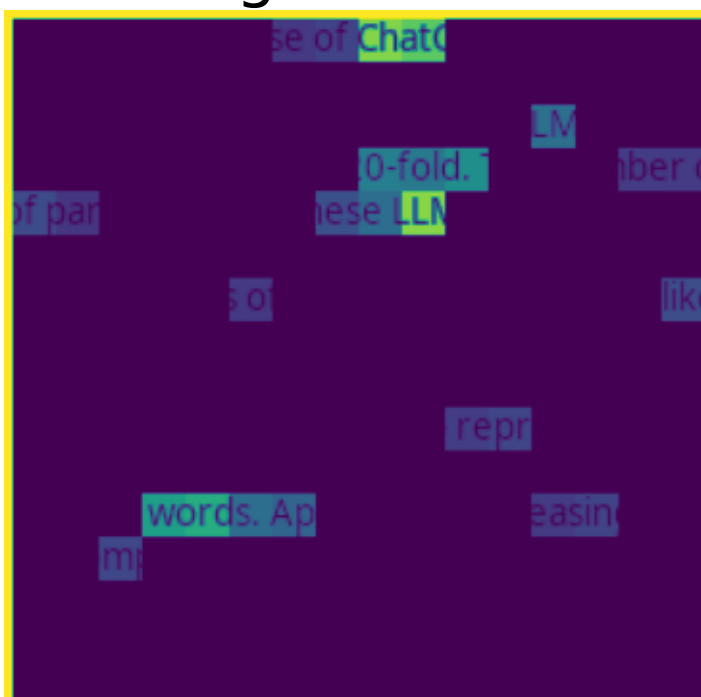


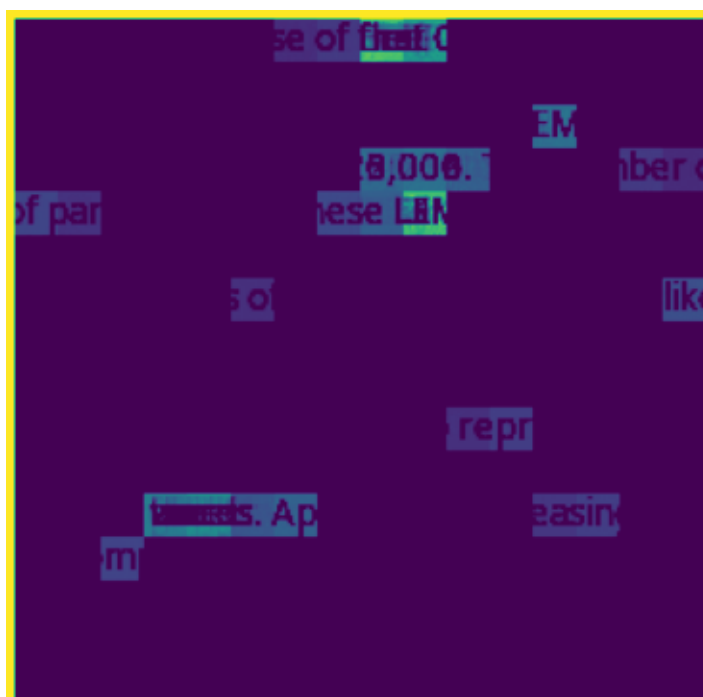
Original

After the release of ChatGPT in 2022, the number of papers published every day about Large Language Models (LLMs) has increased more than 20-fold. The number of parameters in these LLMs jumped from 340 millions in implementations such as BERT to billions of parameters in models like GPT-3 or LLaMA. A large part of these parameters come from the word-embedding layers which are used to represent a finite vocabulary of characters, sets of characters or words. Apart from increasing model complexity, a fixed vocabulary is also responsible for brittle models, which cannot deal with out-of-vocabulary inputs and cannot generalize to new languages. As a

