Eliciting Model Steering Interactions via Data and Design Probes

Anamaria Crisan
University of Waterloo
ana.crisan@uwaterloo.ca

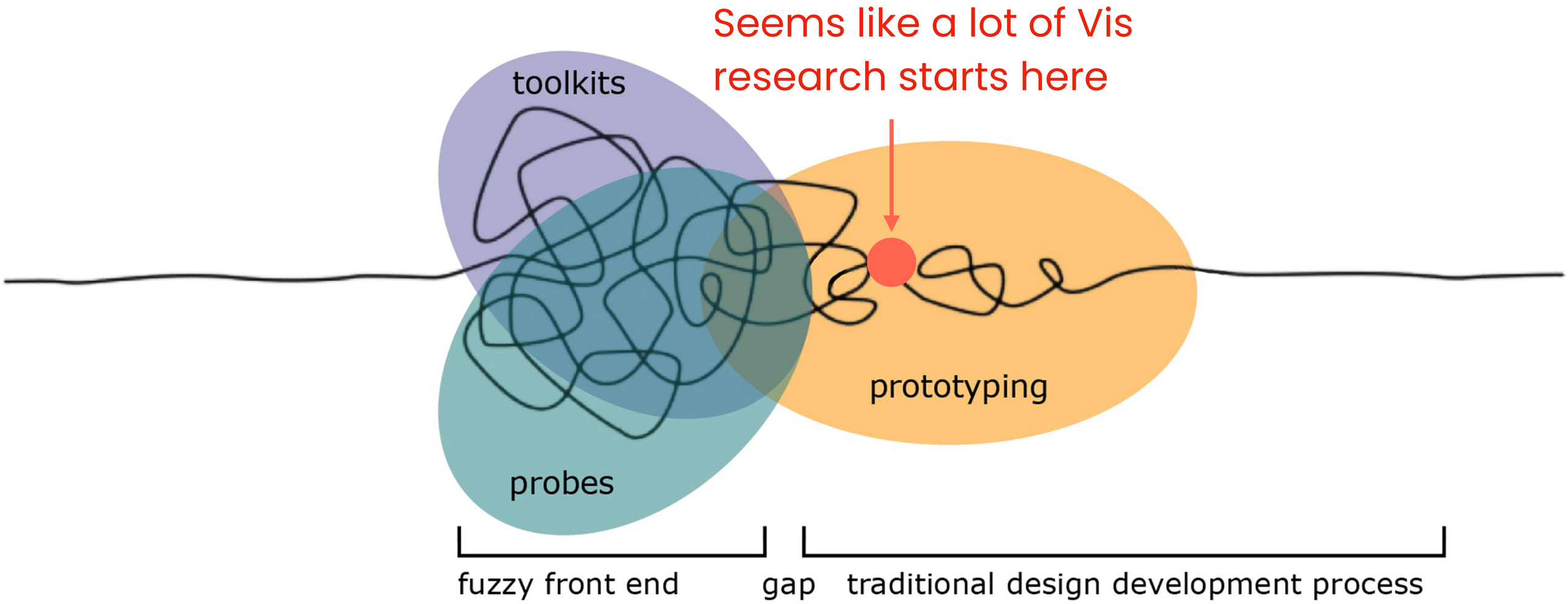
Maddie Shang
Tableau Software, USA

Eric Brochu
Tableau Software, USA

Its Inherently Hard to Design for Human-AI Interaction (HAI)

