O2ARC 3.0: A Platform for Solving and Creating ARC Tasks

Suyeon Shim , **Dohyun Ko** , **Hosung Lee** , **Seokki Lee** , **Doyoon Song** , **Sanha Hwang** , **Sejin Kim** and **Sundong Kim**

Gwangju Institute of Science and Technology

{suyeonshim, doridori, gitpush-force, sklee1103, sd990307, hsh6449j}@gm.gist.ac.kr, {sejinkim, sundong}@gist.ac.kr

Abstract

We introduce the O2ARC 3.0 interface for the Abstraction and Reasoning Corpus (ARC). O2ARC 3.0 gamifies the experience, fostering user engagement through competitive features and community-driven problem creation and evaluation. Built with a React frontend and NestJS backend, the platform provides a responsive and intuitive interface for efficient rule inference. This approach not only improves data collection for AI training but also enhances the problem-solving process, offering a scalable solution for advancing cognitive AI research. O2ARC is available at https://o2arc.com.

1 Introduction

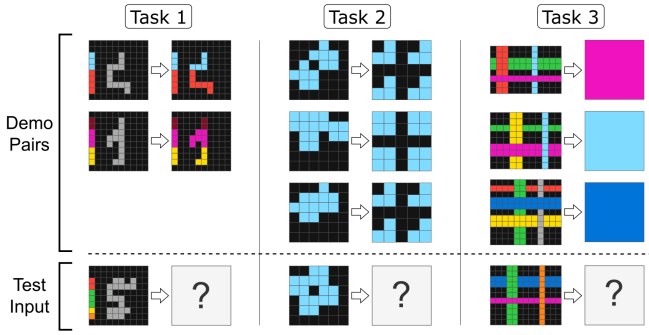


Figure 1: Three different ARC tasks. The ARC dataset is designed to assess the generalization and learning abilities of AI systems across novel tasks, which is crucial for measuring true intelligence.

The Abstraction and Reasoning Corpus (ARC), introduced by François Chollet [2019b], presents grid-based tasks that target both humans and AI systems to infer rules from given input-output pairs and apply them to a test input. This task demands abductive reasoning skills and a level of common sense—capabilities where human cognition excels. In contrast, current AI systems struggle, achieving no more than 30% accuracy on these tasks [Lab42, 2024a; Bober-Irizar and Banerjee, 2024]. To bridge this gap, researchers have highlighted the need for interfaces to collect additional tasks [Moskvichev *et al.*, 2023] and to understand the human-solving process [Johnson *et al.*, 2021; Kim *et al.*, 2022]. Building upon the original interface [Chollet, 2019a], tools like Language-Annotated ARC [Acquaviva *et al.*, 2022], Lab42's ARCreate [2024b], and Object-Oriented ARC (O2ARC 1.0) [Kim *et al.*, 2022] have been developed for this purpose.

While these platforms were a step towards addressing the data collection challenge for AI learning, they fell short in fostering widespread user engagement and were not successful in eliciting spontaneous participation. Consequently, data collection has often depended on extrinsic motivators, such as financial incentives, which are expensive to train data-centric models. The user experience frequently involved tedious single-pixel editing, contrary to the intended object-centric strategy [Xu *et al.*, 2023; Park *et al.*, 2023].

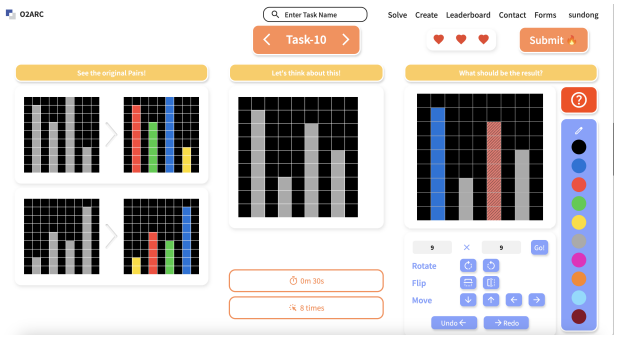
In response, we present O2ARC 3.0, an evolution of the tool that gamifies the problem-solving experience. O2ARC 3.0's interface is designed to mimic the engagement of puzzle games, encouraging voluntary involvement and diminishing the need for external rewards. We have added competitive elements like leaderboards and a system for creating and peer-reviewing tasks to enhance user engagement. The interface is designed to discourage inefficient single-pixel editing and to promote object-based manipulations, ensuring the generation of high-quality human traces.

