

# 인지 및 추론 연구를 위한 테스트베드 (Test Bed for Abstraction and Reasoning)

김수빈<sup>†</sup>    프린푸나피반<sup>\*\*</sup>    안동현<sup>†</sup>    김선동<sup>\*\*\*</sup>  
(Subin Kim)    (Phunyaphibarn Prin)    (Donghyun Ahn)    (Sundong Kim)

**요약** 프랑소와 솔레(François Chollet)가 제안한 Abstraction and Reasoning Corpus (ARC)는 특정 문제에 매몰되지 않고, 일반화가 가능한 지능의 개발을 위해 디자인된 아이큐 테스트 형태의 벤치마크로, 인간과 컴퓨터 모두의 인지 능력을 측정하기에 적합하다. 대부분의 문제를 풀 수 있는 사람에 반해, 30% 이상의 문제를 풀 수 있는 컴퓨팅 기반 ARC-Solver는 알려지지 않았다. 이 연구에서는 기존 ARC의 난이도를 유지하면서도 모델의 복잡도를 간소화하기 위해 탐색 공간을 최소화한 벤치마크 데이터 Mini-ARC를 소개한다. Mini-ARC의 수집을 위해, 인간의 풀이 과정을 추적할 수 있는 인터페이스인 O2ARC를 고안하였으며, 이를 통해 25명에게 총 3,000여 개의 풀이를 수집하는 데 성공했다. 이 연구는 간소화된 인간의 인지 과정과 그 풀이 과정을 대량으로 확보하는 시스템을 구축하여, 컴퓨팅 기반 ARC-Solver 개발의 새로운 접근법을 제시한다.

**키워드:** abstract and reasoning corpus, 범용 인공 지능, Mini-ARC, O2ARC

**Abstract** Abstraction and Reasoning Corpus (ARC) proposed by François Chollet is a benchmark designed to develop generalizable intelligence, suitable for measuring cognitive abilities of both humans and computers. While most problems can be solved by humans, a computing-based ARC-Solver that can solve more than 30% of the problems is not yet known. In this study, a benchmark dataset, Mini-ARC, was introduced to simplify the model complexity while maintaining the difficulty level of the original ARC. To collect Mini-ARC, O2ARC was designed. It is an interface that can track the human solution process. A total of 3,000 solutions were collected from 25 people. This study proposed a new approach to developing a computing-based ARCSolver by constructing a system that could massively secure the simplified cognitive solution process. The Mini-ARC dataset can be found at <https://github.com/ksb21ST/Mini-ARC>.

**Keywords:** abstract and reasoning corpus, artificial general intelligence, Mini-ARC, O2ARC

## 1. 서론

구글의 LaMDA, Open AI의 DALL·E 등을 필두로 개발되고 있는 인공지능 모델은 인간이 평생 접하지 못

할 크기의 지식을 습득한 후 패턴화하여 분류, 이상 탐지, 이미지 생성 등 특정 임무에서 인간을 능가하는 성능을 보여주고 있다[1].

하지만, 아직 접하지 못한 문제에 대한 해결 능력을 갖

· 이 논문은 과학기술정보통신부의 재원으로 한국연구재단과 정보통신기획평가원의 지원을 받아 수행된 연구임(RS-2023-00240062, RS-2023-00216011, 2019-0-01842)

· 이 논문은 2022 한국소프트웨어종합학술대회에서 '인지 및 추론 연구를 위한 Mini-ARC 벤치마크 데이터'의 제목으로 발표된 논문을 확장한 것임

† 학생회원 : 한국과학기술원 전산학부 학생  
21supersoo@kaist.ac.kr  
segaukwa@kaist.ac.kr

\*\* 비회원 : 한국과학기술원 전산학부 학생  
prin10517@kaist.ac.kr

\*\*\* 비회원 : 광주과학기술원 AI대학원 교수(GIST)  
sundong@gist.ac.kr  
(Corresponding author임)

논문접수 : 2023년 3월 16일

(Received 16 March 2023)

논문수정 : 2023년 9월 19일

(Revised 19 September 2023)

심사완료 : 2023년 10월 7일

(Accepted 7 October 2023)

Copyright©2024 한국정보과학회: 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.  
정보과학회논문지 제51권 제1호(2024. 1)

