

User Interface Software Technology :Discussion

Presenters

Shenshen Han

Hana Cho

Generative Agents: Interactive Simulacra of Human Behavior

Joon Sung Park et al.
2023
Stanford University



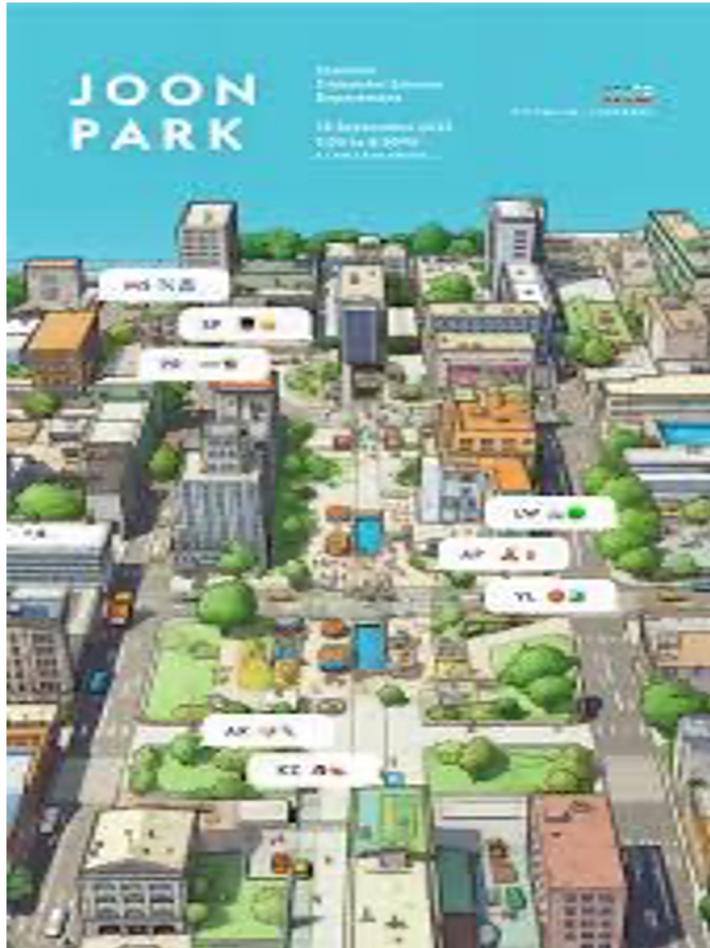
Words to define

- **Simulacra**: Something that replaces reality with its representation
- **Proxies**: Agent authorized to act on behalf of a user
- **LLM** (Large Language Model): advanced artificial intelligence system trained on vast amounts of text data, capable of understanding and generating human language
- **NLP** (Natural Language Processing)
 - Human language
 - Technology that gives computers ability to interpret, manipulate, comprehend human language

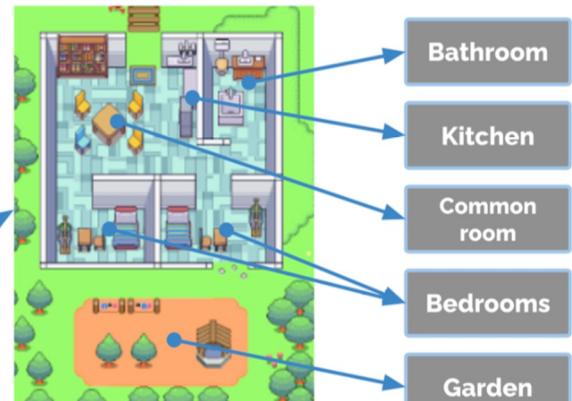


The SIMS

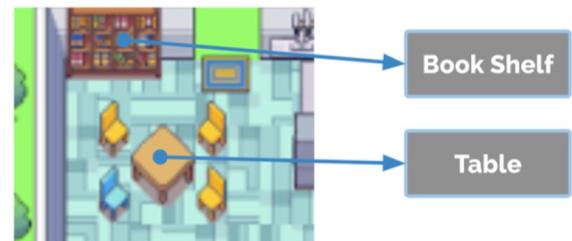




<https://www.youtube.com/watch?v=ZdoU9vI2yCg&pp=ygUiZ2VuIGFnZW50cyBqb29uIHN1bmcgcGFyaybWcmV2aWIIdw%3D%3D>