2 Functionalities and User Engagement

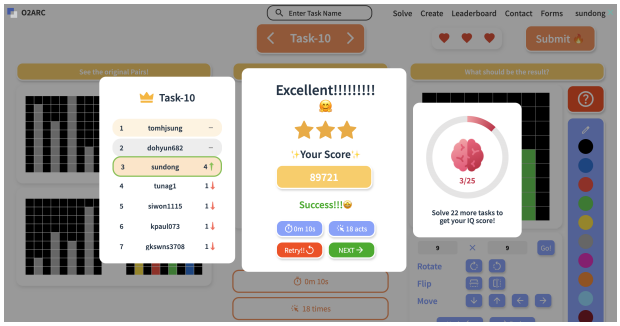
Within O2ARC 3.0, users can engage in various activities:

Solving ARC Tasks The platform allows users to engage with a random task via the navigation bar and presents subsequent tasks in sequence. The editor panel offers various operations, such as rotate, flip, resize grid, move, undo, and redo, to facilitate the construction of the predicted output. Incorrect attempts reduce the user's 'lives', whereas correct solutions display task-specific scores and rankings on the leaderboard, rewarding efficiency and accuracy. To foster continued participation, an IQ revelation feature is unlocked after 25 tasks are completed. Trial and time restrictions are in place to deter excessive focus on any single task.

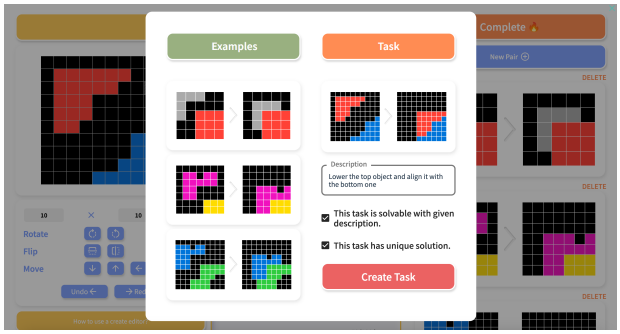
Creating ARC Tasks Users can construct their tasks, requiring a minimum of three demonstrations and a test input-output pair for submission. The creation process mirrors the



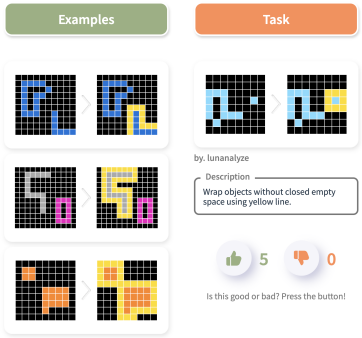
(a) Solve a task with editor panel.



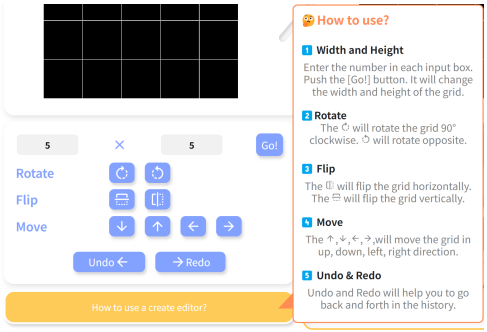
(b) Check your score after solving the task.



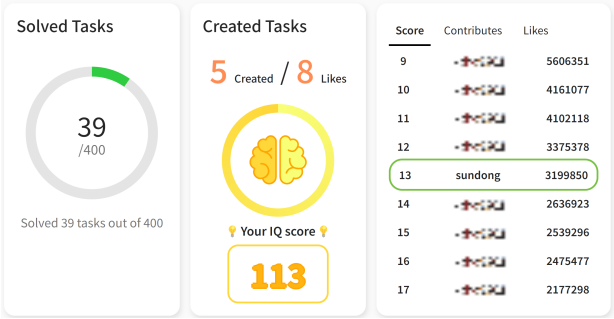
(c) Create a task with four or more pairs.



(d) Evaluate the tasks that others have created.



(e) View the instructions by pressing the '?' button.



(f) See your achievements in the profile tab.

Figure 2: Content that users experience in the O2ARC tool.

solving interface, with additional functionalities like input-output reset. Before submission, creators verify the uniqueness of their solutions and the adequacy of the accompanying descriptions. Community members can then assess the quality of these tasks on the 'created' page.

Profile and Leaderboard The profile tab provides users with a personalized summary of their problem-solving activity, likes received, and leaderboard position. The leaderboard dynamically updates to reflect user scores and rankings, encouraging competition with visual animations and offering recognition for both problem-solving proficiency and creative contributions.

Main Interface The main interface displays the 400 ARC training tasks in a masonry grid, adapting the layout to accommodate the diversity in size and complexity of the problem sets.

