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

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)

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

Dataset Statistics

Data Collection
- Crowworkers provided with public Excel spreadsheets
- Workers explain & perform actions, recording OfficeScript code

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)

Case Study in Conditional Formatting

NL to Code Examples

<table>
<thead>
<tr>
<th>Query</th>
<th>Gold Standard Output</th>
</tr>
</thead>
</table>
| Find unique from column C11:C42 on G17 on sheet A1 | Function main(Workbook: ExcelWorkbook) {
  let selectedSheet = workbook.getActiveWorkbook();
  // Set range G17 on selectedSheet, getRange("F*?"), setFormulaLocal("unique(C11:C42")
} |

<table>
<thead>
<tr>
<th>Query</th>
<th>Gold Standard Output</th>
</tr>
</thead>
</table>
| Find top 2 value from range B16:P22 on sheet CVs | Function main(Workbook: ExcelWorkbook) {
  let selectedSheet = workbook.getActiveWorkbook();
  // Create top 2 values from range B16:P22 on selectedSheet
  let formulaLocal = "=index(B16:P22,1)"
  let formulaLocal = "=index(B16:P22,2)"
} |

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

Error Analysis of GPT4+DP

<table>
<thead>
<tr>
<th>Error</th>
<th>EM</th>
<th>SM</th>
<th>ExM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use instruction understood/unexpected task</td>
<td>Change the legend position of sheet data in chart</td>
<td>selectedSheet.getChart()</td>
<td>getChartLegend().setTitle(&quot;New Title&quot;)</td>
</tr>
<tr>
<td>Incorrect target for data operation</td>
<td>Data Values for drop down for column Lab from range AI1:A111 on sheet A1</td>
<td>selectedSheet.getDropDown().setSource(&quot;AI1:A111&quot;)</td>
<td></td>
</tr>
<tr>
<td>Incorrect format output</td>
<td>Apply proper function to cell C4 in sheet Supplementary Data 2</td>
<td>selectedSheet.getFormula(&quot;=SUM(A1:B1)&quot;)</td>
<td></td>
</tr>
<tr>
<td>Overwriting important data</td>
<td>Apply number format to cell B7 of column price of sheet 8.5k: Airline Demand</td>
<td>selectedSheet.getFormula(&quot;=&quot;+formatNumber(8.5k) &quot; Ai&quot;:Airline Demand)</td>
<td></td>
</tr>
</tbody>
</table>