Family House



Common Room

1. Memory stream
2. Reflection
3. Planning



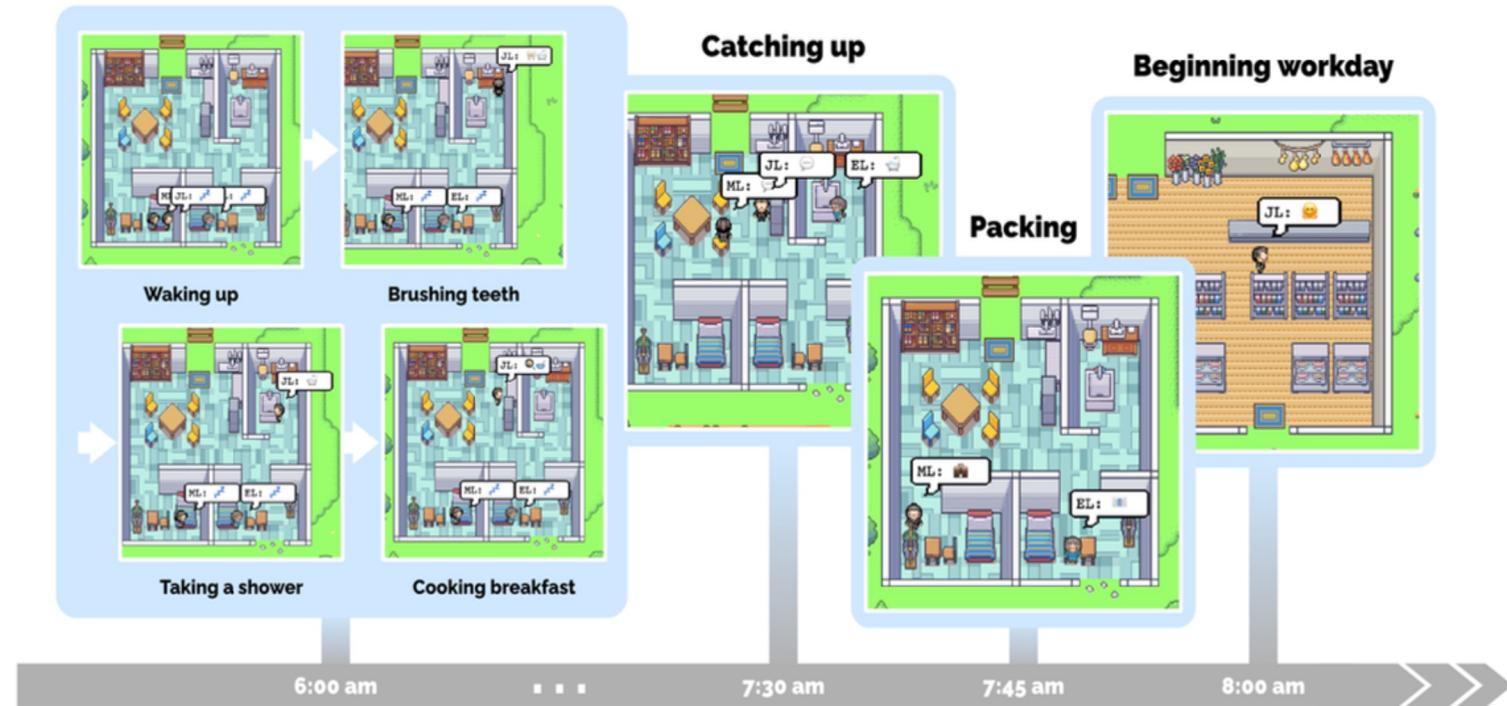
chatGPT LLM

John Lin is a pharmacy shopkeeper at the Willow Market and Pharmacy who loves to help people. He is always looking for ways to make the process of getting medication easier for his customers; John Lin is living with his wife, Mei Lin, who is a college professor, and son, Eddy Lin, who is a student studying music theory; John Lin loves his family very much; John Lin has known the old couple next-door, Sam Moore and Jennifer Moore, for a few years; John Lin thinks Sam Moore is a kind and nice man; John Lin knows his neighbor, Yuriko Yamamoto, well; John Lin knows