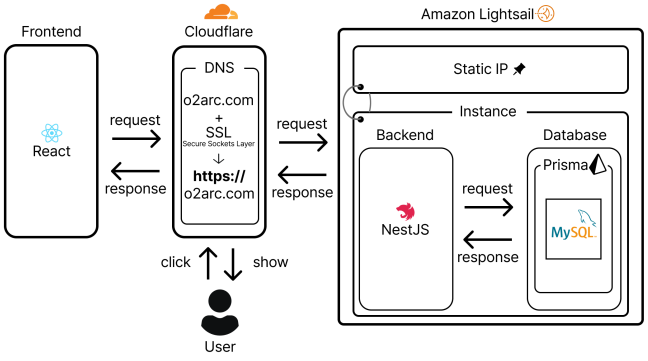


Figure 3: System design of the O2ARC 3.0.

3 System Design

O2ARC 3.0 uses React and NestJS to build the frontend and backend, with the frontend hosted on Cloudflare and the backend hosted on AWS LightSail. Figure 3 illustrates the O2ARC 3.0 system design.

Frontend (React) The frontend of O2ARC 3.0 utilizes React, a leading frontend framework, to create seamless user interfaces. By employing a Single Page Application (SPA) architecture alongside frontend caching, the application delivers an immersive experience with minimal loading times. The use of TypeScript, combined with a functional React architecture, aids in simplifying maintenance and ensuring robustness. Operations within the editor grid are handled locally to reduce backend communication delays. Additionally, the application features dynamic animations on both the leaderboard and scoreboards to maintain user interest and engagement.

Backend (NestJS) NestJS [2024] is a Node.js framework used for building server-side applications. Also, it supports Express, providing an HTTP server. O2ARC 3.0 utilized it for the backend, with code written in TypeScript.

NestJS provides a uniform structure consisting of a controller, a service, a repository, and a module, which makes the code easier to understand and maintain. Additionally, the service file defines which repository functions to call. A service function can call multiple repository functions. The repository file utilizes the Prisma service to perform operations on a specific table in the database, such as finding, inserting, and deleting rows.

Database (Prisma and MySQL) Prisma is an ORM (Object Relational Mapping) framework. It automatically converts the Prisma schema syntax, which contains the database structure, into MySQL statements and executes them.

Backend Hosting and Storage (Amazon Lightsail) Amazon Lightsail is a cloud computing service that provides a virtual server (instance) for backend hosting. O2ARC 3.0 tools utilized a Lightsail instance with a Linux/Unix platform and Ubuntu operating system. The instance has 2GB of RAM, 2 vCPUs, and a 60GB SSD. Pins a static IP to the instance so that it always uses the same IP even if the instance is restarted. It contains the backend system and database tables, where ARC task data and user-solving logs are stored.

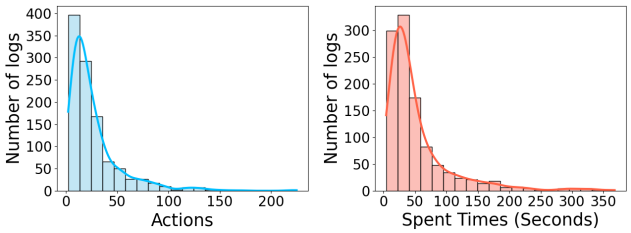
Https Hosting (Cloudflare) Cloudflare is a CDN service, specialized in deploying static frontends. When a user accesses o2arc.com through a browser, Cloudflare delivers the prepared frontend build files to the browser, which then interprets them to render O2ARC's UI and communicate with the Backend. Additionally, we have integrated Cloudflare with our GitHub repository to enable automatic deployment for each branch, automating the tasks necessary for deployment.

Design System (Figma) The UI for each page of the O2ARC 3.0 tool was designed using Figma, a collaborative interface design tool that provides various design actions. Figma also offers a developer mode that is useful for building UI on the frontend.

4 User Engagement and Impact

UX Design In our UX design, we prioritize object-oriented operations over single-pixel manipulations to better reflect human problem-solving strategies. The interface initially featured Edit, Select, and Flood-Fill modes [Chollet, 2019a], each toggled via specific buttons. To minimize single-pixel edits, we removed the Edit mode and modified the Flood-Fill mode to activate through a double-click, now named Flood Selection. This adjustment inherently encourages users to adopt object-oriented operations by limiting the granularity allowed by the previous Edit mode and making object manipulation more accessible.

User Study A user study involving 50 participants evaluated the solve and create functions of O2ARC 3.0, with 24 providing detailed feedback. The tool received high satisfaction ratings (Likert scale from 1–10): 8.7 overall, 8.6 for solving, 8.7 for creating, and 8.8 for evaluating tasks. Key highlights include the leaderboard's role in enhancing competition and motivation, the tool's ease of use and intuitive design, and its effectiveness in engaging users with ARC concepts. This feedback emphasizes the improvements in user engagement and data collection compared to previous versions.