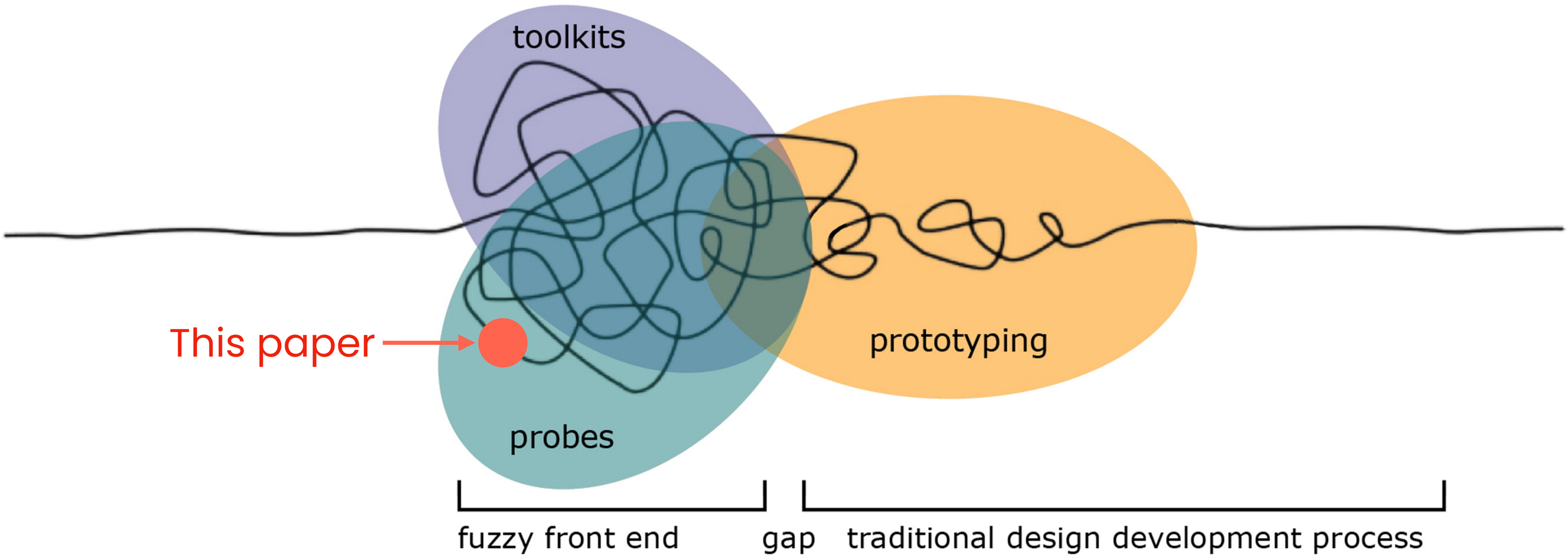
- We **propose** the idea of using **data** and **design probes** to help
- We **reify** this idea in a **co-design study** with participants
- We **identify** effects of interaction + encodings on HAI
- We **summarize** research challenges on **semantic interaction x HAI**

We Make A Lot of Assumptions About HAI



Sanders et. Al. (2014) Probes, Toolkits and Prototypes: three approaches to making in codesgning.

Probes Reveal User's Preferences *before Prototyping*



Sanders et. Al. (2014) Probes, Toolkits and Prototypes: three approaches to making in codesgning.

Probes Reveal User's Preferences *before Prototyping*

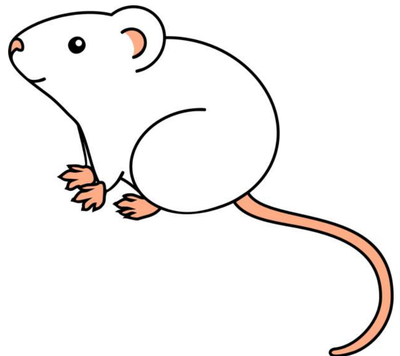
- The objectives of our data and design probes is to elicit model steering interactions from a diverse group of data workers

Data Probes

- A tool for **encouraging divergent thinking** about the behaviours on AI/ML systems
- **Controls for dataset effects** (e.g., familiarity, specific characteristics)
- Provides a **common baseline**

Visual Design Probes

- A tool for **encouraging exploration and co-creation** for interacting with AI/ML systems
- **Controls for attribution errors** (e.g., AI/ML effects vs prototype)
- **Isolates effects** of specific encodings



Data Probes

- Dataset of movies sources from IMDB/Rotten Tomatoes
- We used a small number of examples (n=50)
- We introduced deliberate errors into the dataset