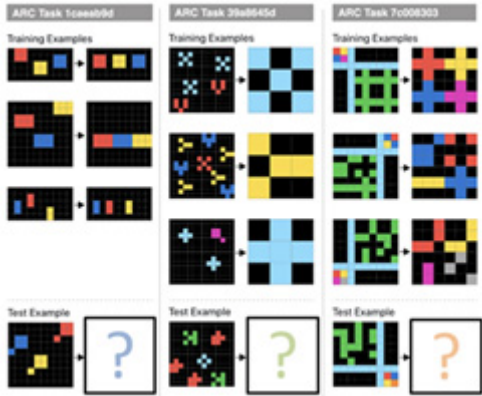


그림 1 ARC 문제 예시  
Fig 1 Example of ARC

준 범용 인공지능 모델(Artificial General Intelligence; AGI)을 위한 연구는 비교적 부진하다. 일례로, ‘내부 구조를 모르는 집안에 들어가서 커피를 끓이는 문제’로 알려진 스티브 위즈니악의 커피 테스트는 현존하는 AI기술을 활용해서는 해결이 어려운 문제로, 사전에 준비되지 않은 일에 대응할 수 있는 기반 기술의 중요성을 역설한다. 이러한 환경을 단순화한 문제가 바로 Abstraction and Reasoning Corpus (ARC)이다[2,3]. ARC는 적은 양의 데이터로부터 문제에 담긴 패턴을 추론하고 새로운 문제에 적용하는 인지 능력을 측정하기 위해 고안되었다. 이를 위해 그림 1에서 볼 수 있듯이 각 ARC 문제에서 문제를 푸는 주체는 최대 3쌍의 입출력 예제를 통해 패턴을 추론하여 마지막 열에 제시된 입력에 대응하는 추론 값을 예측해야 한다.

ARC는 다른 IQ 테스트들과 마찬가지로, 살면서 축적해 온 다양한 상식을 바탕으로 문제에 담긴 패턴을 발견하고, 이를 적용할 수 있는 능력을 확인한다. 정확한 정답 그리드를 생성해야 하는 ARC 문제의 특성상 출제자의 의도를 정확히 파악한 후 논리에 근거하여 추론하는 방법론의 개발이 필수적이며, Raven’s Progressive Matrices와 같은 오지선다형 IQ 테스트의 솔루션들에서 활용된 유사도 기반의 신경망 모델을 활용하기 어렵다 [4, 5]. 아울러 ARC 문제에서 요구하는 다양한 입출력 형태는 일정 수준 이상의 복잡도를 요구하여, 해결책을 고안하는 과정에서 고려해야 할 부분이 배가되었다.

이 연구에서는 더욱 간소화된 Mini-ARC 벤치마크 테스트를 소개하고, 다양한 기능을 활용해 Mini-ARC를 풀어볼 수 있는 인터페이스인 O2ARC 및 참여자들의 솔루션인 Mini-ARC trace를 소개한다. 입력과 출력의 크기가 다양한 기존의 ARC와 달리, Mini-ARC는 입출력 그리드 공간을 5x5로 제한하여 인공지능 모델의 훈

련에 대한 실효성을 높였다. 더불어, 여섯 개의 기능이 추가된 O2ARC 인터페이스는 기존에 비해 높은 수준의 편의성을 제공한다. O2ARC를 통해 수집된 풀이 과정인 Mini-ARC trace는 추후 모방 학습, 프로그램 합성에 기반한 솔루션 개발에 활용할 수 있을 것으로 기대된다.

## 2. Mini-ARC

이 연구에서 소개하는 Mini-ARC 데이터는 150개의 문제로 구성되었으며, 입출력 크기를 5x5로 제한하여, 인공지능 모델 개발 단계에서의 성능을 효율적으로 확인하고자 한다.

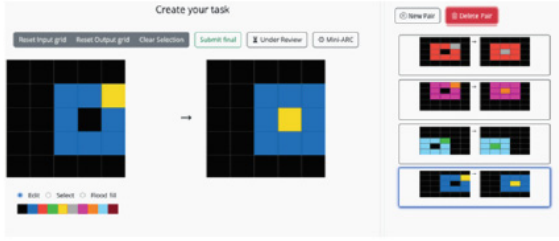
### 2.1 제작 원칙

기존의 ARC의 입력과 출력의 크기는 1x1에서 30x30 까지 다양하며, 모델 학습에 필요한 탐색 공간은 입출력의 크기에 비례하여 커진다. 따라서 이 연구에서는 학습의 탐색 크기를 제한하여 더욱 가벼운 훈련 환경을 제공함과 동시에, 모든 제작자가 동일한 조건에서 문제를 구상할 수 있도록 Mini-ARC 문제의 입출력 크기를 5x5로 단일화하였다. 크기의 설정 배경은 다음과 같다.

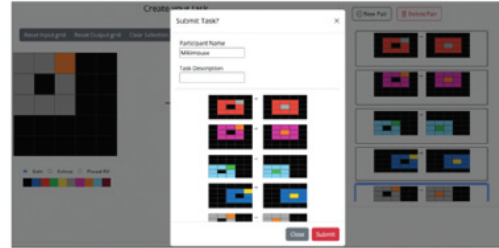
- 1) 입출력 공간이 중심점을 가지도록 크기를 홀수로 설정하였다.
- 2) 정사각형 모양의 입출력 공간은 문제를 풀 때 전체 그리드의 회전, 뒤집기와 같은 기본적인 변형을 사용하도록 유도한다.
- 3) 5x5 미만의 경우 회전과 대칭을 구별하기 어렵기 때문에 사용하기 어렵다. 문제 제작을 위해, 25명의 참가자들을 모집하였고, 5x5로 고정된 크기 내에서 ARC 문제의 의도에 맞는 명확하고 유일한 해결책을 가지는 Mini-ARC를 구상하도록 지시하였다. 패턴 파악을 위해 문제당 최소 4개의 입출력 쌍을 포함하도록 했다. 참가자들은 4시간 동안 새로운 문제 제작에 임하였고, 그 결과 총 150개의 Mini-ARC 문제를 수집할 수 있었다.

