix	1			
DISCO	1			
text-embedding	1			
Flan	1			
CodeBERTa	1			
SynCoBERT	1			
CodeLlama	1			
Transformer	1			
GIN	1			
flan-alpaca	1			
UnifiedQA	1			
GrammarT5	1			
GPT-C	1			
VRepair	1			
CodeReviewer	1			
VulMaster	1			

VulRepair	1
PanguCoder	1
TypeFix	1
CoTexT	1
CodeT5+	1
AlphaRepair	1
PyTER	1
CoCoNuT	1
LANCE	1
IRGen	1
Deep-SE	1
FAIR	1
OSCAR	1
sentenceBERT	1
GPT2SP	1
CodeGen-Mono	1
CodeGen-NL	1
CodeGen-Multi	1
CodeGen2	1
PyCodeGPT	1
GPT-Code-Clippy	1
Name: count dtyne:	int

Name: count, dtype: int64

2025 - Number of Papers per	Model Family:
model_families_list	-
GPT-4	39
GPT-3.5	35
CodeLlama	25
Llama	14
StarCoder	14
CodeBERT	12
DeepSeekCoder	12
CodeGen	9
Claude	7
Gemini	7
CodeT5	7
GraphCodeBERT	7
ChatGPT	6
CodeQwen	6
DeepSeek	6
UniXcoder	5
UnixCoder	3
LineVul	3
WizardCoder	3
ChatGLM	3
Mistral	3
Gemma	3
RoBERTa	3
InCoder	3
Codestral	2
Qwen	2
CodeT5+	2
CodeGemma	2
GPT3.5	2
OpenDevin	2
CodeGPT	2
Incoder	2
text-embedding	2
Phi	2
GPT-3	2
AthenTest	1
ChatDev	1
SVulD	1
Code-davinci	1
Multi-Turn Program Synthesis	
GPT4	1
AgentCoder	1
Self-collaboration	1
Copilot	1
AutoGPT	1
MetaGPT	1
Magicoder	1
CoditT5	1
Unixcoder	1
Longformer	1
Shipwright	1
MagiCoder	1
TransCoder	1
Parfum	1
UniTrans	1
SantaCoder	1
ALL-MINILM-L6-V210	1
CodeBert	1

TOGA	1
Poro	1
BERT	1
SEQ Graph& HYBRID	1
Devign	1
•	
VULGEN	1
COME	1
CCT5	1
NNGen	1
ReVeal	1
Mixtral	1
VGX	1
LILAC	1
LLM-Parser	1
	1
Starcoder	
GPTSniffer	1
FitRepair	1
ChatRepair	1
RAP-Gen	1
AlphaRe-pair	1
GLTR	1
Sapling	1
OpenCodeInterpreter	1
Stable-Code	1
BigBird	1
ContraBERT	1
CuBERT	1
VulBERTa	1
ChatUniTest	1
	1
TestGen-LLM	
RustAssistant	1
Sonnet	1
GPTZero	1
GPT-2 Output Detector	1
DetectGPT	1
	_
Repilot	1
Aider	1
AppMap Naive	1
AutoCodeRover	1
SWE-Agent	1
Vercel	1
Agentless	1
TitanFuzz	1
Moatless Tools	1
HedgeCode	1
GPT-4, CodeBERT	1
Codex	1
SYNCOBERT	1
CoCoSoDa	1
CodeRetriever	1
LLama	1
FuzzGPT	1
CodeShell	1
PLBART	1
Pythia	1
OPT	1
Exlong	1
DeBERTa	1
CodeParrot	1
CAT-LM	1
	_

PolyCoder	1
StarChat	1
PaLM	1
Vicuna	1
Guanaco	1
Tulu	1
GPT3-5	1
StableCode	1
Name: count, dtype:	int64

## **Targeted Programming Languages**

## Which Programming Languages are Evaluated

Overall - Number of Papers per Programming Language:

```
programming_languages_list
Java
        67
Python
              56
C
               30
C++
               23
JavaScript
               13
PHP
               10
Go
              9
NM
              8
                7
Rust
             7
Ruby
C#
               5
             3
Kotlin
SQL
                3
R
TypeScript 2
Haskell 2
Objective-C 2
Scala 2
Swift
              2
Prolog
Erlang
              1
Solidity
Bash
               1
CSharp
                1
                1
Perl
                1
SCRATCH
HTML
```

Name: count, dtype: int64

```
2023 - Number of Papers per Programming Language:
programming_languages_list
Java
              14
Python
C
              7
JavaScript
              4
PHP
               4
C++
               2
               2
NM
Ruby
               2
               2
Go
SCRATCH
C#
               1
Name: count, dtype: int64
2024 - Number of Papers per Programming Language:
programming_languages_list
Java
             22
Python
              18
C
              11
C++
              9
               4
Go
Ruby
               3
JavaScript
PHP
               3
Rust
               3
               3
C#
Kotlin
               3
               2
Scala
SQL
               2
NM
               1
```