Original + SD



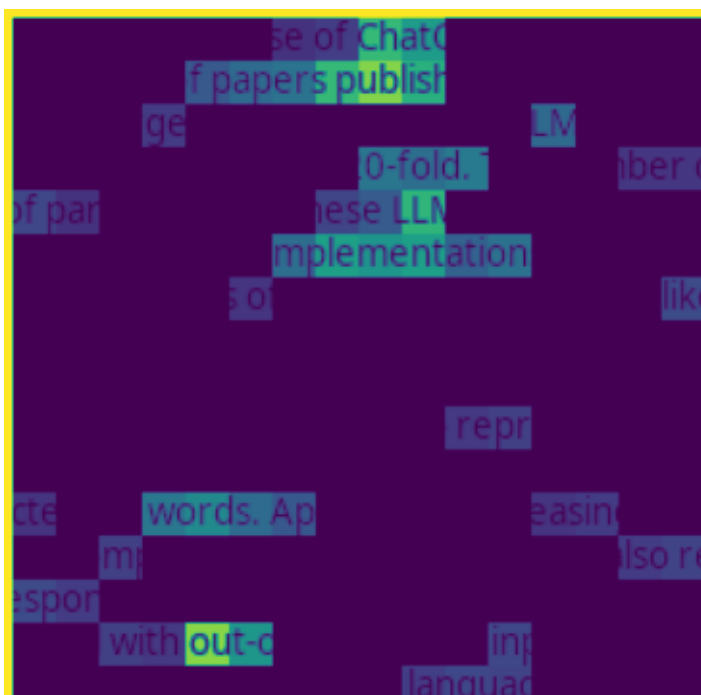
Mean Predictions + SD



250

mask ratio = 10%

After the release of ChatGPT in 2022, the number of papers published every day about Large Language Models (LLMs) has increased more than 20-fold. The number of parameters in these LLMs jumped from 340 millions in implementations such as BERT to billions of parameters in models like GPT-3 or LLaMA. A large part of these parameters come from the word-embedding layers which are used to represent a finite vocabulary of characters, sets of characters or words. Apart from increasing model complexity, a fixed vocabulary is also responsible for brittle models, which cannot deal with out-of-vocabulary inputs and cannot generalize to new languages. As a

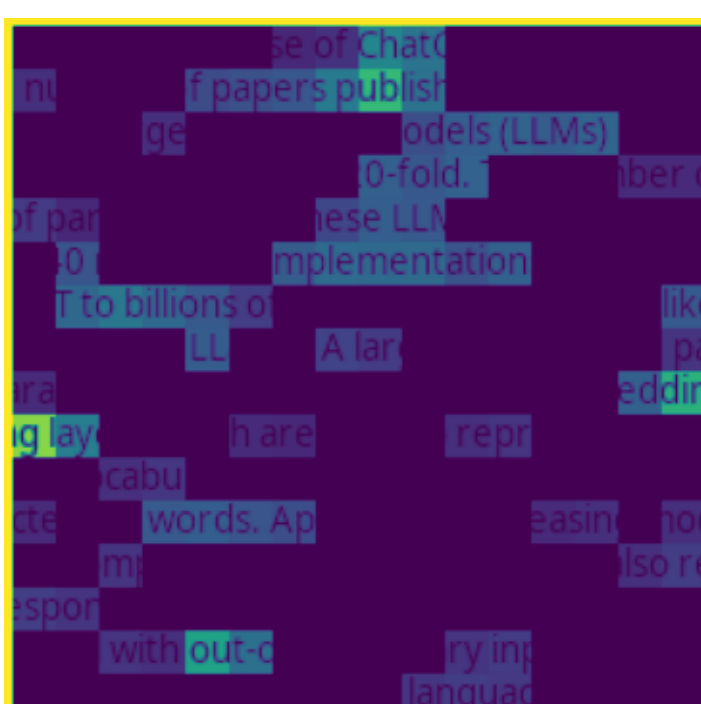


se of that C
 f b establish
 ge EM
 0,000. nber c
 of par these LEM
 nthe establish
 so nlike
 repr
 cte wards. Ap measin
 m also re
 respon
 with link c inp
 languag

- 225

mask ratio = 20%

After the release of ChatGPT in 2022, the number of papers published every day about Large Language Models (LLMs) has increased more than 20-fold. The number of parameters in these LLMs jumped from 340 millions in implementations such as BERT to billions of parameters in models like GPT-3 or LLaMA. A large part of these parameters come from the word-embedding layers which are used to represent a finite vocabulary of characters, sets of characters or words. Apart from increasing model complexity, a fixed vocabulary is also responsible for brittle models, which cannot deal with out-of-vocabulary inputs and cannot generalize to new languages. As a