### 2.2 Mini-ARC 수집 과정

Mini-ARC 수집 과정은 다음과 같다: Mini-ARC 문제를 만들기 위한 그림 2와 같이 웹 브라우저 기반 인터페이스를 직접 만들었다. 참가자들은 각각 최소 4쌍의 입출력 쌍을 포함하고 있는 문제를 제출할 수 있으며, 제출된 작업은 관리자에 의해 심사되도록 대기열에 포함된다. 관리자는 제출된 문제들을 확인하고, 3명의 관리자 중 2명 이상이 승인한 문제는 Mini-ARC에 포함된다. 참가자들은 작업을 하는 동안 다른 제출물들을 확인할 수 있으며, 기존의 제출물을 새로운 문제로 수정할 수 있다. 그림 3은 인터페이스를 통해 수집된 하나의 Mini-ARC 문제에 대한 json 형식의 메타데이터를 나타낸다. 문제를 만든 사람과 문제에 대한 설명, 승인되었는지의 여부를 포함하고 있다. Testjson 항목에는 4쌍의 입출력이 포함되어 있으며, 하나의 입력 또는 출력



(a) Mini-ARC 제작 화면



(b) Mini-ARC 제출 화면

그림 2 Mini-ARC 제작에 사용된 인터페이스

Fig. 2 Interface for curating Mini-ARC

```
{id : 5948,
user_id : Subin,
test_id : l69906wfy3uwotg3svi,
testjson :
[[input cells: [[0, 0, 1, 2, 3], [...], [...], [...], [...]],
output cells: [[0, 1, 2, 4, 0], [...], [...], [...], [...]]],
{input cells: [...],
output cells: [...],
...},
Approve : 1,
Description : "bouncing ball"}
```

그림 3 수집된 Mini-ARC 의 메타데이터 예시  
Fig. 3 Meta-data of Mini-ARC in json format

항목에는 숫자로 이루어진 5x5 배열이 포함되어 있고, 각 숫자는 입출력에서 하나의 칸이 가지는 색을 나타낸다.

### 2.3 Mini-ARC 유형 분석

Mini-ARC는 각 문제를 구성하는 대표적인 개념에 따라 움직임, 색깔, 객체, 숫자, 기하, 상식의 6가지 범주로 분류할 수 있다. 그림 4에서 각 유형별로 대표적인 예시를 보이고 있다.

움직임 유형은 뒤집기, 회전 및 옆으로 이동 등과 같은 동적 움직임을 기반으로 한다. 색깔 유형은 각 칸의 색깔을 기반으로 두 칸의 색깔을 교환하는 등의 패턴이 존재한다. 객체 유형은 객체 단위의 상호작용이나 움직임을 파악한다. 여기에서 객체는 배경과 직관적으로 구별할 수 있는 색칠된 영역을 의미한다. 숫자 유형에서는 같은 색깔의 칸의 개수, 또는 한 객체 내에 포함된 칸의 개수 등에 기반한 셈을 요한다. 기하 유형에서는 점, 선, 면 등의 기하학적 개념을 필요로 한다. 상식 유형에서는 높은 수준의 인지 기능을 필요로 하는 문제가 속한다 (테트리스, 미로찾기 등). 문제 해결에 쓰이는 기능의 조합 등이 유형별로 다를 수 있다.

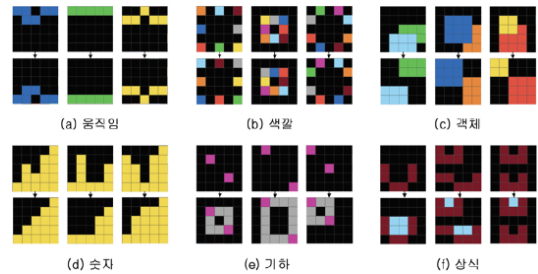


그림 4 Mini-ARC 문제 유형 별 예시

Fig. 4 Representative Mini-ARC examples of each category

### 2.4 데이터셋 평가

각 참가자는 다른 참가자들이 생성한 문제를 평가하고 독창성 및 난이도 측면에서 1-5점 사이의 점수를 부여했다. 점수가 높을수록 독창적인 문제이거나, 난이도가 높은 문제라는 것을 의미한다. 문제 난이도는 두 가지로, 평가자가 직접 문제를 풀었을 때 느낀 문제의 난이도와, 해당 문제를 푸는 프로그램을 개발할 때 예상되는 난이도로 나뉜다. 평가자가 직접 풀었을 때 느낀 난이도가 높을수록 고도의 인지 능력을 필요로 한다는 가정을 기반으로, 각 문제가 인간의 인지 능력을 얼마나 활용하는지 측정하기 위한 과정이다. 만약 해당 문제의 프로그램을 개발하는 것이 어렵다면, 문제에 활용된 규칙이 코드를 통해 해결하기엔 직관적이지 않다는 점을 의미한다. 각 문제는 최소 1회 이상의 평가를 받았으며, 총 208회의 평가가 이루어졌다.

