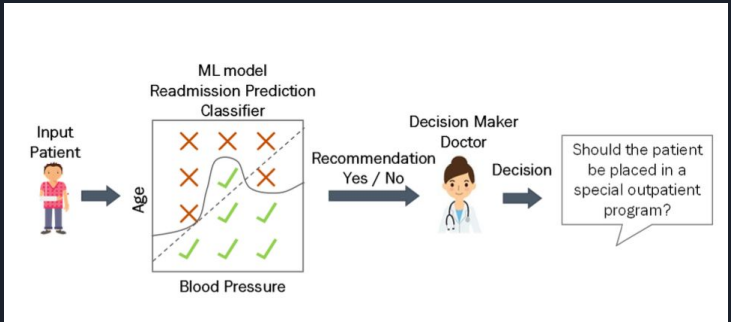
Human-AI Collaboration

Alex Bakus & Danielle Larson

Motivation for Human AI teams

- Issues around ML and AI decision making algorithms
  - Bias
  - Unfairness
- Different expertise
  - Doctor may know info missing from electronic health records.
  - AI smay have access to most recent results and trends.
- Complements => Highest team performance
  - Computer = fast computation
  - People = ethical decipher



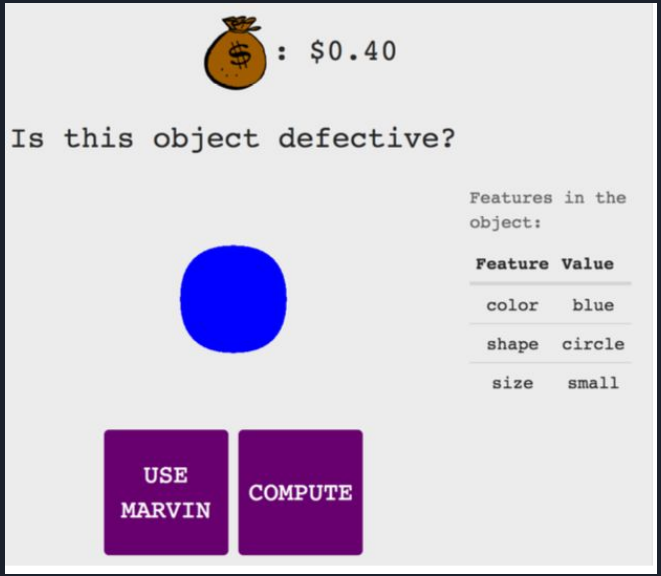
Discussion Question?

- What information about an AI model would you need to know in order to trust the responses for the following situations:
    - Blind date compatibility matches on a dating website
    - Health prognosis
    - Riding in a self driving vehicle
- Think about what separates these sort of situations that make you more prone to trust the AI.

Recommended Considerations for Human-Centered AI

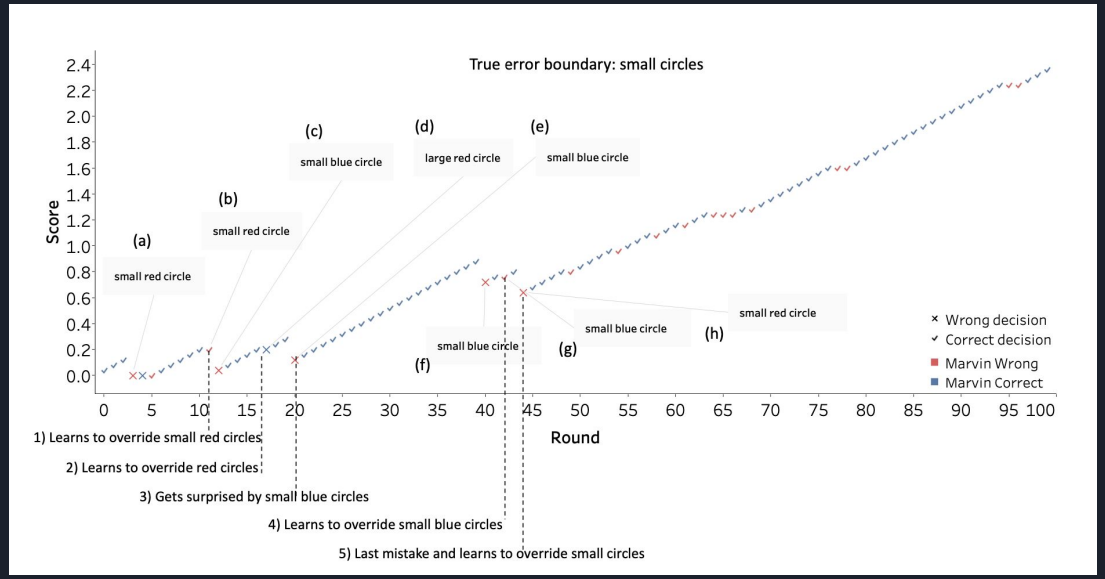
1. Parsimony
   a. The parsimony of an error boundary is inversely related to it representational complexity. (more complex = less parsimonious)
   b. For AI error boundaries formulated in mathematical logic using disjunctive normal form, complexity depends on the number of conjuncts and literals in the function considered. (literals = values & features)
   c. Ex. {(age = old $\wedge$ blood Pressure = high) $\vee$ (age = young $\wedge$ blood Pressure = low)}
2. Stochasticity
   a. An error boundary is non-stochastic if it separates all mistakes from correct predictions.
   b. 3 reasons: generalization, representation mismatch between the AI and human, and inherent stochasticity in the outcome being predicted.
3. Task Dimensionality
   a. eliminate irrelevant features
   b. analyzing trade-off between marginal gain of performance vs marginal loss of the accuracy
4. Backwards Compatibility
   a. regularizing in order to minimize the introduction of new errors on instances where the user has learned to trust the system.

