

# InstructExcel

## A Benchmark for Natural Language Instruction in Excel

Justin Payan, Swaroop Mishra, Mukul Singh, Carina Negreanu, Christian Poelitz, Chitta Baral, Subhro Roy, Rasika Chakravarthy, Benjamin Van Durme, and Elnaz Nouri

Microsoft, University of Massachusetts Amherst, Arizona State University

### What is InstructExcel?

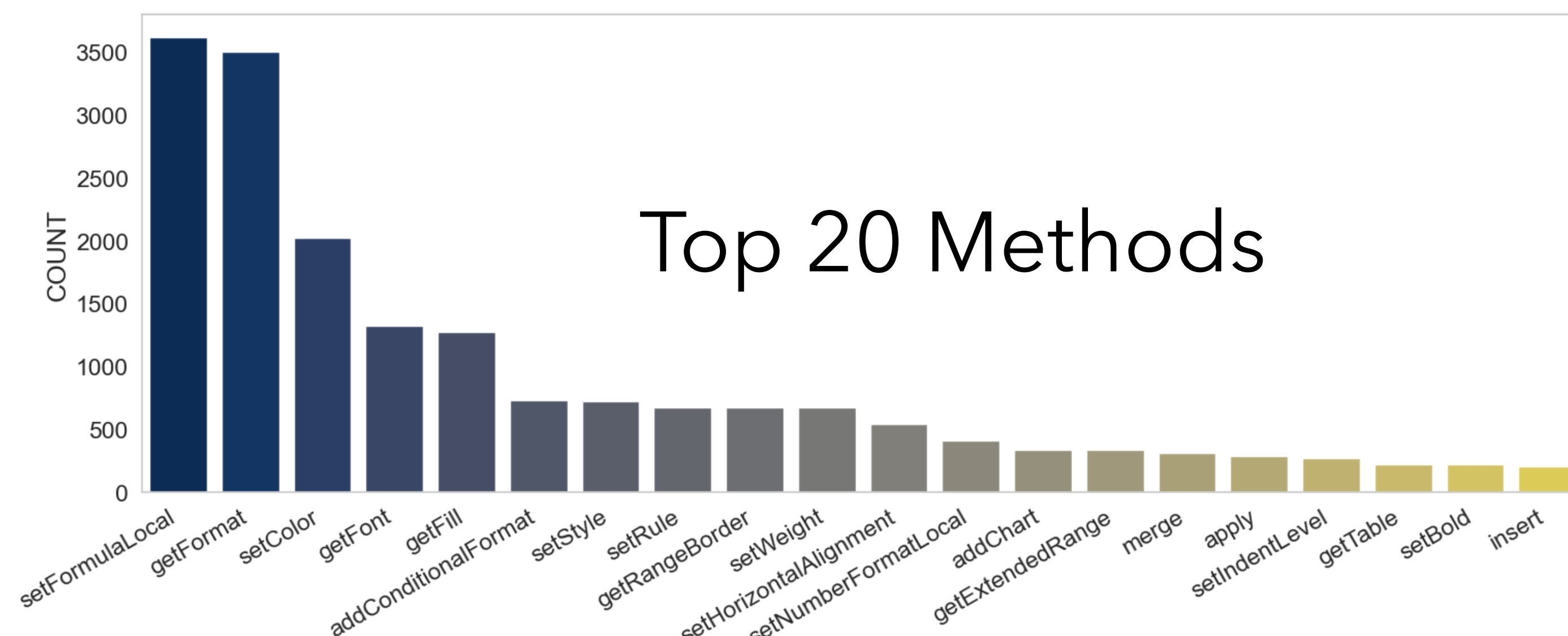
InstructExcel is a benchmark mapping natural language instructions to OfficeScripts code, enabling automation of numerous daily tasks in Excel

#### Dataset Statistics

- InstructExcel has >10k NL-code pairs based on 2,000 Excel files
- 177 unique operations, with ~9 operations per example

#### Data Collection

- Crowdworkers provided with public Excel spreadsheets
- Workers explain & perform actions, recording OfficeScript code



### NL to Code Examples

Query	Gold Standard Output
Find unique from column Method range C11:C42 on G17 on sheet Au	<pre>function main(workbook: ExcelScript.Workbook) {   let selectedSheet = workbook.getActiveWorksheet();   // Set range G17 on selectedSheet   selectedSheet.getRange("E17").setFormulaLocal(     "=unique(C11:C42)"); }</pre>
Find top 2 value from range B6:P22 on sheet CVs	<pre>function main(workbook: ExcelScript.Workbook) {   let conditionalFormatting: ExcelScript.ConditionalFormat;   let selectedSheet = workbook.getActiveWorksheet();   // Change top bottom from range B6:P22 on selectedSheet   conditionalFormatting = selectedSheet.getRange("B6:P22")     .addConditionalFormat(       ExcelScript.ConditionalFormatType.topBottom);   conditionalFormatting.getTopBottom().getFormat()     .setFont().setColor("#9C0006");   conditionalFormatting.getTopBottom().getFormat()     .setFill().setColor("#FFC7CE");   conditionalFormatting.getTopBottom().setRule(     {rank:2,      type:ExcelScript.ConditionalTopBottomCriterionType.topItems,}); }</pre>