그림 5-(a)에서 볼 수 있듯이, 25명의 참여자들의 기여로 우리는 Mini-ARC를 충분히 독창적인 문제들로 구성할 수 있었다(평균 2.59). 그럼에도 불구하고, 그림 5-(b)에서 볼 수 있듯이 응답자의 대부분은 주어진 문제를 직관적으로 해결할 수 있다고 평가하였는데 (평균 2.02), 이는 인간은 뛰어난 인지 능력을 토대로 창의적인 문제도 해결할 수 있다는 점을 시사한다. 참여자들

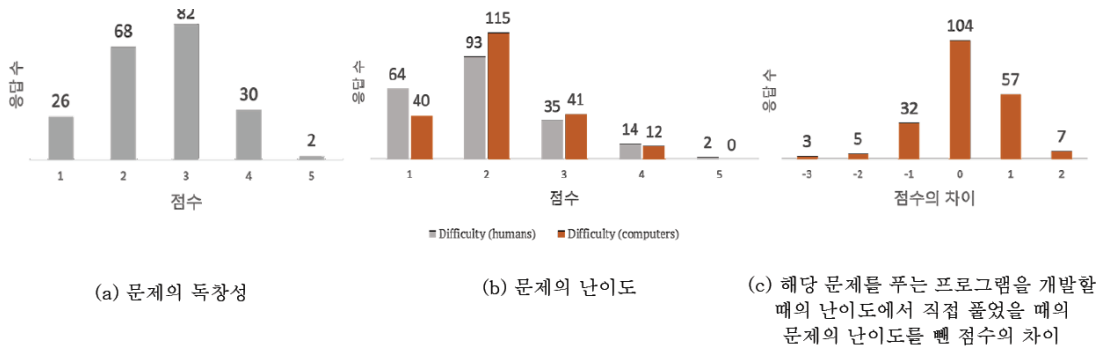


그림 5 Mini-ARC에 대한 설문조사 결과  
Fig. 5 Survey results of Mini-ARC

은, 각각의 문제를 해결하는 프로그램을 개발하는 것이 직접 푸는 것보다 더 어렵다고 느꼈다 (평균 2.12). 이를 보여주기 위한 그림 5-(c)로, 각 응답에서 평가자가 해당 문제에 대한 프로그램을 개발할 때 예상되는 난이도에서 평가자가 느낀 문제의 난이도를 뺀 결과를 보여준다. 두 난이도가 동일하게 느껴졌다는 응답이 전체 응답 수의 절반인 104개이었지만, 나머지 104개의 중 64개의 응답의 경우, 직접 문제를 푸는 것보다 프로그램을 구현하기가 더 어렵다는 의견이 있었다.

2.5 다른 ARC 데이터셋과의 비교

ARC를 확장한 데이터셋인 1D-ARC는 기존의 ARC를 단순화하여, 대형 언어 모델이 더 쉽게 접근할 수 있는 기준을 제공하기 위해 입력 및 출력 이미지의 차원을 하나로 제한한다[6]. 이는 비순차적 텍스트에서 객체 일관성을 유지하는 어려움을 효과적으로 해소한다. Concept ARC는 특정 개념에 따라 일반화하는 수준을 측정할 수 있도록 기존 ARC를 16가지 유형으로 나누고 각 유형별 10개의 문제를 제작했다[7]. 1D-ARC와 Mini-ARC는 모델의 용이한 개발을 위해 차원을 축소하였고, Concept ARC와 Mini-ARC는 모델의 능력을 체계적으로 측정하기 위해 난이도를 나누었다.

표 1 다른 ARC 데이터셋과의 비교

Table 1 Comparison with other ARC-like datasets

	Number of problems	Number of shapes	Reducing dimensions	Distinct levels
1D-ARC	50	8	o	x
Concept ARC	160	16	x	o
Mini-ARC	150	8	o	o

2.6 5x5 공간이 가진 한계 및 의의

5x5로 입출력의 크기를 제한하였기에, 더 큰 공간을 필요로 하는 일부 아이디어는 기각되었다. 그럼에도 불

구하고 참가자들은 5x5의 공간 안에서 다양한 종류의 문제를 만들기 위해 노력하였다. 데이터의 크기가 작은 만큼 보다 가벼운 환경에서 Mini-ARC를 위한 인공지능 모델을 개발한 이후, 입출력 크기의 제한이 없는 ARC 데이터셋에 맞는 형태로 모델을 확장해 볼 수 있다. 그러나 5x5 크기의 공간 안에서 기존의 입출력 크기 제한이 없는 ARC 처럼 다양한 규칙을 만들어 내는 것에 한계가 있다.

3. O2ARC

참가자들이 ARC 및 Mini-ARC 문제를 풀어볼 수 있도록 브라우저 기반 인터페이스인 Object-oriented ARC(이하 O2ARC)를 설계하였다. 많은 문제가 객체 지향적이라는 것을 토대로, O2ARC는 ARC 제작자가 제공한 기본적인 툴 위에 문제 해결 과정을 보조하는 6가지 기능을 추가하였다. 그림 6은 O2ARC의 기능을 보이고 있다.