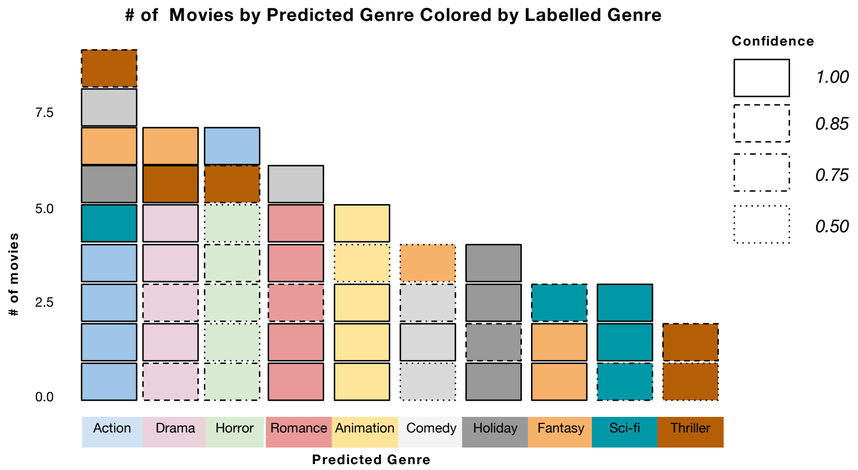
Dataset		Model Output			
Movie Title	**Labelled Genre**	**Predicted Genre**	**x**	**y**	**Probability**
Austin Powers : International Man of Mystery	Comedy	Action	0.5359921364	6.916128641	0.999
The Empire Strikes Back	Sci-fi	Action	1.200007496	5.663269463	1
Die Hard	Holiday	Action	3.579763422	9.399525235	1
Black Panther	Action	Action	2.48520129	7.134850243	1
The Terminator	Action	Action	3.577290086	6.314528258	0.999
The Dark Knight	Action	Action	0.3799107153	7.712339076	0.995
Mission Impossible	Action	Action	3.738254133	8.109444115	1

Visual Design Probes of Common Encodings

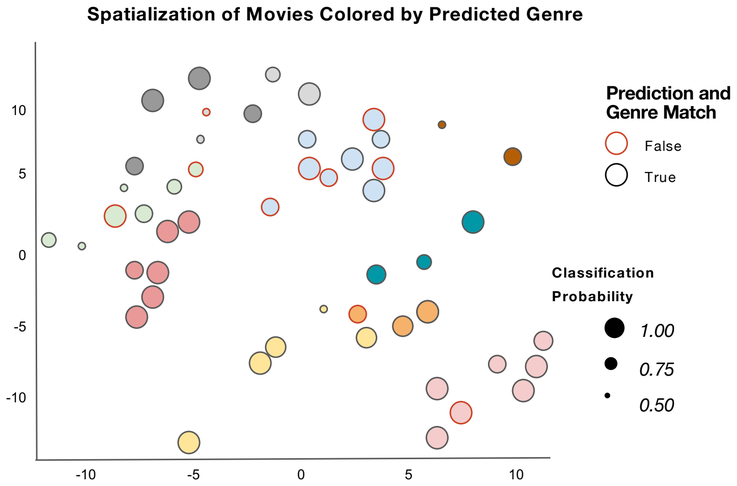
Table

Movie	Labelled Genre	Predicted Genre	Classification Probability
Austin Powers : International Man of Mystery	Comedy	Action	0.999
Skyfall	Thriller	Action	1.000
Die Hard	Holiday	Action	0.795
Alien	Horror	Horror	0.870
Home Alone	Holiday	Holiday	0.842
Romeo and Juliet	Romance	Romance	1.000
Dr. Strangelove	Comedy	Romance	1.000
Aliens	Action	Horror	0.869