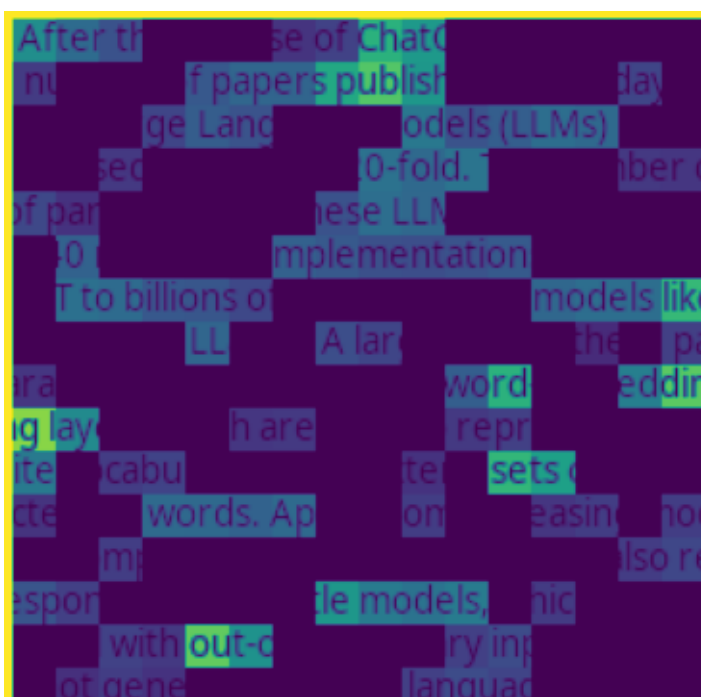


ni
 ge
 of par
 8
 T
 ara
 d
 cte
 m
 respon
 with
 se of that C
 of
 edab
 5,000.
 ber o
 here
 n
 o
 A lar
 h are
 repr
 cabu
 s. Ap
 easing
 no
 also re
 ry in
 langua

200

mask ratio = 30%

After the release of ChatGPT in 2022, the number of papers published every day about Large Language Models (LLMs) has increased more than 20-fold. The number of parameters in these LLMs jumped from 340 millions in implementations such as BERT to billions of parameters in models like GPT-3 or LLaMA. A large part of these parameters come from the word-embedding layers which are used to represent a finite vocabulary of characters, sets of characters or words. Apart from increasing model complexity, a fixed vocabulary is also responsible for brittle models, which cannot deal with out-of-vocabulary inputs and cannot generalize to new languages. As a



After the release of that O
 nline if I publish today
 ge Land B
 sed \$8,000. Number o
 of part these M
 T B o
 A lar the pa
 ara word the title
 ch are repr
 ite cabu ite
 cte s. Ap om easing no
 m
 respon der nic
 with t c ry in
 ot gene langua

175

mask ratio = 40%

After the release of ChatGPT in 2022, the number of papers published every day about Large Language Models (LLMs) has increased more than 20-fold. The number of parameters in these LLMs jumped from 340 millions in implementations such as BERT to billions of parameters in models like GPT-3 or LLaMA. A large part of these parameters come from the word-embedding layers which are used to represent a finite vocabulary of characters, sets of characters or words. Apart from increasing model complexity, a fixed vocabulary is also responsible for brittle models, which cannot deal with out-of-vocabulary inputs and cannot generalize to new languages. As a



After the release of the 2000 Census, the Census Bureau published a series of reports on the language spoken at home by the population. The reports show that the number of people who speak a language other than English at home has increased significantly since 1990. The most common languages spoken at home are Spanish, Chinese, and Vietnamese. The reports also show that the number of people who speak a language other than English at home has increased significantly since 1990. The most common languages spoken at home are Spanish, Chinese, and Vietnamese. The reports also show that the number of people who speak a language other than English at home has increased significantly since 1990. The most common languages spoken at home are Spanish, Chinese, and Vietnamese.