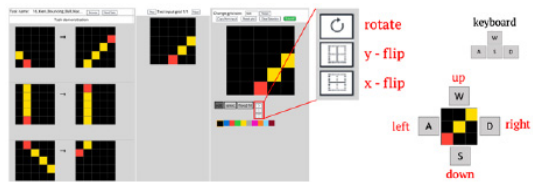


그림 6 O2ARC 예시  
Fig. 6 Example of O2ARC

3.1 원본 도구를 재설계한 동기

원본 ARC 데이터셋에서 제공하는 도구에 포함된 기능은 매우 기본적인이다. 세 개의 심볼 컨트롤(edit, select, flood fill)과 세 개의 그리드 컨트롤(copy from input, resize grid, reset grid to null)이 포함되어 있다. 이러한 기본적인 기능으로 모든 ARC 문제를 해결

할 수 있지만, 사용자는 문제를 해결하는 동안 버튼을 반복해서 클릭해야 하는 불편함이 있었다. 따라서 이 도구를 이용해서 문제를 푸는 과정은 길고 복잡하며, 때로는 문제를 한눈에 해결하는 직관성을 반영하지 못한다.

**3.2 새로운 6개의 기본 기능 추가**

사람들의 직관을 반영할 수 있는 기능을 위해 파워포인트, 포토샵 등의 편집 도구에서 영감을 받아 6가지 기능을 추가했다. 이는 상하좌우 움직임, 잘라내기, 되돌리기, 뒤집기, 회전, 여러 칸 동시 선택이다. 상하좌우 움직임은 방향키를 이용해서 한 칸 씩 이동할 수 있다. 잘라내기는 마우스 드래그를 통해 도형의 일부를 잘라낼 수 있다. 되돌리기는 실행 취소를 할 수 있다. 뒤집기는 선택한 칸들을 가로축, 또는 세로축에 대해 선택한 칸을 대칭이 되게 뒤집을 수 있다. 회전은 시계 방향으로 90도씩 기울게 된다. 또한, 기존에 한 칸 씩 밖에 선택하지 못하던 기능을 개선하여 shift 키를 누르고 여러 칸을 클릭하면 동시에 선택이 가능하게 했다. 더불어, 문제 해결 과정에서 사용한 기능들의 조합과 순서, 그리고 그에 따른 답안을 추적할 수 있다. 알파 테스트를 통해 세부 사항을 개선한 O2ARC는 보다 높은 수준의 편의성을 제공하였다.

**3.3 레이어 기능의 추가 및 삭제**

레이어 기능은 포토샵과 유사하게 여러 겹의 수정 가능한 레이어를 조작하는 기능으로, 여러 겹의 레이어를 개별적으로 수정 가능하고, 각 레이어를 합치거나 삭제

할 수 있는 기능이다. 레이어로 사용자는 각 구분되는 물체를 레이어 단위로 조작함으로써 Object-oriented 풀이 방식을 장려하고자 했다. 그러나 ARC 보다 단순한 5x5 크기의 Mini-ARC 문제에서 이 기능이 많이 사용되지 않는 것이 관찰되어 레이어 기능을 제외하였다.

**3.4 인터페이스 디자인의 추가 및 삭제**

3개의 예제 입출력 쌍을 보고 남은 입력에 대한 출력을 예측하는 하나의 ARC 문제를 풀 때, 예제를 한 번에 모두 보여주는 것이 아니라 입력-출력 쌍을 하나씩 보여주는 방식으로 인터페이스 레이아웃을 변경해 보았다. 사용자는 어떤 행동을 취하기 전에 몇 개의 입력-출력 쌍을 보는지, 사용자는 각 쌍을 몇 초 동안 보는지 등을 파악하여 문제의 난이도나 참가자의 능력을 평가하려고 했다. 그러나 예제 입출력 쌍 전체를 보는 것이 더 직관적이라는 피드백을 고려해 이 디자인은 제외되었다.

**3.5 참가자들의 문제 풀이 과정 기록**

O2ARC는 참가자들이 선택한 일련의 기능과 이에 따른 각 그리드 상태의 변화를 기록할 수 있다. 참가자가 한 가지 기능을 선택할 때마다 기능의 이름과 그 기능이 적용된 그리드의 영역이 기록된다. 이 기능이 적용된 전후의 그리드의 상태를 색깔에 따라 숫자로 나타낸다. 이를 통해 참가자들이 어떤 기능을 사용하여 그리드를 어떻게 변화시켰는지 시간에 따른 변화를 파악할 수 있다. 이는 추후에 서술할 Mini-ARC Trace의 수집에 O2ARC를 사용하는 이유이다.