of his neighbors, Tamara Taylor and Carmen Ortiz, but has not met them before; John Lin and Tom Moreno are colleagues at The Willows Market and Pharmacy; John Lin and Tom Moreno are friends and like to discuss local politics together; John Lin knows the Moreno family somewhat well – the husband Tom Moreno and the wife Jane Moreno.

Environmental Interaction

- Agents enter and leave the buildings, navigate, and approach other agents
- Initiate interactions, remember its behavior
- Users and agents can **influence the objects** in Smallville

Morning routine



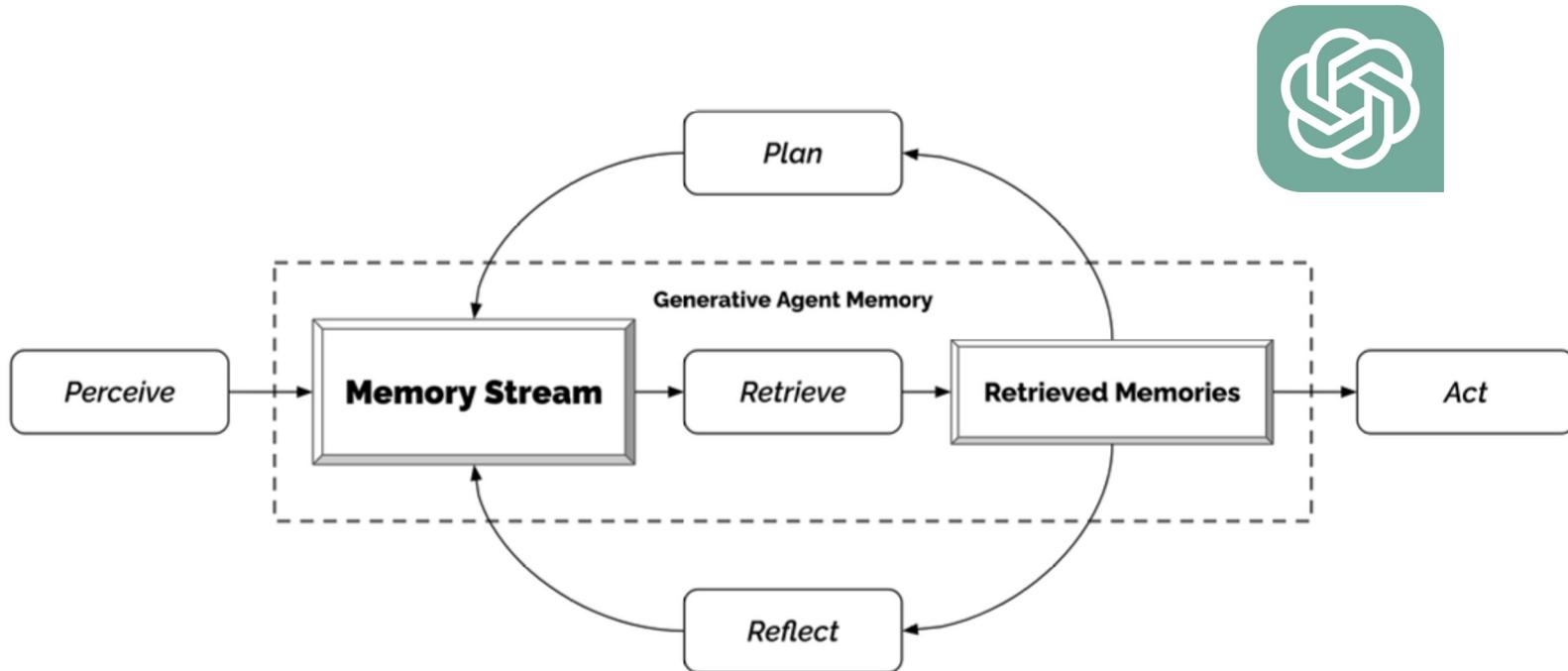
John Lin: "Hey Eddy, how's the music composition project for your class coming along?"