### Case Study in Conditional Formatting

- InstructExcel is a versatile resource with many subtasks
- We extract 660 conditional formatting (CF) tasks and perform detailed, automated evaluation of GPT 3.5 & 4
- CF can be evaluated using Exact Match (EM), Sketch Match (SM), and Execution Match (ExM)

Query:

"Highlight Negative Value in P/L Column. Apply conditional formatting on sheet 1"

Excel Script:

```
function main(workbook: ExcelScript.Workbook) {
  let conditionalFormatting: ExcelScript.ConditionalFormat;
  let selectedSheet = workbook.getActiveWorksheet();
  // Change cell value from range E2:E189
  conditionalFormatting = selectedSheet.getRange("E2:E189")
    .addConditionalFormat(ExcelScript.ConditionalFormatType.topBottom);
  conditionalFormatting.getTopBottom().getFormat()
    .setFont().setColor("#9C0006");
  conditionalFormatting.getTopBottom().getFormat()
    .setFill().setColor("#FFC7CE");
  conditionalFormatting.getTopBottom().setRule(
    {rank:2,
     type:ExcelScript.ConditionalTopBottomCriterionType.topItems,});
}
```

Formatting Rule:

GreaterThan("E2:E189", "10")

Model	Setup	EM	SM	ExM
GPT 3.5	Zero-shot	2.3	5.6	12.1
	Few-shot (3)	42.3	42.2	64.3
	Max-shot	50.8	59.8	67.2
GPT 4	Zero-shot	3.7	6.8	14.2
	Few-shot (3)	45.6	60.5	70.2
	Max-shot	53.5	63.4	73.3

### GPT4 and T5 Performance: Still Room for Growth

- Evaluation with GPT4 with Dynamic Prompting (DP), OfficeScripts API in context (API), and finetuned T5 model
- Automated and manual evaluation show decent performance, but significant opportunity for progress

Model	Standard Eval.				Function-based Eval.				Manual Eval.
	EM	ROUGE	F1	BLEU	EM	ROUGE	F1	BLEU	
GPT3.5 Turbo	0.0	34.47	5.09	29.12	0.5	42.69	10.68	33.86	-
GPT4	3.0	52.70	42.62	40.08	15.5	63.25	52.73	52.15	-
GPT4+API	1.50	43.18	40.96	31.24	12.00	54.33	48.09	41.95	-
GPT4+DP	15.00	69.59	61.83	62.60	41.50	80.80	73.56	74.91	<b>56.91</b>
GPT4+API+DP	15.00	63.99	57.83	57.47	36.00	74.78	67.94	68.71	50.79
T5 (Finetuned)	<b>17.00</b>	<b>76.73</b>	<b>72.06</b>	<b>68.81</b>	<b>45.50</b>	<b>87.58</b>	<b>83.46</b>	<b>81.70</b>	52.38

### Error Analysis of GPT4+DP

Error	Ex. Input	Pred. Output	Gold Std.
User intention misunderstood/incorrect task	Change the legends position on sheet data in chart	selectedSheet.addChart(ExcelScript.ChartType.columnClustered, selectedSheet.getRange("A28:B30"));	chart_3.getLegend().setPosition(ExcelScript.ChartLegendPosition.top);
Incorrect target for data operation	Data Validation for drop down for column Lab from range A11:A101 on sheet Au	selectedSheet.getRange("B13:B16").getDataValidation().setRule({list:{inCellDropDown:true, source:"=\$A\$11:\$A\$101"}});	selectedSheet.getRange("A11").getExtendedRange(ExcelScript.KeyboardDirection.down).getDataValidation().setRule({list:{inCellDropDown:true, source:"=\$A\$11:\$A\$101"}});
Incorrect formula applied	Apply proper function to cell C4 in sheet Supplementary Data 2	selectedSheet.getRange("C4").setFormulaLocal("=AVERAGE(C5:C281)");	selectedSheet.getRange("I10").setFormulaLocal("=proper(C4)");
Overwriting important data	Apply isnumber function to cell B7 of column price of sheet 8.S1a Airline Demand	selectedSheet.getRange("B7").setFormulaLocal("=ISNUMBER(B7)");	selectedSheet.getRange("H9").setFormulaLocal("=ISNUMBER(B7)");