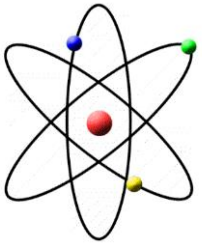
ScienceWorld: Is your Agent Smarter than a 5th Grader?

Wang, Jansen, Côté, and Ammanabrolu (2022)

Presented by Marcus McAllister
University of Maryland, Baltimore County
CMSC 691 (Interactive Fiction and Text Generation)
November 11, 2025

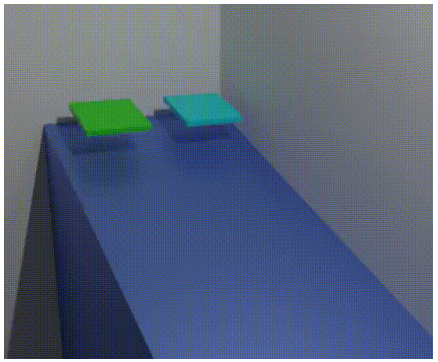
UMBC

We’re going to talk about



SCIENCE!!!

(at least a little bit)



What steps would you go through to demonstrate…

- …whether a battery still has electrical charge?
- …separating aluminum from iron?
- …identifying a plant species from a seed?
- …which of two objects has greater mass?

ScienceWorld is a text-based virtual environment for testing agent scientific reasoning capabilities.

Image sourced from
https://sciworld.apps.allenai.org/

The environment consists of ten locations and dozens of objects.

Locations

- Art studio
- Bathroom
- Bedroom
- Foundry
- Greenhouse
- Hallway
- Kitchen
- Living Room
- Outside
- Workshop

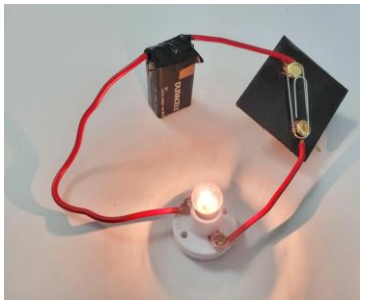
Objects

- Bookcase
- Book
- Tin cup
- Glass jar
- Light bulb
- Motor
- Wire
- Thermometer
- Stove
- Stopwatch
- And many more!

There are a total of 30 tasks available for testing across a range of scientific topics.

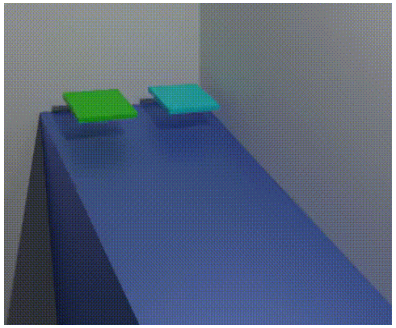
Boiling water!

Mixing chemicals!



Building circuits!

Observing life processes!



Testing friction!

And more!

ScienceWorld runs start with a task description and details of the starting location.

```
Task Description: Your task is to boil water. For compounds without a boiling point, combusting
 the substance is also acceptable. First, focus on the substance. Then, take actions that will
cause it to change its state of matter.

This room is called the hallway. In it, you see:
        the agent
        a substance called air
        a picture
You also see:
        A door to the art studio (that is closed)
        A door to the bedroom (that is closed)
        A door to the greenhouse (that is closed)
        A door to the kitchen (that is closed)
        A door to the living room (that is closed)
        A door to the workshop (that is closed)
Reward: 0
Score: 0
isCompleted: False
'help' lists valid action templates, 'objects' lists valid objects, use <tab> to list valid act
ions.
'goals' lists progress on subgoals.
type 'exit' to quit.
Valid Actions: ['connect agent to air', 'connect agent to art studio', 'connect agent to art st
udio door', 'connect agent to bedroom', 'connect agent to bedroom door', 'connect agent to door
 to greenhouse', 'connect agent to door to kitchen', 'connect agent to door to living room', 'c
onnect agent to door to workshop', 'connect agent to greenhouse', 'connect agent to hallway', '
connect agent to inventory', 'connect agent to kitchen', 'connect agent to living room', 'conne
```

An agent is expected to perform actions in the environment towards accomplishing the overall task.

```
> open door to kitchen

The door is now open.
Reward: 0
Score: 0
isCompleted: False
'help' lists valid action templates, 'objects' lists valid objects, use <tab> to list valid act
ions.
'goals' lists progress on subgoals.
type 'exit' to quit.
Valid Actions: ['close door to kitchen', 'connect agent to air', 'connect agent to art studio',
 'connect agent to art studio door', 'connect agent to bedroom', 'connect agent to bedroom door
', 'connect agent to door to greenhouse', 'connect agent to door to kitchen', 'connect agent to
 door to living room', 'connect agent to door to workshop', 'connect agent to greenhouse', 'con
```

Strengths

ScienceWorld goes beyond simple question-answering; Agents must demonstrate steps involved.