Eddy Lin: "Hey Dad, it's going well. I've been taking walks around the garden to clear my head and get some inspiration."

Evaluation



- 1. Agent interviews**
- 2. Evaluation conditions**
- 3. Human evaluators**
- 4. Statistical analysis**
- 5. Qualitative analysis**



Memory Stream

```
2023-02-13 22:48:20: desk is idle  
2023-02-13 22:48:20: bed is idle  
2023-02-13 22:48:10: closet is idle  
2023-02-13 22:48:10: refrigerator is idle  
2023-02-13 22:48:10: Isabella Rodriguez is stretching  
2023-02-13 22:33:30: shelf is idle  
2023-02-13 22:33:30: desk is neat and organized  
2023-02-13 22:33:10: Isabella Rodriguez is writing in her journal  
2023-02-13 22:18:10: desk is idle  
2023-02-13 22:18:10: Isabella Rodriguez is taking a break  
2023-02-13 21:49:00: bed is idle  
2023-02-13 21:48:50: Isabella Rodriguez is cleaning up the kitchen  
2023-02-13 21:48:50: refrigerator is idle  
2023-02-13 21:48:50: bed is being used  
2023-02-13 21:48:10: shelf is idle  
2023-02-13 21:48:10: Isabella Rodriguez is watching a movie  
2023-02-13 21:19:10: shelf is organized and tidy  
2023-02-13 21:18:10: desk is idle  
2023-02-13 21:18:10: Isabella Rodriguez is reading a book  
2023-02-13 21:03:40: bed is idle  
2023-02-13 21:03:30: refrigerator is idle  
2023-02-13 21:03:30: desk is in use with a laptop and some papers on it
```

...



Q. What are you looking forward to the most right now?

Isabella Rodriguez is excited to be planning a Valentine's Day party at Hobbs Cafe on February 14th from 5pm and is eager to invite everyone to attend the party.

retrieval	=	recency	importance	relevance
2.34	=	0.91	+ 0.63	+ 0.80

ordering decorations for the party

2.21	=	0.87	+ 0.63	+ 0.71
------	---	------	--------	--------

researching ideas for the party

2.20	=	0.85	+ 0.73	+ 0.62
------	---	------	--------	--------

...