(a) Distribution of Actions (b) Distribution of Spent Times

Figure 4: User study analysis: After preprocessing the data collected from users, we analyzed 1081 logs. The distribution of the number of actions in each log is shown, with most users taking less than 50 actions to solve a problem. Additionally, the distribution of the amount of time spent in each log is presented, with most users solving the task within two minutes.

Broader Impact Inspired by O2ARC, sister projects like ARCLE [Lee *et al.*, 2024] are emerging, leveraging its state and action spaces for training reinforcement learning agents. This collaboration bridges human cognitive processes with AI learning, enriching the ecosystem. Through ARCLE, O2ARC's rich dataset becomes a fertile ground for developing AI agents with human-like reasoning skills.

5 Conclusion

This paper introduces O2ARC 3.0, an engaging interface for Abstraction and Reasoning Corpus (ARC). O2ARC 3.0 is accessible at https://o2arc.com and will be showcased at our booth during the demo session. Attendees are invited to interactively engage in problem-solving and creating activities throughout the event.

Acknowledgements

This work was supported by the IITP (RS-2023-00216011, No. 2019-0-01842), the National Research Foundation (RS-2023-00240062), and GIST (AI-based Research Scientist Project) funded by the Ministry of Science and ICT, Korea.

Contribution Statement

Suyeon Shim and Dohyun are the co-first authors of the paper. They developed O2ARC 3.0, focusing on the backend and frontend, respectively, and also contributed to the writing of the manuscript. Hosung Lee, as the maintainer of the unpublished O2ARC 2.0, authored the operation manual to facilitate an easier transition to O2ARC 3.0. Seokki Lee led the organization and execution of the full-day workshop and recorded the demo video. Doyoon Song played a key role in the initial stages by sketching and prototyping O2ARC 2.0. Sanha Hwang maintained O2ARC 1.0 and provided valuable support for the project. Sejin Kim offered insightful feedback that influenced the project's development. Sundong Kim is the corresponding author. He oversaw the entire project and wrote the manuscript.

References

[Acquaviva *et al.*, 2022] Samuel Acquaviva, Yewen Pu, Marta Kryven, Theodoros Sechopoulos, Catherine Wong, Gabrielle Ecanow, Maxwell Nye, Michael Tessler, and Joshua B. Tenenbaum. Communicating Natural Programs to Humans and Machines. In *NeurIPS*, 2022.

[Bober-Irizar and Banerjee, 2024] Mikel Bober-Irizar and Soumya Banerjee. Neural Networks for Abstraction and Reasoning: Towards Broad Generalization in Machines. *arXiv:2402.03507*, 2024.

[Chollet, 2019a] François Chollet. ARC Testing Interface. https://github.com/fchollet/ARC/blob/master/apps/testing_interface.html, 2019.

[Chollet, 2019b] François Chollet. On the Measure of Intelligence. *arXiv:1911.01547*, 2019.

[Johnson *et al.*, 2021] Aysja Johnson, Wai Keen Vong, Brenden M. Lake, and Todd M. Gureckis. Fast and Flexible: Human Program Induction in Abstract Reasoning Tasks. In *CogSci*, 2021.

[Kim *et al.*, 2022] Subin Kim, Prin Phunyaphibarn, Donghyun Ahn, and Sundong Kim. Playgrounds for Abstraction and Reasoning. In *NeurIPS Workshop on Neuro Causal and Symbolic AI*, 2022.

[Lab42, 2024a] Lab42. Arcathon leaderboard. https://lab42.global/arcathon/leaderboard/, 2024.

[Lab42, 2024b] Lab42. ARCreate. https://lab42.global/arcreate/, 2024.

[Lee *et al.*, 2024] Hosung Lee, Sejin Kim, Seungpil Lee, Sanha Hwang, Jihwan Lee, Byung-Jun Lee, and Sundong Kim. ARCLE: The Abstraction and Reasoning Corpus Learning Environment for Reinforcement Learning. In *CoLLAs*, 2024.

[Moskvichev *et al.*, 2023] Arseny Moskvichev, Victor Vikram Odouard, and Melanie Mitchell. The ConceptARC benchmark: Evaluating Understanding and Generalization in the ARC Domain. *TMLR*, 2023.

[NestJS, 2024] NestJS. [NestJS] Documentation. https://nestjs.com/, 2024. Accessed: 2024-02-19.

[Park *et al.*, 2023] Jaehyun Park, Jaegyun Im, Sanha Hwang, Mintaek Lim, Sabina Ualibekova, Sejin Kim, and Sundong Kim. Unraveling the ARC Puzzle: Mimicking Human Solutions with Object-Centric Decision Transformer. In *ICML Workshop on Interactive Learning with Implicit Human Feedback*, 2023.

[Xu *et al.*, 2023] Yudong Xu, Elias B. Khalil, and Scott Sanner. Graphs, Constraints, and Search for the Abstraction and Reasoning Corpus. In *AAAI*, 2023.