Question-Answering

Q: What will happen to an ice cube when placed on a stove?

A: It will melt

Demonstration

To melt an ice cube:

1. Acquire ice cube.
2. Acquire cooking pot.
3. Place ice cube in pot.
4. Activate stove.
5. Place pot onto stove.
6. Wait until ice cube melts.

ScienceWorld permits a variety of solutions to address tasks.

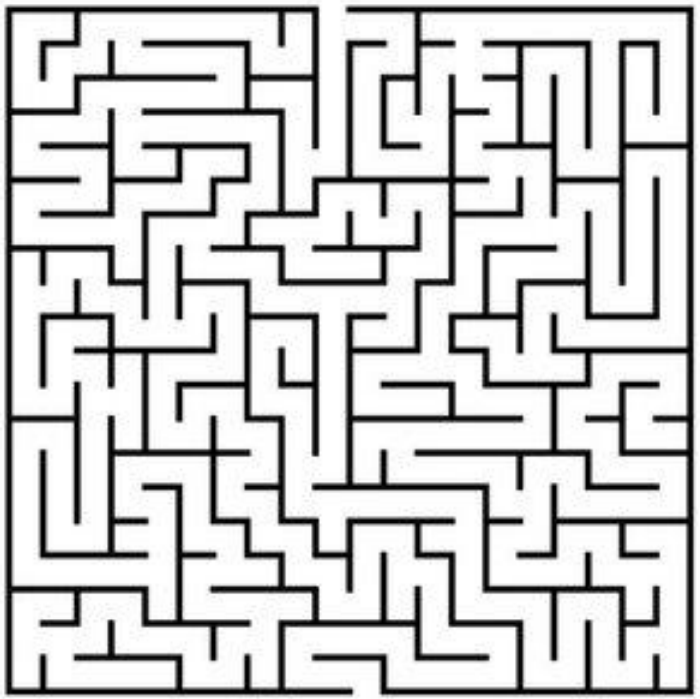
Certain tasks may require time to pass to resolve.

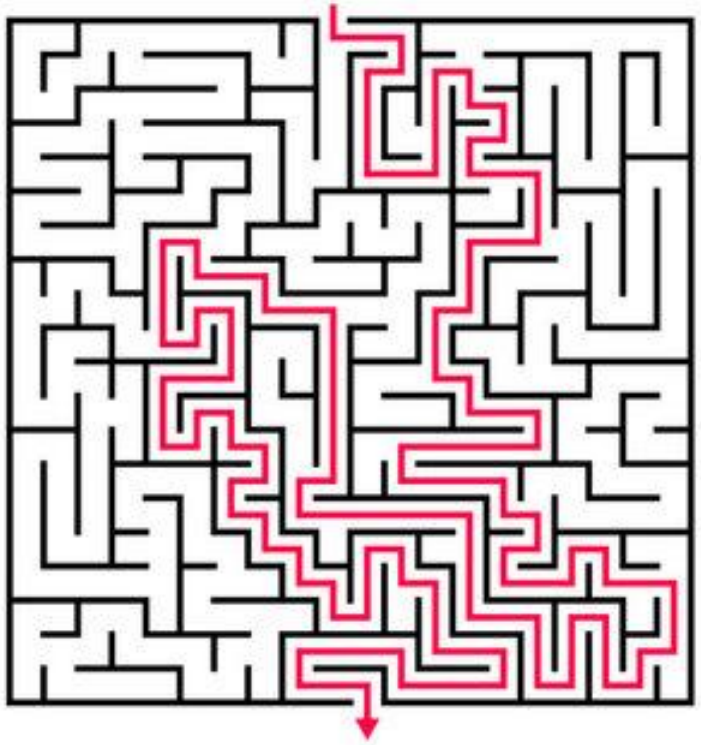
wait1

wait10

Weaknesses

ScienceWorld has limited variability in its environment.