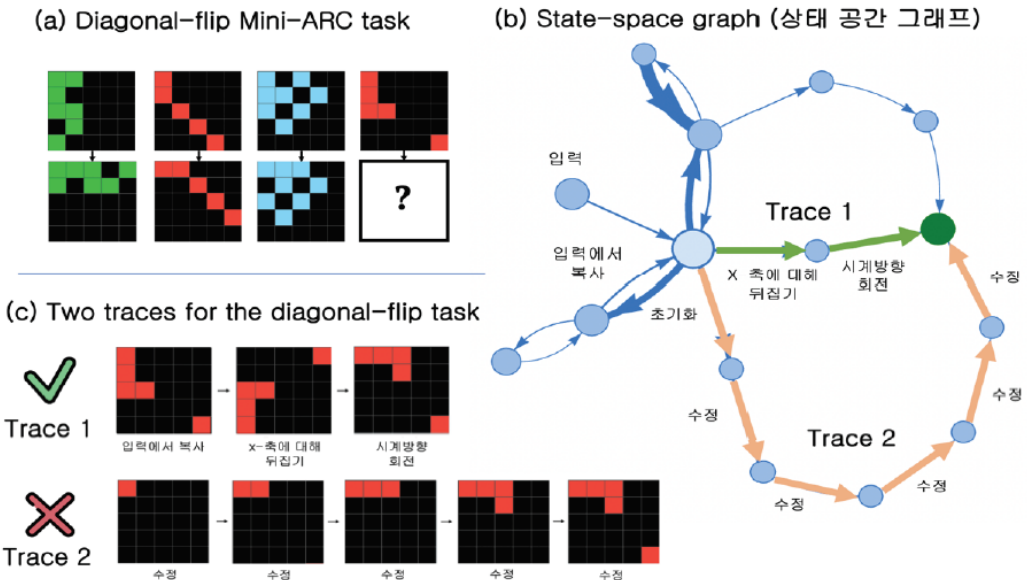


그림 7 대각선 뒤집기 Mini-ARC 문제와 해당 문제를 접근한 두 가지 대표 풀이법을 담은 상태 공간 그래프  
 Fig. 7 Diagonal flip problem logged into different traces, and the state space graph that represents it

## 4. Mini-ARC Trace

참가자들이 문제를 푸는 과정을 분석한다면 사람의 인지 과정을 유추할 수 있을 것이다. 참가자들이 문제를 푸는 과정에서 선택한 일련의 기능과 그에 따른 그리드 상태의 변화, 문제를 풀기 위한 시간을 O2ARC를 통해 수집하여 이를 Mini-ARC trace(이하 trace)로 명명하였다.

### 4.1 Mini-ARC trace 수집 과정

Trace 수집을 위해 20명의 참가자는 2시간 동안 50개의 선별된 Mini-ARC 문제를 해결하였다. 50개의 문제는 3.3장에 언급되었던 6가지 범주 중에서 균일하게 선별되었으나, '상식' 범주에 속한 문제는 최대한 배제하였다. 이 범주에 속한 문제는 높은 수준의 직관을 요구하여 O2ARC에 구현된 기능 만으로는 인지 과정을 제대로 파악하기 힘들기 때문이다. 풀이의 중복을 예방하고 지나치게 단순한 풀이 과정을 지양하기 위해 참가자들은 각 문제 당 최소한 3가지 이상의 다른 풀이 방식을 제출했다. 오답 3개를 연속으로 제출한 경우, 문제의 해결책을 찾지 못하였다고 판단하였다. Mini-ARC trace는 4.1장과 4.2장에서 언급된 기능들 중 하나가 쓰일 때마다 사용된 기능과 이에 따라 변한 출력을 저장하였다. 이는 행동 (action)과 상태 (state)를 기준으로 세밀하게 풀이를 추적하기 위함이다. 추가로 문제를 풀 때까지 걸리는 시간까지 수집했는데, 이는 후에 효율적인 풀이를 구분하기 위해 사용될 수 있다.

### 4.2 Mini-ARC 예제 및 수집된 trace 분석

그림 7-(a)는 그리드 위의 객체를 대각선 축을 중심으로 뒤집는 '대각선 뒤집기' 문제이다. 그림 7-(b)의 상태 공간 그래프는 해당 문제의 서로 다른 풀이법(trace)들을 나타내며, 상태 공간 그래프 속 초록색 선은 trace 1을, 주황색 선은 trace 2를 나타낸다. 구체적인 풀이 방법은 그림 7-(c)를 통해 확인할 수 있으며, 두 참여자가 문제를 풀기 위한 논리를 유추할 수 있다.

Trace 1은 회전, 뒤집기 등 O2ARC에서 새롭게 구현한 기능들을 이용해 대각선 뒤집기 로직을 구현하였다. 반면, Trace 2는 정답에 해당하는 픽셀들을 한 칸씩 빨간 색으로 색칠하는 직관적인 방식으로 문제를 해결하였다. Trace 1에서 구현된 대각선 뒤집기 로직은 모든 예제들에 일반화하여 적용될 수 있지만, Trace 2에 사용된 직관은 다른 예제들에 곧바로 적용될 수 없는 불완전한 해결책이다. 결과적으로, 대각선으로 뒤집어야 하는 출제자의 의도를 논리적으로 반영한 Trace 1이 완벽한 정답이다.

입출력 크기를 5x5로 제한한 Mini-ARC 문제의 경우 패턴이 복잡하더라도 같이 단순한 동작만을 이용하여 출력을 생성하는 것이 어렵지 않기 때문에 Trace 1과

같은 솔루션이 상당히 많았다. 이렇게 수집된 trace는 문제 해결 과정을 반영하지 않기 때문에 모델 학습에 사용하기 어렵다. 상태 공간 그래프 내에 존재하는 정보의 육석을 가리기 위한 고민이 필요하다.