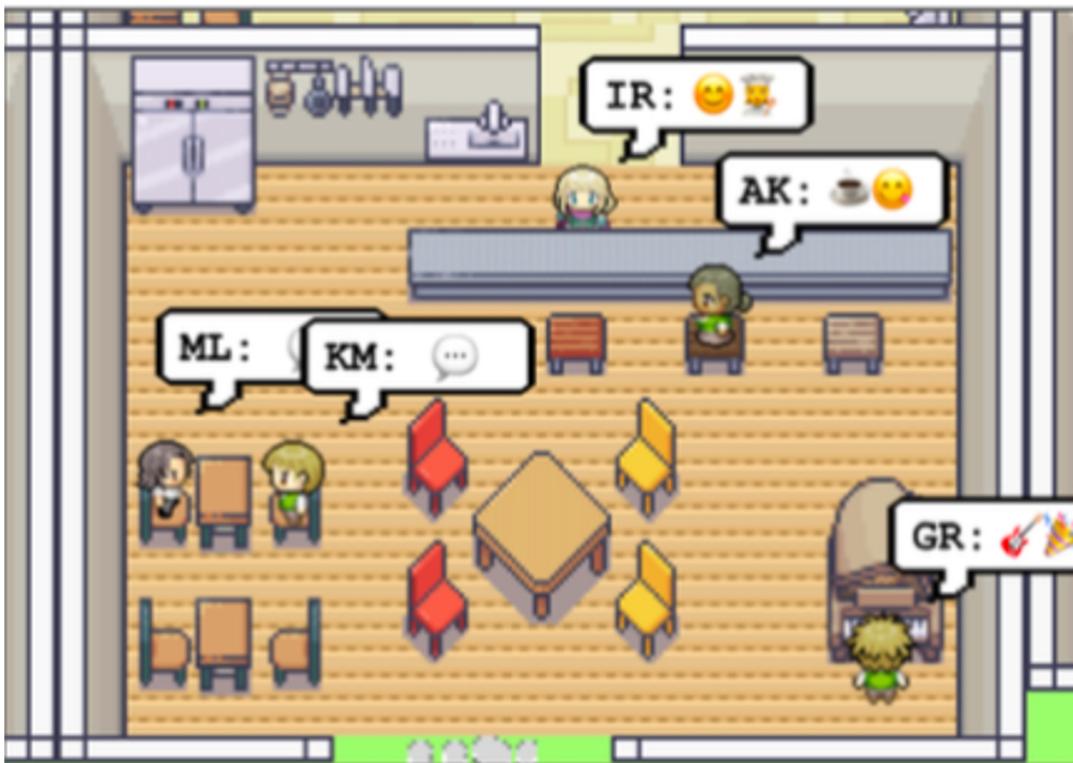


I'm looking forward to the Valentine's Day party that I'm planning at Hobbs Cafe!



Isabella

Valentine's Day party



Rajiv:

"No, I
don't think so. I'm focusing
on my upcoming show,
and I don't
really have time to make
any plans for Valentine's
Day."

I'm planning a Valentine's Day party at Hobbs Cafe on February 14th from 5pm to 7pm!

Isabella

I heard you're planning a Valentine's Day party at Hobbs Cafe. That sounds like a great event!

Tom

Giorgio

Eddy

John

Wolfgang

Sam

Ayesha

Klaus

Latoya

Abigail

Speaking of which, Isabella has invited us to a Valentine's Day party at Hobbs Cafe.

Jennifer

Maria

I'm organizing a reading night for the Valentine's Day party at Hobbs Cafe. You're welcome to join.

I'm actually planning on attending Isabella Rodriguez's Valentine's Day party as well, so maybe we could talk more then!

Election discussion

Isabella: I'm still weighing my options, but I've been discussing the election with Sam Moore. What are your thoughts on him?

Tom: To be honest, I don't like Sam Moore. I think he's out of touch with the community and doesn't have our best interests at heart.

Perfect?

1. Agents failed to retrieve memories
2. Overly formal speech or behavior from the language model
3. Misclassification of social norms

Discussion

- Human-like behavior in interactive environments
- Simulations of **human behavior** to prototype social systems /social theories, human societies
- HCI, AI, Cog Sci, Ethics
- User interaction models & interactive system design
- Versatility and depth of generative agents

Discussion Question

[The assessment of gen agents' behavior in this study was limited to a relatively short time scale and baseline human crowdworker condition]

- 1. If you were the researcher to conduct this study, how would you improve the performance of these gen agents**
- 2. and evaluate them in terms of believability?**

Ethical concerns

1. People forming parasocial relationships with gen agents
2. Impact of errors: wrong inferences
3. Deep fakes, misinformation, tailored persuasion
4. Over-reliance

Human-Computer Interaction Contribution

?

HCI contribution

Interactivity and user engagement

User experience design

Social dynamics in virtual communities

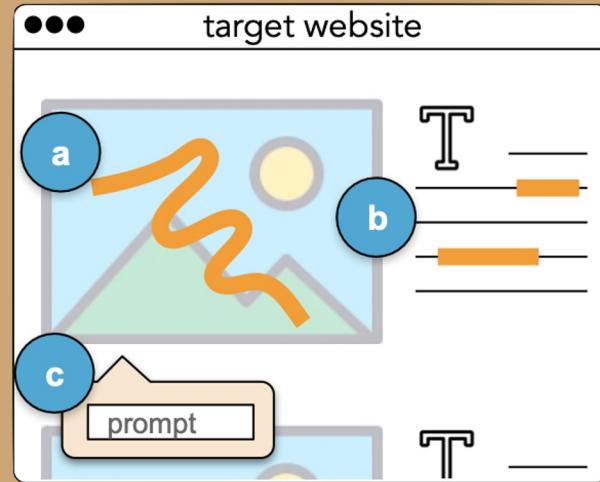
Evaluating AI interaction

Takeaways for HCI community

1. **New framework** for creating more realistic and dynamic agents in various applications
1. **Novel design** interactive system

SemanticOn: Specifying Content-Based Semantic Conditions for Web Automation Programs

Kevin Pu et al
UIST 2022
Honorable Mention



Catchy Title?