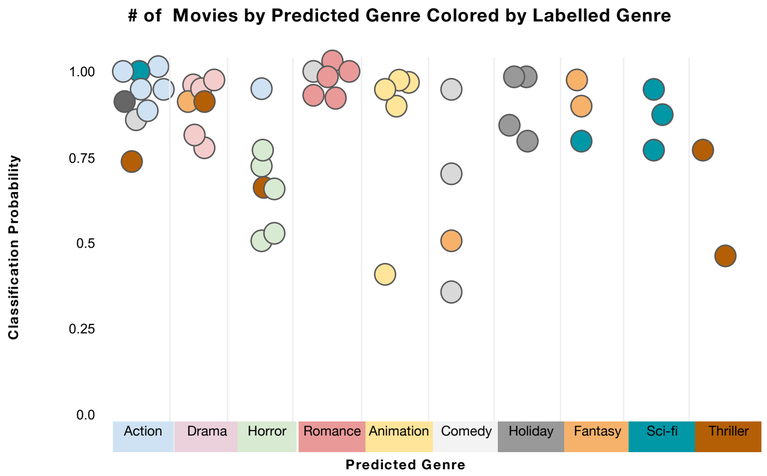
Bar Chart



Scatter Chart



Dot Chart



Visual Design Probes of Common Encodings

Data	Source		Encoding			
	Initial Data	Model Results	Table	Bar Chart	Scatter Chart	Dot Chart
Movie Title	●		Tt	💬	💬	💬
Synopsis	●		✕	✕	✕	✕
Genre - Labelled	●		Tt	■ ■ ■	✕	● ● ●
Genre - Predicted		●	Tt	Y	● ● ●	Y
Genre - Mismatch		●	Tt	Y	○ ○	Y
Probability		●	Tt	□ ⬚ ⬚	○ ◯ ◯	Y
Coordinates		●	✕	✕	Y	✕

Encoding Legend

Text : Tt Tooltip : 💬 Color: ■ ● Position: Y

Size : ○ ◯ Outline: ○ ⬚ n.a: ✕

Limitations: we made initial decisions of what was encoded

Elicitation Study Setup

We recruited 20 participants with a diverse AI/ML background

Step 1:

Background Survey

Step 2:

Instrument Prep

Step 3:

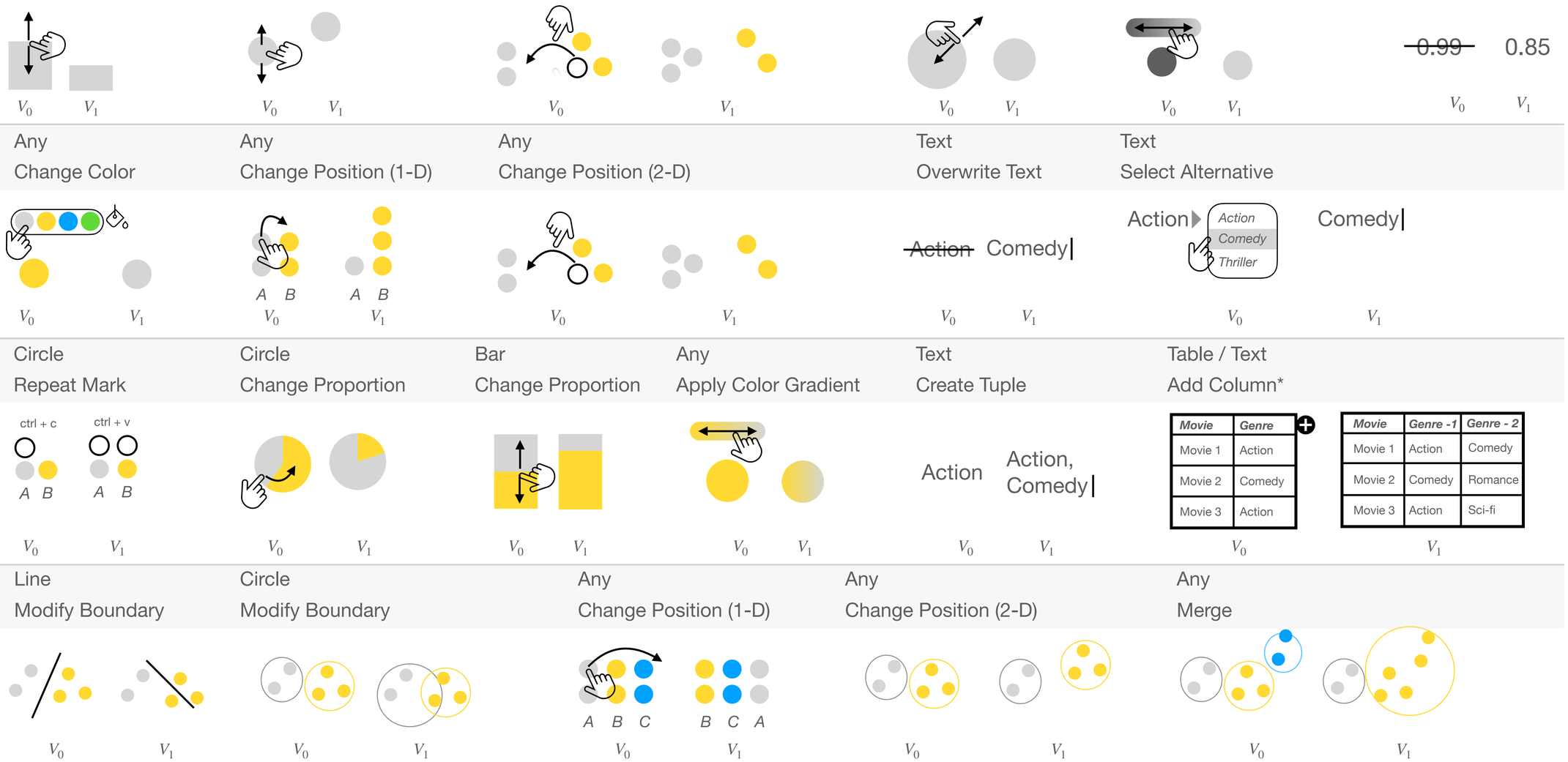
WoZ Co-design

Study Materials are available as online Supplemental Materials

Examples

Design Space of HAI for Classification Tasks

Probability	**Mark**	Bar	Circle	Any	Any	Any	Text
	Interaction	Change Height	Change Position (1-D)	Change Position (2-D)	Change Size	Change Transparency	Over write
Label (Single)	**Mark**	Any	Any	Any	Text	Text	
	Interaction	Change Color	Change Position (1-D)	Change Position (2-D)	Overwrite Text	Select Alternative	
Label (Multiple)	**Mark**	Circle	Circle	Bar	Any	Text	Table / Text
	Interaction	Repeat Mark	Change Proportion	Change Proportion	Apply Color Gradient	Create Tuple	Add Column*
Cluster	**Mark**	Line	Circle	Any	Any	Any	
	Interaction	Modify Boundary	Modify Boundary	Change Position (1-D)	Change Position (2-D)	Merge	



What are People Communicating through Interaction?

- Adding additional dimensions, not in the data

> when I think about what I am feeding the machine learning model there, it's essentially **creating another dimension** for it to run through[...] I am saying 'Here's a user dimension that I want you to consider [...] with the other dimensions that you have'

- Prioritizing the agent's attention

> Don't worry as much about getting these [others ones] right [...] which would allow me to say, **I really care about these and I don't so much care about those**

- Interactive updates are about people AND model refinements

> I would love to prototype models because that would let me refine my own notes and **recommendations to the person that is potentially designing the model**

Hesitations for Interactions

- Concerns about the introduction of bias

Admin: In the previous vis you changed the confidence via the [mark] size, do you want to do that in the table?

Participant: No *[emphatic]* and I don't know why. I mean I guess because it's numbers, don't ever change the numbers.

- Too many degrees of freedom

I think I prefer this vis [bar chart]. **There's just too much freedom in the scatter plot**

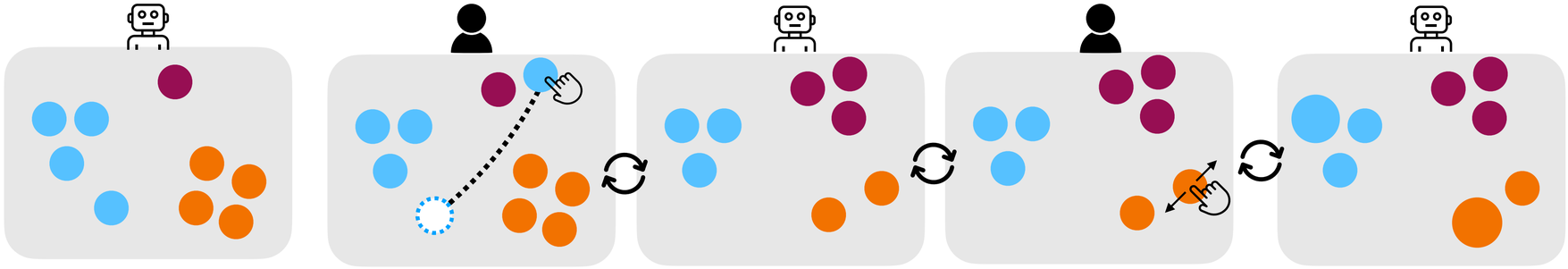
I feel like **engaging with the other charts I feel more confident** [..] because[..] when I change position [on the scatter chart] I will change the distance to all the other points."