### 4.3 Mini-ARC trace를 활용한 추후 연구 방안

우리는 인간의 문제 해결 능력을 효율적으로 학습할 수 있는 인공 지능 모델을 만들기 위해 Mini-ARC trace를 수집하였다. Mini-ARC trace를 이용하는 한 가지 방법으로는, 인간을 모방할 수 있는 모델을 훈련하는 imitation learning이 있다[8]. Object-centric Decision Transformer는 문제 풀이 과정 속 각 단계별로 추출한 물체에 대한 정보를 이용해 Decision Transformer를 훈련하여 특정 유형의 ARC 문제를 60% 이상 풀어내는 성과를 보였다[8,9]. 이처럼 다양한 문제에서 공통으로 사용된 패턴을 토대로, ARC에서 일반적으로 사용되는 기능의 조합을 도출하고, 이를 활용한 효율적인 프로그램 합성 및 강화 학습 모델 개발을 할 수 있다[9,10].

## 5. 배경지식

### 5.1 ARC 를 풀기 위한 다양한 시도

ARC를 풀기 위한 시도들은 다양한 방법을 통해 이루어졌다. 그 중 하나는 DreamCoder로, 이는 프로그래밍 언어의 기본 구조와 개념을 학습하며 새로운 프로그램을 생성하는 인공 지능 시스템이다[10]. 인공 신경망, 확률론적 프로그래밍, 유전 알고리즘 등의 기술을 활용하여 입력과 출력 예제를 기반으로 문제를 해결한다. 또한, 양방향 프로그램 검색 방법을 사용하여 프로그램 검색 공간을 좁히고 빠른 프로그램 생성을 가능케 했다[11]. 그러나 이 방법은 여전히 한정적이고 단순한 유형의 문제만을 해결할 수 있는 한계가 있다.

또 다른 시도는 Language-complete ARC (LARC)로, 이는 ARC 데이터 세트를 확장한 것으로 자연어를 사용하여 문제와 솔루션을 설명하는 것을 목표로 한다[12]. Descriptor 그룹과 Builder 그룹으로 나뉘어 사용자는 예제 입출력 쌍을 보고 문제를 자연어로 표현하거나 이를 기반으로 출력을 생성한다. 이를 통해 인간-인간과 인간-기계 간의 소통 차이를 분석했는데, 한정된 개념의 사용이 큰 차이임을 발견했다. 이 외에도 대형 언어 모델(LLM)을 이용해 ARC를 풀려는 시도는 꾸준히 지속되고 있다[13-15].

또한, Abstract Reasoning with Graph Abstractions (ARGA)는 객체 중심적 접근 방식을 사용하여 그래프 기반 이미지 표현 방법을 제안한다[16]. 이는 추상화된 그래프 공간을 기반으로 한 DSL에서 올바른 프로그램을 찾기 위한 검색을 수행한다. 실험 결과에서는 ARGA가 검색 효율성에서 더 뛰어나며 훈련과 테스트 과정에

서 해결된 문제 수의 차이가 작았다.

ARCathon은 스위스 다보스에서 개최되며, ARC 문제를 해결하는 인공지능을 경합하는 대회이다[17]. 현재까지 알려진 알고리즘은 ARC 문제의 30.5%만을 해결할 뿐이지만, 대회에서 정답률 31% 이상을 얻으면 상금을 받을 수 있으며, 42%에 도달한 팀은 스위스 AI 어워드에서 포상을 받을 수 있다.

## 5.2 ARC를 응용한 데이터셋

1D-ARC는 ARC를 단순화한 데이터셋으로, 입력과 출력 이미지의 차원을 하나로 제한하여 대형 언어 모델이 더 쉽게 다룰 수 있도록 하였다. 이로써 비순차적 텍스트에서 객체 일관성을 유지하는 어려움을 효과적으로 해결할 수 있다[6]. Concept ARC는 기존 ARC를 16가지 유형으로 분류하고 각 유형별로 10개의 문제를 생성하여, 특정 개념에 따라 일반화 수준을 측정할 수 있도록 했다[7]. 또한, Mini-ARC와 Concept ARC는 모델의 용이한 개발을 위해 차원을 축소하고, 난이도를 나누어 모델의 능력을 체계적으로 평가할 수 있도록 조정되었다.

ARCreate는 ARC 문제를 활용하여 5,000개의 고품질 데이터셋을 클라우드소싱하여 만드는 프로젝트로, 이 데이터셋은 ARC의 기본 구조인 Core Knowledge에 부합해야 한다.

## 6. 결론

이 논문은 간결한 범용 인공 지능 벤치마크인 Mini-ARC 데이터셋과 편리한 문제 해결을 위한 인터페이스인 O2ARC, 그리고 인간의 해결 과정을 담은 Mini-ARC trace를 소개한다. 이를 활용하여 범용 인공 지능에 한 단계 다가갈 수 있는 프로그램 합성 및 강화 학습 모델을 연구할 수 있으리라 기대한다. Mini-ARC 데이터셋은 <https://github.com/ksb21ST/Mini-ARC> 에서 확인할 수 있다.