Name: count, dtype: int64

1

1

1

1

1

1

1

CSharp

Bash

Perl

R

Swift

Solidity

TypeScript

Objective-C 1

```
2025 - Number of Papers per Programming Language:
programming_languages_list
Java
               31
Python
               29
C
               12
C++
               12
JavaScript
                6
MM
                4
Rust
Go
PHP
                3
                2
Ruby
Haskell
                2
C#
                1
TypeScript
Objective-C
Swift
                1
Erlang
                1
Prolog
                1
SQL
HTML
                1
Name: count, dtype: int64
```

#### Number of Programming Languages Evaluated per Paper

```
Overall - Distribution of Number of Programming Languages per Paper:
programming_languages_list
1
      81
      48
0
2
      24
3
      10
6
       4
4
       3
10
       2
5
       2
16
       1
7
13
       1
Name: count, dtype: int64
Papers covering multiple programming languages: 48
Percentage covering multiple programming languages: 27.1%
2023 - Distribution of Number of Programming Languages per Paper:
programming_languages_list
1
     13
0
2
      5
4
      1
6
      1
7
      1
5
      1
3
      1
Name: count, dtype: int64
Papers covering multiple programming languages: 10
```

Percentage covering multiple programming languages: 31.2%

```
2024 - Distribution of Number of Programming Languages per Paper:
programming_languages_list
      22
0
      15
2
      10
       3
5
       1
6
4
16
       1
13
       1
Name: count, dtype: int64
Papers covering multiple programming languages: 18
Percentage covering multiple programming languages: 32.7%
2025 - Distribution of Number of Programming Languages per Paper:
programming_languages_list
      46
1
0
      24
2
       9
6
       2
       1
Name: count, dtype: int64
Papers covering multiple programming languages: 20
Percentage covering multiple programming languages: 22.2%
```

# RQ2 - How well do authors tackle the problem of data leakage/contamination?

```
Overall - Contamination reported in 58 out of 177 papers (32.8%)

2023 - Contamination reported in 6 out of 32 papers (18.8%)

2024 - Contamination reported in 14 out of 55 papers (25.5%)

2025 - Contamination reported in 38 out of 90 papers (42.2%)
```

# RQ3 - How replicable are LLM-based studies?

## **Model Configuration Reporting**

```
Overall - Number of Papers Reporting on Inference (Generation) Configuration/Parameters: 89 out of 177 papers (50.3%)

2023 - Number of Papers Reporting on Inference (Generation) Configuration/Parameters: 10 o ut of 32 papers (31.2%)

2024 - Number of Papers Reporting on Inference (Generation) Configuration/Parameters: 27 o ut of 55 papers (49.1%)

2025 - Number of Papers Reporting on Inference (Generation) Configuration/Parameters: 52 o ut of 90 papers (57.8%)
```

## **Artefact Availability**

#### **Artifact Badges**

```
Relevant Papers with Artifact Available Badge: 33 out of 177 papers (18.6%)
Non-Relevant Papers with Artifact Available Badge: 213 out of 515 papers (41.4%)
Relevant Papers with Artifact Reusable Badge: 21 out of 177 papers (11.9%)
Non-Relevant Papers with Artifact Reusable Badge: 150 out of 515 papers (29.1%)
Relevant Papers with Artifact Functional Badge: 13 out of 177 papers (7.3%)
Non-Relevant Papers with Artifact Functional Badge: 70 out of 515 papers (13.6%)

2023 - Artifact Available Badge for 6 out of 32 Relevant Papers (18.8%)
2023 - Artifact Reusable Badge for 3 out of 32 Relevant Papers (9.4%)
2023 - Artifact Functional Badge for 1 out of 32 Relevant Papers (3.1%)

2024 - Artifact Available Badge for 9 out of 55 Relevant Papers (16.4%)
2024 - Artifact Reusable Badge for 8 out of 55 Relevant Papers (14.5%)
2024 - Artifact Functional Badge for 0 out of 55 Relevant Papers (0.0%)

2025 - Artifact Available Badge for 18 out of 90 Relevant Papers (11.1%)
2025 - Artifact Functional Badge for 10 out of 90 Relevant Papers (11.1%)
2025 - Artifact Functional Badge for 12 out of 90 Relevant Papers (13.3%)
```

#### Manual Artefact Evaluation

```
Overall - Number of Papers with Artefacts (Manual Check): 144 out of 177 (81.4%)
Overall - Number of Papers with no Artefact (Manual Check): 24 out of 177 (13.6%)
Overall - Number of Papers with Dead Links (Manual Check): 9 out of 177 (5.1%)
```

#### Number of Papers with Artifact Badges that have Dead Links

```
Number of papers with Artifact Available Badges but DEAD links: 2
Number of papers with Artifact Reusable Badges but DEAD links: 1
Number of papers with Artifact Functional Badges but DEAD links: 0
```

# RQ4 - What are the costs of LLM-based SE research?

Frequency of Different Costs Reported Across All Papers: cost\_list gpu 78 68 time 48 18 13 money content 10 hw memory 6 invocations 3 tpu 2 tokens 1 operations 1

Name: count, dtype: int64

Number of Papers Reporting both Time and Hardware: 36