- SemanticOn: Specifying Content-Based Semantic Conditions for Web Automation Programs (ChatGPT)
 - “SemanticOn” → Catchy
 - “Specifying Content-Based Semantic” → Focus on “Semantic”
 - “Web Automation Programs” → Application domain

Paper Requirements from UIST

Types of Contribution

- Everything regarding artifact? (algorithms, techniques, systems, tools, hardware, devices, models and applications)

Paper Requirements from UIST

Amount of New Content

- Novelty Over Existing Archival Work (i.e., previously published papers): a paper has to include about 70% new material to be considered a novel contribution over existing archival work.
- Novelty Over Existing Non-Archival Work (i.e., previously shown demos, posters): a paper has to include about 30% new material to be considered a novel contribution over existing non-archival work.

Guidelines from UIST

- What problem are you solving?
- What were the previous solutions?
- How well did you solve your problem?
- What does this work contribute to the field?
- Is the paper complete?

Quick Recap

The screenshot shows the Microsoft Power Automate designer interface. At the top, there are three green bars with labels: 'My flows (1)', 'My connectors (1)', and 'My triggers (1)'. Below these is a blue bar labeled 'Condition Panel'. Underneath the blue bar are two buttons: 'My connectors' and 'My triggers'. A message box says 'Select an automation from the next page'. The main area is titled 'Condition Panel' and contains a table:

Action	Condition	Else
Get item	A flow (single step) (Condition-based)	✓ (Edit, Delete)
Get item	A flow (multiple steps) (Condition-based)	✓ (Edit, Delete)

Below the table is another section titled 'Action Panel' with three buttons: 'My connectors', 'My triggers', and 'My flows'. At the bottom left is a 'Run' button with a play icon. On the right side, there is a 'Copilot AI' icon and a 'Feedback' link.

Quick Recap

- What problem are you solving?
 - Web Automation
- What were the previous solutions?
 - Yes, but...
- How well did you solve your problem?
 - 80.7% Accuracy, Survey (Likert scale)
- What does this work contribute to the field?
 - Human-AI Collaboration, Programming-by-Demonstration/Example (End-User Programming)
- Is the paper complete?
 - Yes, I guess...

Picky Committee



Say something bad...

- ***What problem are you solving?***

Data scientists, researchers, and clerks often create web automation programs to perform repetitive yet essential tasks, such as data scraping and data entry. However, existing web automation systems lack mechanisms for defining conditional behaviors where

We identified a need for web automation with semantic conditions through prior user studies [24] and analysis of real user requests in online forms [3, 7]. This includes vision-related seman-

Performing these tasks manually can often lead to human errors (e.g., data duplicates, missed entries), which can cause inefficiencies.

Say something bad...

- ***What were the previous solutions?***

Past research has developed techniques to help users of all expertise levels to quickly and accurately create their intended web automation programs [25–27, 44, 68]. However, these techniques are limited to creating programs with requirements at the website syntax or structural level (e.g., scraping the first two items in each row of a table). Tools capable of creating logic based on the meaning of the content (semantics) remain unexplored. For instance,

structure) and programming experience. Commonly used tools like Puppeteer [4], Selenium [6], Scrapy [5], and Beautiful Soup [1] require users to learn code syntax, understand the task content architecture (e.g., DOM tree hierarchy), and have software test-

Using machine learning (ML) models to refine intent has been a recent focus in the field of interactive ML. One common interactive

Say something bad...