## References

- [1] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, ... & I. Sutskever, "Zero-shot text-to-image generation.", In *International Conference on Machine Learning*, PMLR, pp. 8821-8831, July, 2021.
- [2] F. Chollet, "On the measure of intelligence.", *arXiv preprint arXiv:1911.01547*, 2019.
- [3] E. S. Spelke & K. D. Kinzler, "Core knowledge.", *Developmental science*, Vol. 10 No. 1, pp. 89-96, 2007.
- [4] M. Małkiński & J. Mańdziuk, "Deep Learning Methods for Abstract Visual Reasoning: A Survey on Raven's Progressive Matrices.", *arXiv preprint arXiv:2201.12382*, 2022.
- [5] A. Johnson, W.K. Vong, B. M. Lake & T. M. Gureckis, T. M., "Fast and flexible: Human program induction in abstract reasoning tasks.", *arXiv preprint arXiv:2103.05823*, 2021.
- [6] Y. Xu, W. Li, P. Vaezipoor, S. Sanner, & E. B. Khalil, "LLMs and the Abstraction and Reasoning Corpus: Successes, Failures, and the Importance of Object-based Representations." *arXiv preprint arXiv:2305.18354*, 2023.
- [7] A. Moskvichev, V. V. Odouard & M. Mitchell, "The ConceptARC Benchmark: Evaluating Understanding and Generalization in the ARC Domain.", *arXiv preprint arXiv:2305.07141*, 2023.
- [8] J. Park, J. Im, S. Hwang, M. Lim, S. Ualibekova, S. Kim, S& S. Kim, "Unraveling the ARC Puzzle: Mimicking Human Solutions with Object-Centric Decision Transformer." *arXiv preprint arXiv:2306.08204*, 2023.
- [9] L. Chen, K. Lu, A. Rajeswaran, K. Lee, A. Grover, M. Laskin, ... & I. Mordatch, "Decision transformer: Reinforcement learning via sequence modeling.", *Advances in neural information processing systems*, Vol. 34, pp. 15084-15097, 2021.
- [10] K. Ellis, C. Wong, M. Nye, M. Sablé-Meyer, L. Morales, L. Hewitt, ... & J. B. Tenenbaum, "Dreamcoder: Bootstrapping inductive program synthesis with wake-sleep library learning.", In *Proceedings of the 42nd acm sigplan international conference on programming language design and implementation*, PLDI, pp. 835-850, 2021.
- [11] S. Alford, A. Gandhi, A. Rangamani, A. Banburski, T. Wang, S. Dandekar, ... & P. Chin, "Neural-Guided, Bidirectional Program Search for Abstraction and Reasoning.", In *Complex Networks & Their Applications X: Volume 1, Proceedings of the Tenth International Conference on Complex Networks and Their Applications COMPLEX NETWORKS 2021 10*. Springer International Publishing, pp. 657-668, 2022.
- [12] S. Acquaviva, Y. Pu, Y. M. Kryven, T. Sechopoulos, T. C. Wong, G. Ecanow, ... & J. Tenenbaum, "Communicating natural programs to humans and machines.", *arXiv preprint arXiv:2106.07824*, 2021.
- [13] S. Mirchandani, F. Xia, P. Florence, P. B. Ichter, D. Driess, M. G. Arenas, ... & A. Zeng, "Large language models as general pattern machines." *arXiv preprint arXiv:2307.04721*, 2023.
- [14] D. Huang, D. Z. Nan, Z. X. Hu, P. Jin, S. Peng, Y. Wen, ... & Y. Chen, "ANPL: Compiling Natural Programs with Interactive Decomposition." *arXiv preprint arXiv:2305.18498*, 2023.
- [15] G. Camposampiero, L. Houmar, B. Estermann, J. Mathys & R. Wattenhofer, "Abstract Visual Reasoning Enabled by Language." In *Proceedings of the IEEE/CVF Conference on Computer Vision*

and *Pattern Recognition*, pp. 2642-2646, 2023.

- [16] Y. Xu, E. B. Khalil & S. Sanner. "Graphs, constraints, and search for the abstraction and reasoning corpus.", In *Proceedings of the AAAI Conference on Artificial Intelligence* Vol. 37, No. 4, pp. 4115-4122, June, 2023.
- [17] "ARCathon 2023", Available: <https://lab42.global/arcathon/> (last modified Feb 22, 2023, accessed Feb 22, 2023, Lab42)
- [18] "ARCreate", <https://lab42.global/arcreate/> (last modified Feb 22, 2023, accessed Feb 22, 2023, Lab42)



김 수 빈

2023년 한국과학기술원 전산학과 졸업 (학사). 2023년~현재 한국과학기술원 전산학과 석사과정. 관심분야는 강화 학습, 표현학습, 지식추론.



Prin Phunyaphibarn

2023년 한국과학기술원 전산학과 재학 (학사). 관심분야는 강화 학습, 표현학습, 지식추론.



안 동 현

2018년 KAIST 전산학부 공학사(수리 과학과 복수전공) 2019년~현재 KAIST 전산학부(석박통합과정). 관심분야는 컴퓨터 비전, 표현학습, 지식추론.



김 선 동

2019년 KAIST 지식서비스공학대학원 (공학박사). 2019~2022년 기초과학연구원 차세대연구리더. 2022년~현재 광주과학기술원 AI대학원 조교수. 관심분야는 강인공지능을 위한 표현 학습, 지식 추론, 예측 분석