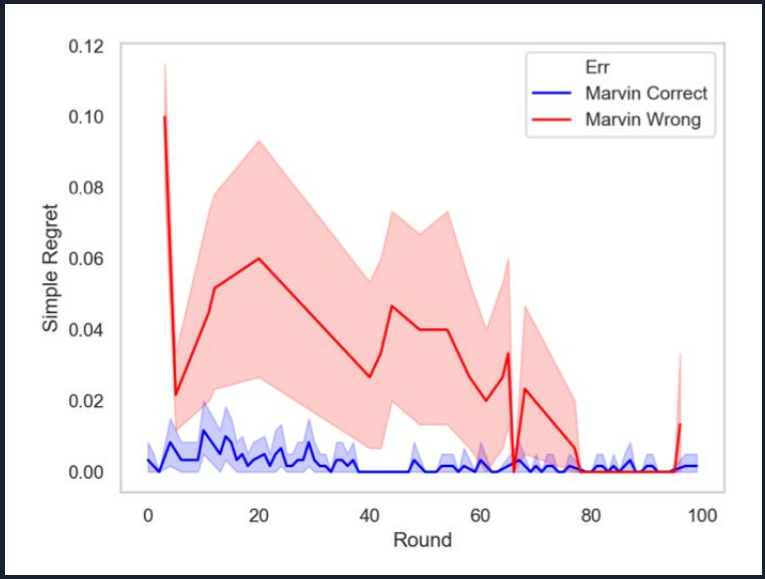
Mental Model Experiment

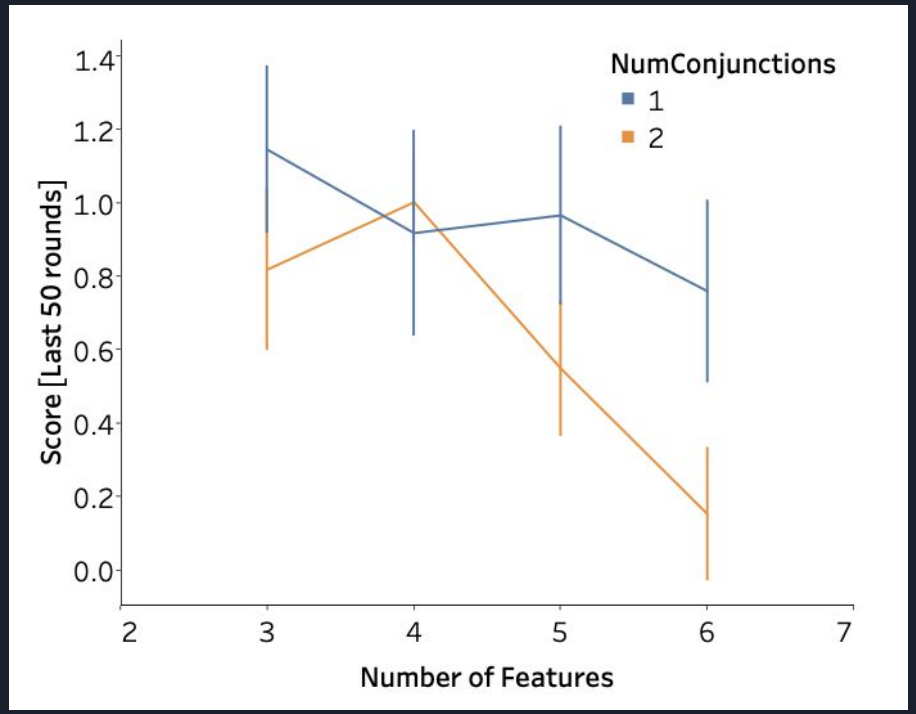


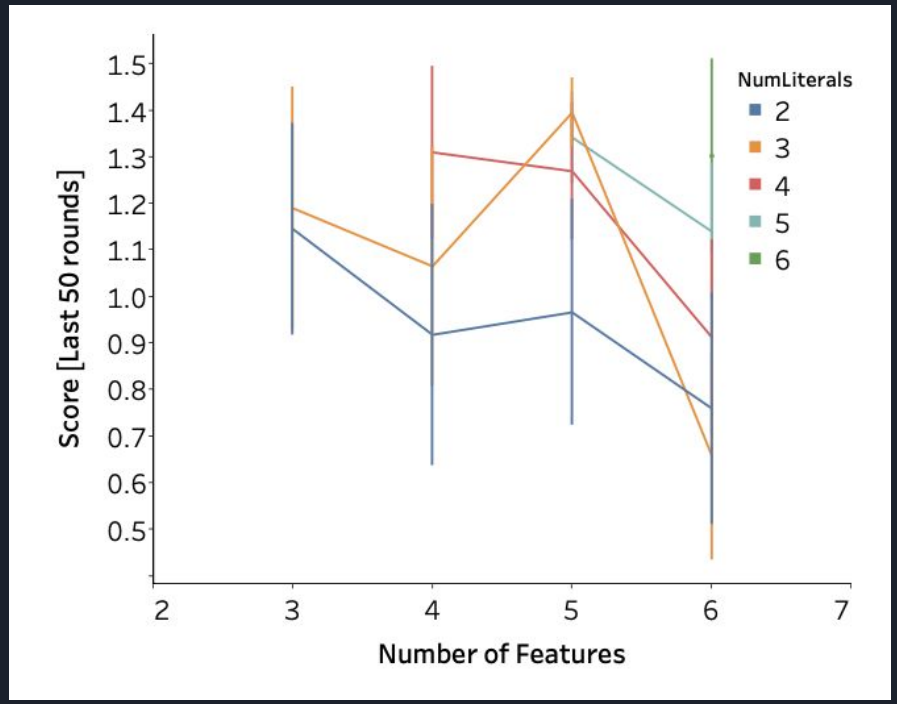
	Marvin Correct	Marvin Wrong
Accept	$0.04	-$0.16
Compute	0	0

Mental Model Results






NumConjunctions
1
2
Score [Last 50 rounds]
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
2
3
4
5
6
7
Number of Features