- ***How well did you solve your problem?***

Question	Likert Scale (Mean, SD)
User Enters Experience	5.3 (1.0)
System Suggests Experience	5.1 (1.2)
User Enters Ease of Use	5.1 (1.0)
System Suggests Ease of Use	5.7 (0.8)
Coordination to Refine Specifications	4.7 (1.6)
Usefulness of Generated Prompts	5.6 (0.9)
Conditions Displayed Clearly	6.0 (1.0)
Trust in Full Automation	4.5 (1.5)
Success in Completing Task	5.0 (0.9)

Question	Likert Scale (Mean, SD)
Mental Demand	4.6 (1.7)
Effort to Achieve Results	4.9 (1.3)
Feelings of Insecurity and Stress	5.2 (1.7)
Feelings of Being Hurried or Rushed	4.9 (1.6)

Table 2: Survey Responses. For section one (top), 1 is very negative, and 7 is very positive. For section two (bottom), 1 is very high mental demand, effort, insecurity and stress, and feelings of being hurried and rushed.

Task index	Completion time (mean, SD in mm:ss)	Accuracy
1	05:06 (01:43)	85.3%
2	06:58 (02:28)	68.9%
3	06:44 (01:50)	81.3%
4	05:56 (01:50)	86.0%

Table 1: Average time spent on each task. Tasks 1 and 3 are image tasks; tasks 2 and 4 are text tasks.

We recruited 10 people (5F5M, mean age 24.3, mean coding experience 4.6 years) at a large public university. Participants are denoted

Say something bad...

- ***What does this work contribute to the field?***

One emphasis of this project is to compare the usability of the two semantic condition specification methods. To summarize the findings, participants shifted their preference between *System Suggests* and *User Enters* based on two factors: perceived effort and the type of content presented. When participants expressed their prefer-

A main contribution of SemanticOn is introducing a new interaction paradigm for users to continuously add/refine semantic conditions in programming-by-demonstration (PBD) systems. In particular, it

Suggests regarding user effort and the sense of control. Our work can point directions to future system and interaction designs for user-intent specification and refinement in a continuous human-AI collaboration setting.

Say something bad...

- ***Is the paper complete?***

We implemented SemanticOn as a Chrome browser extension, incorporating the core program synthesis engine from an existing system, WebRobot. Primarily, it uses plain JavaScript for the front-end interactions. For semantic condition comprehension, we adopted two off-the-shelf Transformer models.

to add to the condition table (Fig. 4.f,g,h). We used these models through the Microsoft Azure Cognitive Services and Cloud platform.² Guided by the design of other systems [17, 31], we also implemented edit, delete, and logical operations for multiple condition specifications, allowing the user to modify, remove, or set the AND/OR logic switch on the conditions (Fig. 4.e). They can also

mouse (Fig. 6.a) to illustrate. The image is fed through an off-the-shelf pre-trained multi-layer Transformer model that learns to align image-level tags with their corresponding image region features [32]. The model detects objects for the illustrated section and generates a caption for the entire image. For texts, the user highlights relevant phrases or sentences with their mouse (Fig. 6.b). Similarly, the text is processed by an off-the-shelf unsupervised language model where the noun phrases in the input text are first detected and then ranked based on frequency and co-occurrence.

We did it!