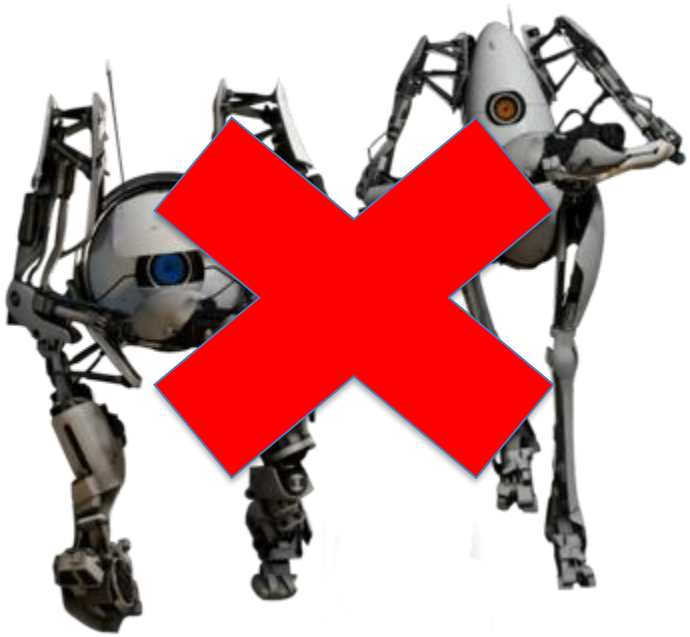


ScienceWorld has a limited action space.

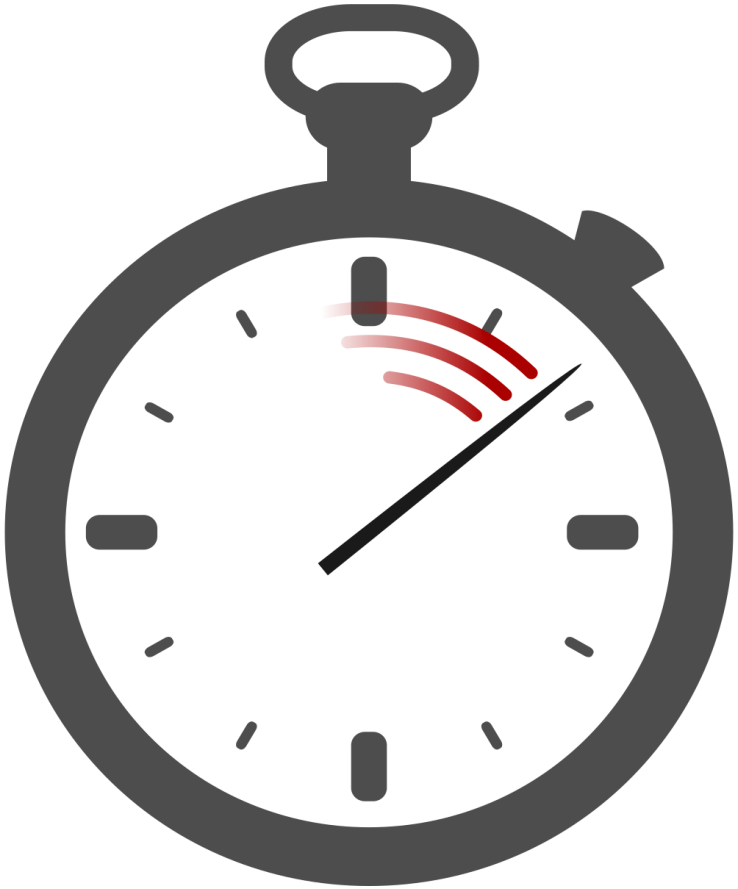
Action	Description
`open/close OBJ`	open/close a container
`de/activate OBJ`	activate/deactivate a device
`connect OBJ to OBJ`	connect electrical components
`disconnect OBJ`	disconnect electrical components
`use OBJ [on OBJ]`	use a device/item
`look around`	describe the current room
`look at OBJ`	describe an object in detail
`look in OBJ`	describe a container's contents
`read OBJ`	read a note or book
`move OBJ to OBJ`	move an object to a container
`pick up OBJ`	move an object to the inventory
`put down OBJ`	drop an inventory item
`pour OBJ into OBJ`	pour a liquid into a container
`dunk OBJ into OBJ`	dunk a container into a liquid
`mix OBJ`	chemically mix a container
`go to LOC`	move to a new location
`teleport to LOC` *	teleport to a specific room
`eat OBJ`	eat a food
`flush OBJ`	flush a toilet
`focus on OBJ`	signal intent on a task object
`wait [DURATION]`	take no action for some duration
`task`	describe current task
`inventory`	list agent's inventory

Image sourced from
https://sciworld.apps.allenai.org/

ScienceWorld only supports testing for one agent per environment instance.



Success criteria defined solely by steps to task completion.



Key takeaways

- ScienceWorld is text-based virtual environment for evaluating AI reasoning
- Various scientific tasks available for testing
- Models performed poorly in this environment.

Image sourced from
https://sciworld.apps.allenai.org/

References

Wang, R., Jansen, P., Côté, M.-A., & Ammanabrolu, P. (2022). ScienceWorld: Is your Agent Smarter than a 5th Grader? Retrieved from https://arxiv.org/abs/2203.07540.

Questions?