150

mask ratio = 50%

After the release of ChatGPT in 2022, the number of papers published every day about Large Language Models (LLMs) has increased more than 20-fold. The number of parameters in these LLMs jumped from 340 millions in implementations such as BERT to billions of parameters in models like GPT-3 or LLaMA. A large part of these parameters come from the word-embedding layers which are used to represent a finite vocabulary of characters, sets of characters or words. Apart from increasing model complexity, a fixed vocabulary is also responsible for brittle models, which cannot deal with out-of-vocabulary inputs and cannot generalize to new languages. As a



After the release of *Frank Ocean* in 2010, the initial response was a mix of praise and criticism. The album's genre-blending sound, which fused R&B, soul, and hip-hop, was a departure from the traditional R&B sound of the time. Some critics praised the album's emotional depth and Ocean's vocal range, while others criticized its experimental sound and lack of commercial appeal. However, over time, the album's reputation grew, and it is now considered one of the greatest R&B albums of all time. The album's success was a testament to Ocean's talent and the power of authentic storytelling in music.

125

mask ratio = 60%

After the release of ChatGPT in 2022, the number of papers published every day about Large Language Models (LLMs) has increased more than 20-fold. The number of parameters in these LLMs jumped from 340 millions in implementations such as BERT to billions of parameters in models like GPT-3 or LLaMA. A large part of these parameters come from the word-embedding layers which are used to represent a finite vocabulary of characters, sets of characters or words. Apart from increasing model complexity, a fixed vocabulary is also responsible for brittle models, which cannot deal with out-of-vocabulary inputs and cannot generalize to new languages. As a

[illegible]

100

mask ratio = 70%

After the release of ChatGPT in 2022, the number of papers published every day about Large Language Models (LLMs) has increased more than 20-fold. The number of parameters in these LLMs jumped from 340 millions in implementations such as BERT to billions of parameters in models like GPT-3 or LLaMA. A large part of these parameters come from the word-embedding layers which are used to represent a finite vocabulary of characters, sets of characters or words. Apart from increasing model complexity, a fixed vocabulary is also responsible for brittle models, which cannot deal with out-of-vocabulary inputs and cannot generalize to new languages. As a