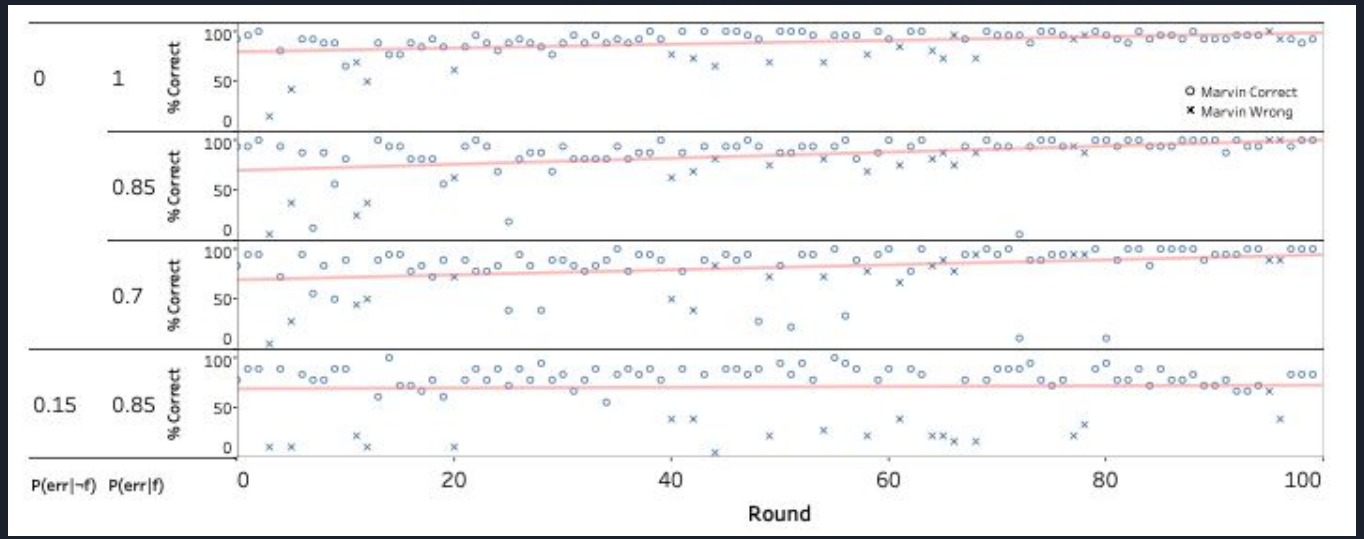

NumLiterals
2
3
4
5
6
Score [Last 50 rounds]
1.5
1.4
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
2
3
4
5
6
7
Number of Features

Questions proposed

Q1: *Do people create mental models of the error boundary? How do mental models evolve with interaction?*

Q2: *Do more parsimonious error boundaries facilitate mental model creation?*

Q3: *Do less stochastic error boundaries lead to better mental models?*



Related Work

- Mental models for collaboration.
- Backward compatibility.
- Interpretability for decision-making.
- Modeling and communicating uncertainty in ML

Optional Reading:
Guidelines for Human-AI Interaction

- AI-infused systems can be disruptive, offensive, confusing, or even dangerous
- Since AI is new, there are not set laws to govern its use

Why we need this

- Small and large error potential
  - Autocomplete issues: Man is to Computer Programmer as Woman is to Homemaker

4 main sections: Breakout Rooms (6 minutes)

- Group 1: Initially & During interaction
- Group 2: When wrong
- Group 3: Over time

Questions

- Would you add or remove any of the guidelines?
- Were any of the guidelines unclear?
- Can you think of products that follow or violate any of the guidelines?

Breakout Room #1: Initially and During Interaction

		AI Design Guidelines	Example Applications of Guidelines
Initially	G1	**Make clear what the system can do.** Help the user understand what the AI system is capable of doing.	[Activity Trackers, Product #1] "Displays all the metrics that it tracks and explains how. Metrics include movement metrics such as steps, distance traveled, length of time exercised, and all-day calorie burn, for a day."
	G2	**Make clear how well the system can do what it can do.** Help the user understand how often the AI system may make mistakes.	[Music Recommenders, Product #1] "A little bit of hedging language: 'we think you'll like'."
During interaction	G3	**Time services based on context.** Time when to act or interrupt based on the user's current task and environment.	[Navigation, Product #1] "In my experience using the app, it seems to provide timely route guidance. Because the map updates regularly with your actual location, the guidance is timely."
	G4	**Show contextually relevant information.** Display information relevant to the user's current task and environment.	[Web Search, Product #2] "Searching a movie title returns show times in near my location for today's date"
	G5	**Match relevant social norms.** Ensure the experience is delivered in a way that users would expect, given their social and cultural context.	[Voice Assistants, Product #1] "[The assistant] uses a semi-formal voice to talk to you - spells out "okay" and asks further questions."
	G6	**Mitigate social biases.** Ensure the AI system's language and behaviors do not reinforce undesirable and unfair stereotypes and biases.	[Autocomplete, Product #2] "The autocomplete feature clearly suggests both genders [him, her] without any bias while suggesting the text to complete."