- The encoding does not naturally support interaction

"not sure how moving things around [in the scatter plot] change anything [and I] see a **table as a natural place to provide row-level feedback**"

Frequency and Cadence of Updates

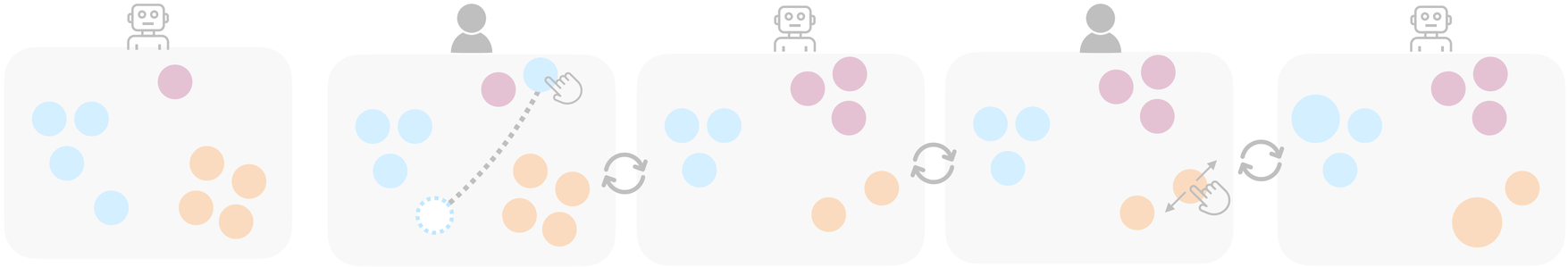
Automated Updates



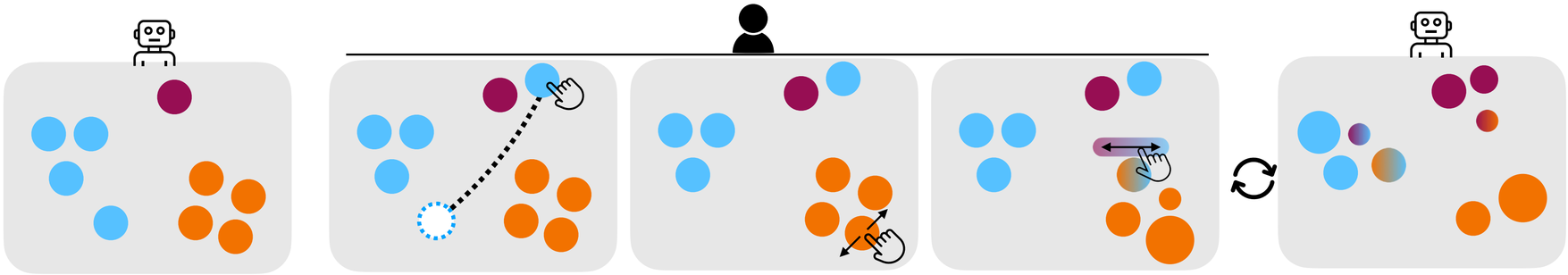
AI/ML agent decides when to respond
(immediately, or after a sequence of actions)