1. **Introduction**
 2. **Background**
 3. **Methodology**
 4. **Results**
 5. **Discussion**
 6. **Conclusion**
 7. **References**
 8. **Appendix**
 9. **Notes**
 10. **Tables**
 11. **Figures**
 12. **Tables**
 13. **Figures**
 14. **Tables**
 15. **Figures**
 16. **Tables**
 17. **Figures**
 18. **Tables**
 19. **Figures**
 20. **Tables**
 21. **Figures**
 22. **Tables**
 23. **Figures**
 24. **Tables**
 25. **Figures**
 26. **Tables**
 27. **Figures**
 28. **Tables**
 29. **Figures**
 30. **Tables**
 31. **Figures**
 32. **Tables**
 33. **Figures**
 34. **Tables**
 35. **Figures**
 36. **Tables**
 37. **Figures**
 38. **Tables**
 39. **Figures**
 40. **Tables**
 41. **Figures**
 42. **Tables**
 43. **Figures**
 44. **Tables**
 45. **Figures**
 46. **Tables**
 47. **Figures**
 48. **Tables**
 49. **Figures**
 50. **Tables**
 51. **Figures**
 52. **Tables**
 53. **Figures**
 54. **Tables**
 55. **Figures**
 56. **Tables**
 57. **Figures**
 58. **Tables**
 59. **Figures**
 60. **Tables**
 61. **Figures**
 62. **Tables**
 63. **Figures**
 64. **Tables**
 65. **Figures**
 66. **Tables**
 67. **Figures**
 68. **Tables**
 69. **Figures**
 70. **Tables**
 71. **Figures**
 72. **Tables**
 73. **Figures**
 74. **Tables**
 75. **Figures**
 76. **Tables**
 77. **Figures**
 78. **Tables**
 79. **Figures**
 80. **Tables**
 81. **Figures**
 82. **Tables**
 83. **Figures**
 84. **Tables**
 85. **Figures**
 86. **Tables**
 87. **Figures**
 88. **Tables**
 89. **Figures**
 90. **Tables**
 91. **Figures**
 92. **Tables**
 93. **Figures**
 94. **Tables**
 95. **Figures**
 96. **Tables**
 97. **Figures**
 98. **Tables**
 99. **Figures**
 100. **Tables**
 101. **Figures**
 102. **Tables**
 103. **Figures**
 104. **Tables**
 105. **Figures**
 106. **Tables**
 107. **Figures**
 108. **Tables**
 109. **Figures**
 110. **Tables**
 111. **Figures**
 112. **Tables**
 113. **Figures**
 114. **Tables**
 115. **Figures**
 116. **Tables**
 117. **Figures**
 118. **Tables**
 119. **Figures**
 120. **Tables**
 121. **Figures**
 122. **Tables**
 123. **Figures**
 124. **Tables**
 125. **Figures**
 126. **Tables**
 127. **Figures**
 128. **Tables**
 129. **Figures**
 130. **Tables**
 131. **Figures**
 132. **Tables**
 133. **Figures**
 134. **Tables**
 135. **Figures**
 136. **Tables**
 137. **Figures**
 138. **Tables**
 139. **Figures**
 140. **Tables**
 141. **Figures**
 142. **Tables**
 143. **Figures**
 144. **Tables**
 145. **Figures**
 146. **Tables**
 147. **Figures**
 148. **Tables**
 149. **Figures**
 150. **Tables**
 151. **Figures**
 152. **Tables**
 153. **Figures**
 154. **Tables**
 155. **Figures**
 156. **Tables**
 157. **Figures**
 158. **Tables**
 159. **Figures**
 160. **Tables**
 161. **Figures**
 162. **Tables**
 163. **Figures**
 164. **Tables**
 165. **Figures**
 166. **Tables**
 167. **Figures**
 168. **Tables**
 169. **Figures**
 170. **Tables**
 171. **Figures**
 172. **Tables**
 173. **Figures**
 174. **Tables**
 175. **Figures**
 176. **Tables**
 177. **Figures**
 178. **Tables**
 179. **Figures**
 180. **Tables**
 181. **Figures**
 182. **Tables**
 183. **Figures**
 184. **Tables**
 185. **Figures**
 186. **Tables**
 187. **Figures**
 188. **Tables**
 189. **Figures**
 190. **Tables**
 191. **Figures**
 192. **Tables**
 193. **Figures**
 194. **Tables**
 195. **Figures**
 196. **Tables**
 197. **Figures**
 198. **Tables**
 199. **Figures**
 200. **Tables**
 201. **Figures**
 202. **Tables**
 203. **Figures**
 204. **Tables**
 205. **Figures**
 206. **Tables**
 207. **Figures**
 208. **Tables**
 209. **Figures**
 210. **Tables**
 211. **Figures**
 212. **Tables**
 213. **Figures**
 214. **Tables**
 215. **Figures**
 216. **Tables**
 217. **Figures**
 218. **Tables**
 219. **Figures**
 220. **Tables**
 221. **Figures**
 222. **Tables**
 223. **Figures**
 224. **Tables**
 225. **Figures**
 226. **Tables**
 227. **Figures**
 228. **Tables**
 229. **Figures**
 230. **Tables**
 231. **Figures**
 232. **Tables**
 233. **Figures**
 234. **Tables**
 235. **Figures**
 236. **Tables**
 237. **Figures**
 238. **Tables**
 239. **Figures**
 240. **Tables**
 241. **Figures**
 242. **Tables**
 243. **Figures**
 244. **Tables**
 245. **Figures**
 246. **Tables**
 247. **Figures**
 248. **Tables**
 249. **Figures**
 250. **Tables**
 251. **Figures**
 252. **Tables**
 253. **Figures**
 254. **Tables**
 255. **Figures**

75

mask ratio = 80%

After the release of ChatGPT in 2022, the number of papers published every day about Large Language Models (LLMs) has increased more than 20-fold. The number of parameters in these LLMs jumped from 340 millions in implementations such as BERT to billions of parameters in models like GPT-3 or LLaMA. A large part of these parameters come from the word-embedding layers which are used to represent a finite vocabulary of characters, sets of characters or words. Apart from increasing model complexity, a fixed vocabulary is also responsible for brittle models, which cannot deal with out-of-vocabulary inputs and cannot generalize to new languages. As a



F 50