Starcoder testing:

Demonstration of Starcoder2’s code generation ability:

Model comes from Replicate: cjwbw/starcoder2-15b (this might be a fine-tuned version of Starcoder2, I wasn’t able to find a good api platform that runs Starcoder2)

Asking the LLM to solve a Leetcode medium level question:

Prompt:

| Given a string s, return the longest palindromic substring in s.  Example 1:  Input: s = "babad"  Output: "bab"  Explanation: "aba" is also a valid answer.  Example 2:  Input: s = "cbbd"  Output: "bb"  example 3: "abcda"  output = "a" |
| --- |

The code generated by Starcoder2 passed the leetcode tests

图形用户界面, 文本

描述已自动生成

Now we want to test the LLM given Leetcode hard level questions:

Prompt:

| Given two sorted arrays nums1 and nums2 of size m and n respectively, return the median of the two sorted arrays.  The overall run time complexity should be O(log (m+n)).    Example 1:  Input: nums1 = [1,3], nums2 = [2]  Output: 2.00000  Explanation: merged array = [1,2,3] and median is 2.  Example 2:  Input: nums1 = [1,2], nums2 = [3,4]  Output: 2.50000  Explanation: merged array = [1,2,3,4] and median is (2 + 3) / 2 = 2.5.    Constraints:  nums1.length == m  nums2.length == n  0 <= m <= 1000  0 <= n <= 1000  1 <= m + n <= 2000  -106 <= nums1[i], nums2[i] <= 106 |
| --- |

The code generated by Starcoder2 passed the leetcode tests

图形用户界面, 文本

描述已自动生成

Prompt:

| Given a rows x cols binary matrix filled with 0's and 1's, find the largest rectangle containing only 1's and return its area.  Example 1:  Input: matrix = [["1","0","1","0","0"],["1","0","1","1","1"],["1","1","1","1","1"],["1","0","0","1","0"]]  Output: 6  Explanation: The maximal rectangle is shown in the above picture.  Example 2:  Input: matrix = [["0"]]  Output: 0  Example 3:  Input: matrix = [["1"]]  Output: 1 |
| --- |

The code generated by Starcoder2 passed the leetcode tests

文本

描述已自动生成

Now we want to see if the model can fit into our approach of code generation:

1. Llam2-70B will first find the device parameters in the database including a default function call (ex. c = gf.components.mmi1x2(width\_taper=1.0, length\_taper=10.0, length\_mmi=5.5, width\_mmi=2.5, gap\_mmi=0.25, width=0.0, cross\_section='xs\_sc') )
2. Starcoder2 will replace the parameters (ex. Width\_taper, Length\_taper, etc.) in the default function call with the numbers for that specific device.

Prompt tested:

| Layout a 1x2 MMI that operates at a wavelength of 1550nm. |
| --- |

Result:

* It generated 40 times and all of them matched what was expected.
* the format of the code doesn’t match the sample output since in the prompt we gave to the llm, we asked it to follow the format of a sample output, we also told the llm that the second line of the sample output should be replaced by the llm.

sample output:

| import gdsfactory as gf  c = gf.components.wire\_straight()  c.plot() |
| --- |

Output from Starcoder:

|  |
| --- |

The llm generated this last part numerous times and didn’t seem to stop, so I had to manually terminate the output. In the end it generated 40 times and all of them are correct, however the model didn’t seem to follow the format of the output. Similar things happened when I tested it with leetcode questions.

Next steps:

1. Figure out why the model outputs have wrong format.
2. Try to run the model locally and compare the results.