Frequency and Cadence of Updates

Automated Updates



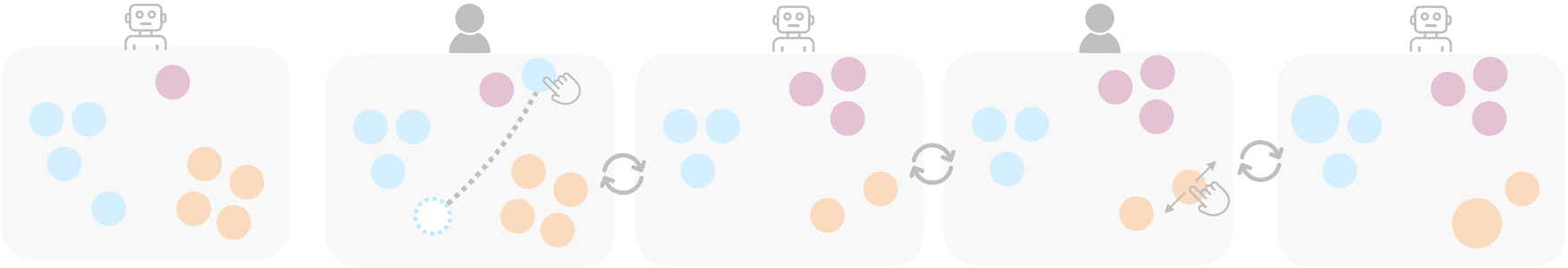
Update on Demand



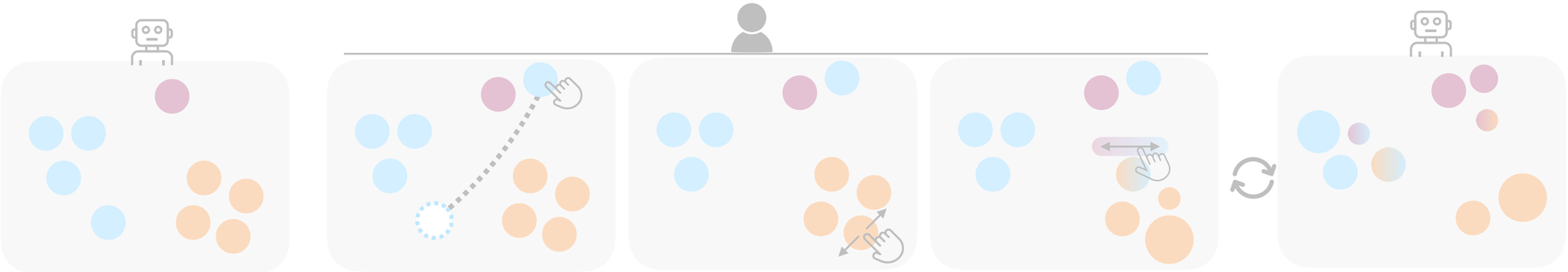
Humans explicitly tell AI/ML agent when to update

