We invited 25 colleagues for 4 hours to generate novel 5x5 tasks, including at least four input-output pairs, and instructed to build a task with a clear and unique solution—Figure 2(a). Then, the generated tasks were submitted to the administrator for approval—Figure 2(b). We pruned around 100 suggested tasks that shared similar concepts during the verification phase and finalized 150 Mini-ARC tasks.

Six Categories
• **Movement** tasks are based on dynamic modifications such as flip, rotation, and sliding sideways.
• **Color** tasks are highly dependent on the color aspect of each pixel, such as swapping colors.
• **Object** tasks are dependent on the movement of the object or agent, where an object refers to an area that can be intuitively distinguished from the background.
• **Number** tasks count something, such as the number of pixels of the same color.
• **Geometry** tasks include problems that require the concept of geometric structures.
• **Common-sense** tasks, like maze-pathfinder or Tetris, require high-level induction even though they may be intuitively evident to people.

Quality Evaluation
In order to rate the novelty and difficulty of the task, participants analyze the created data set and assign a score between 1 and 5. Two grading scales are used, one for human solvers and the other for building AI models. We received 208 answers.

Contributions from 25 participants allowed us to construct Mini-ARC tasks as sufficiently difficult to solve. Six Categories

<table>
<thead>
<tr>
<th>Category</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Movement</td>
<td>Figure 3</td>
</tr>
<tr>
<td>Color</td>
<td></td>
</tr>
<tr>
<td>Object</td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td></td>
</tr>
<tr>
<td>Geometry</td>
<td></td>
</tr>
<tr>
<td>Common-sense</td>
<td></td>
</tr>
</tbody>
</table>

Principles of Curating Mini-ARC Tasks
The primary reason to curate the compact 5x5 data set is to reduce the modeling budget.

O2ARC: Tools for Collecting Expert Demonstrations
Object-oriented ARC (O2ARC), a browser-based interface, was designed so that participants could solve ARC and Mini-ARC problems.

Challenges and suggestions
- **Challenges**
  - As seen in Figure 6B–6J, not all traces contain the high-level process of human intuition. Trace 1 reflects the intuition of diagonal flips using compact movements, and this solution can be generalized to the other pairs following the same rules, while Trace 2 cannot. The restricted input and output grid size of the Mini-ARC tasks may take part in increasing naive actions. How to gather traces that reflect intuition is still an open question.
  - Utilize the traces as a replay buffer for training agents through imitation learning. Analyze the traces based on action sequences to find new primitives that can be generally used for solving multiple Mini-ARC tasks.

**Suggestions**
- The purpose of the Mini-ARC trace is to collect reference trajectories of the Mini-ARC tasks. The following are suggestions for utilizing Mini-ARC in future research.
  - **Object:** use flip and rotate.
  - **Color:** use up and down.

Mini-ARC Trace: Compiled Expert Demonstrations

Figure 1. The Mini-ARC dataset is a condensed version of the ARC problem with a 5x5 grid size.

Figure 2. Interface for curating the Mini-ARC dataset.

Figure 3. Representative Mini-ARC examples for each category.

Figure 4. Survey results of Mini-ARC.

Figure 5. O2ARC tool includes multiple functions to solve Mini-ARC problems.

Figure 6. Different traces are logged for the same diagonal flip problem.

Figure 7. O2ARC tool includes multiple functions to solve Mini-ARC problems.