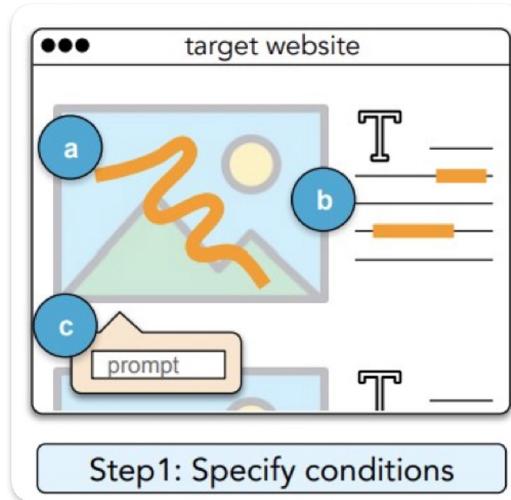


Why not CHI

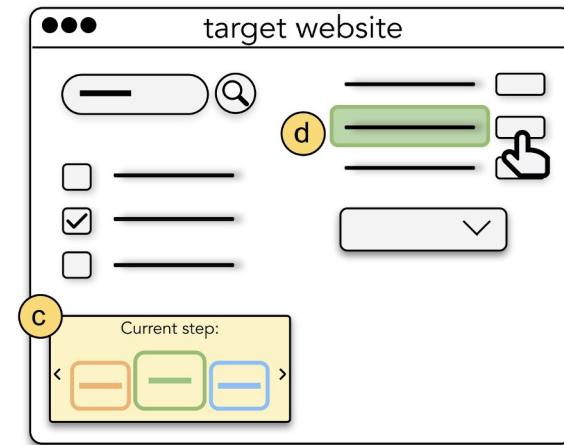
- Developing Novel Devices: Hardware, Materials, and Fabrication
- Interacting with Devices: Interaction Techniques & Modalities
- Blending Interaction: Engineering Interactive Systems & Tools

My Final Doubt

SemanticOn: Specifying Content-Based Semantic Conditions for Web Automation Programs. (UIST '22).



DiLogics: Creating Web Automation Programs with Diverse Logics. (UIST '23).



My Final Doubt

SemanticOn: Specifying Content-Based Semantic Conditions for Web Automation Programs. (UIST '22).

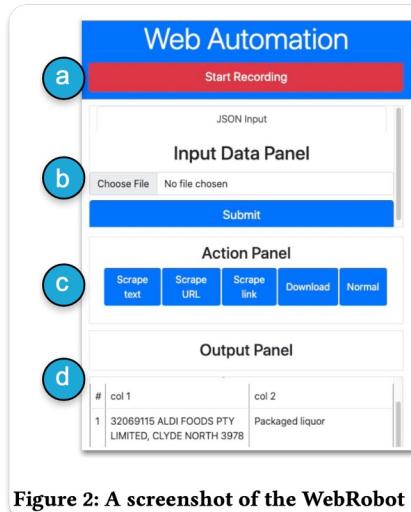


Figure 2: A screenshot of the WebRobot system UI.

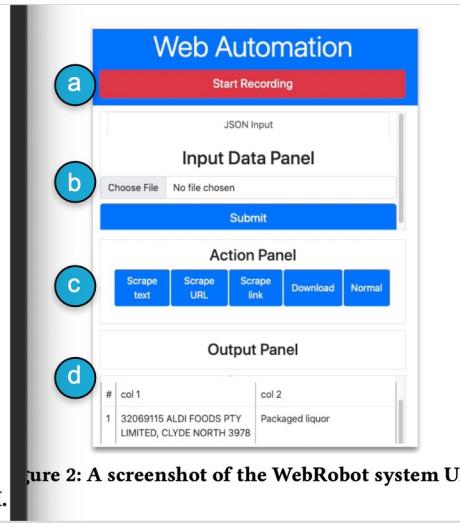


Figure 2: A screenshot of the WebRobot system UI.

DiLogics: Creating Web Automation Programs with Diverse Logics. (UIST '23).

My Final Doubt

SemanticOn: Specifying Content-Based Semantic Conditions for Web Automation Programs. (UIST '22).

ABSTRACT

Data scientists, researchers, and clerks often create web automation programs to perform repetitive yet essential tasks, such as data scraping and data entry. However, existing web automation sys-

7 CONCLUSION

In this work, we designed and developed SemanticOn, a collaborative system that allows users to specify and refine visual and textual conditions through user-entered descriptions and system-suggested prompts in a web automation program. In a system evaluation, we

DiLogics: Creating Web Automation Programs with Diverse Logics. (UIST '23).

ABSTRACT

Knowledge workers frequently encounter repetitive web data entry tasks, like updating records or placing orders. Web automation in-

8 CONCLUSION

To support creating web automation with diverse specifications, we designed and developed DiLogics, a PBD tool that assists users in segmenting task requests and synthesizes programs based on user demonstration of example steps. The steps are mapped to

Q & A

THANK YOU