Frequency and Cadence of Updates

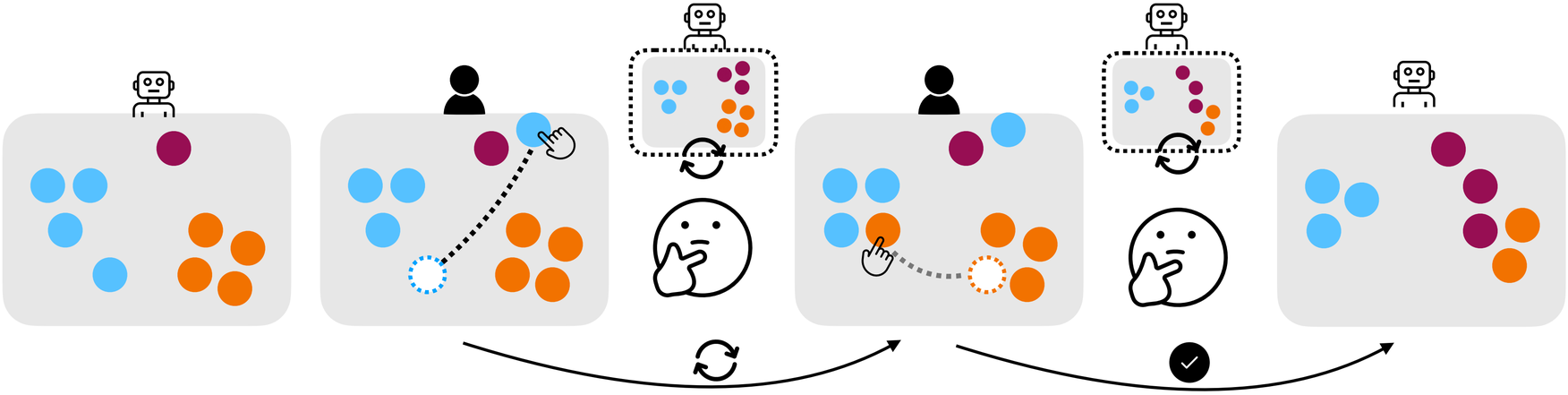
Automated Updates



Update on Demand



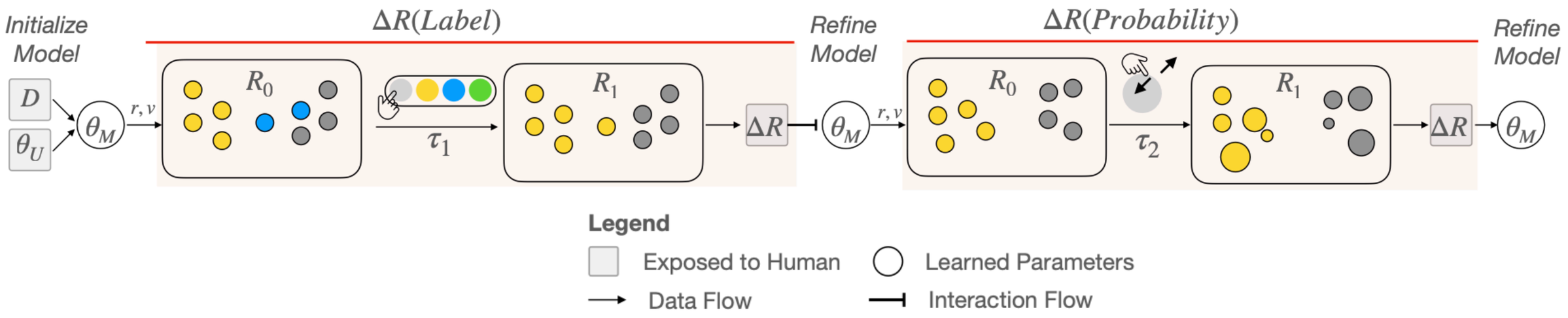
See and Choose



Lesson Learned and Next Steps

- Leading by prototyping considered harmful
- People want to do things AI/ML models may not handle
- People don't want to interact with too many data points

Next Steps: Exploring New Feedback Paradigms in Vis-mediated HAI



Eliciting Model Steering Interactions via Data and Design Probes

Anamaria Crisan
University of Waterloo
ana.crisan@uwaterloo.ca

Maddie Shang
Tableau Software, USA

Eric Brochu
Tableau Software, USA

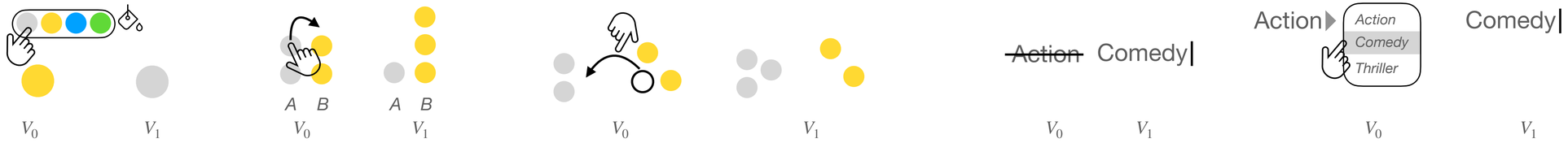
I'm Hiring PhD and Masters student at Waterloo!
Reach out if interested! Application Deadline Dec 1st.

Design Space of HAI for Classification Tasks

Participants wanted to change a marks class

(This is was not too surprising)

Label (Single)	**Mark**	Any	Any	Any	Text	Text
	Interaction	Change Color	Change Position (1-D)	Change Position (2-D)	Overwrite Text	Select Alternative



Design Space of HAI for Classification Tasks

Participants wanted to merge and define new clusters

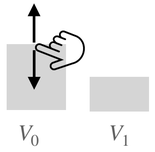
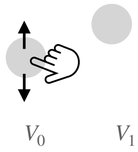
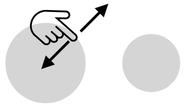
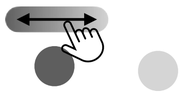
(This was also not too surprising)

Cluster	**Mark**	Line	Circle	Any	Any	Any
	Interaction	Modify Boundary	Modify Boundary	Change Position (1-D)	Change Position (2-D)	Merge

Design Space of HAI for Classification Tasks

Participants wanted to manipulate the models probability

(This is no longer a classification problem 😬)

Probability	**Mark**	Bar	Circle	Any	Any	Any	Text
	Interaction	Change Height	Change Position (1-D)	Change Position (2-D)	Change Size	Change Transparency	Over write
							~~0.99~~ 0.85
		V_0 V_1	V_0 V_1	V_0 V_1	V_0 V_1	V_0 V_1	V_0 V_1

Design Space of HAI for Classification Tasks

Participants wanted to introduce multiple classes the models certainty

(This is no longer a simple classification problem 😬)

Label (Multiple)	**Mark**	Circle	Circle	Bar	Any	Text	Table / Text
	Interaction	Repeat Mark	Change Proportion	Change Proportion	Apply Color Gradient	Create Tuple	Add Column*