Breakout Room #2: When Wrong

When wrong	G7	**Support efficient invocation.** Make it easy to invoke or request the AI system's services when needed.	[Voice Assistants, Product #1] "I can say [wake command] to initiate."
	G8	**Support efficient dismissal.** Make it easy to dismiss or ignore undesired AI system services.	[E-commerce, Product #2] "Feature is unobtrusive, below the fold, and easy to scroll past...Easy to ignore."
	G9	**Support efficient correction.** Make it easy to edit, refine, or recover when the AI system is wrong.	[Voice Assistants, Product #2] "Once my request for a reminder was processed I saw the ability to edit my reminder in the UI that was displayed. Small text underneath stated 'Tap to Edit' with a chevron indicating something would happen if I selected this text."
	G10	**Scope services when in doubt.** Engage in disambiguation or gracefully degrade the AI system's services when uncertain about a user's goals.	[Autocomplete, Product #1] "It usually provides 3-4 suggestions instead of directly auto completing it for you"
	G11	**Make clear why the system did what it did.** Enable the user to access an explanation of why the AI system behaved as it did.	[Navigation, Product #2] "The route chosen by the app was made based on the Fastest Route, which is shown in the subtext."

Breakout Room #3: Over Time

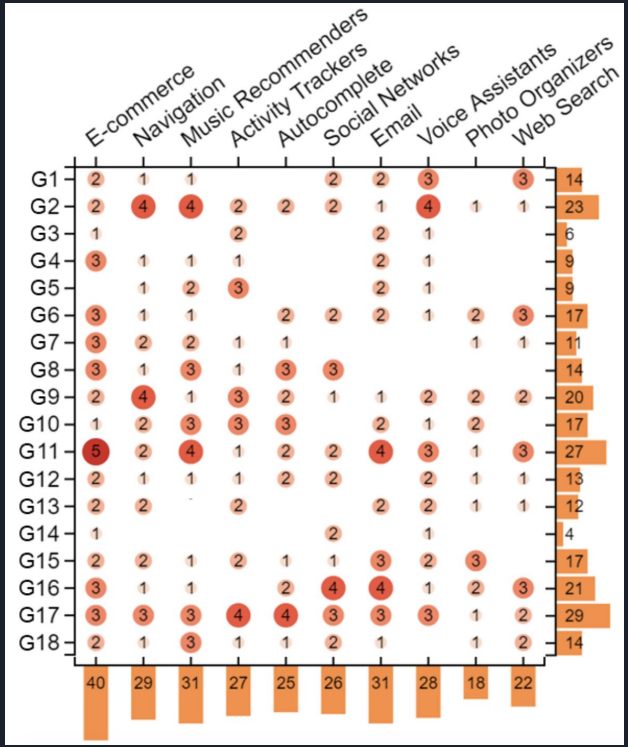
Over time	G12	**Remember recent interactions.** Maintain short term memory and allow the user to make efficient references to that memory.	[Web Search, Product #1] "[The search engine] remembers the context of certain queries, with certain phrasing, so that it can continue the thread of the search (e.g., 'who is he married to' after a search that surfaces Benjamin Bratt)"
	G13	**Learn from user behavior.** Personalize the user's experience by learning from their actions over time.	[Music Recommenders, Product #2] "I think this is applied because every action to add a song to the list triggers new recommendations."
	G14	**Update and adapt cautiously.** Limit disruptive changes when updating and adapting the AI system's behaviors.	[Music Recommenders, Product #2] "Once we select a song they update the immediate song list below but keeps the above one constant."
	G15	**Encourage granular feedback.** Enable the user to provide feedback indicating their preferences during regular interaction with the AI system.	[Email, Product #1] "The user can directly mark something as important, when the AI hadn't marked it as that previously."
	G16	**Convey the consequences of user actions.** Immediately update or convey how user actions will impact future behaviors of the AI system.	[Social Networks, Product #2] "[The product] communicates that hiding an Ad will adjust the relevance of future ads."
	G17	**Provide global controls.** Allow the user to globally customize what the AI system monitors and how it behaves.	[Photo Organizers, Product #1] "[The product] allows users to turn on your location history so the AI can group photos by where you have been."
	G18	**Notify users about changes.** Inform the user when the AI system adds or updates its capabilities.	[Navigation, Product #2] "[The product] does provide small in-app teaching callouts for important new features. New features that require my explicit attention are pop-ups."

Phases of Guideline Creation

- Phases 1 & 2 to create these guidelines
  - 1: Consolidating guidelines.
  - 2: Modified heuristic evaluation

Phase 3: User Story

Product Category	Feature	Participants
E-commerce (Web)	Recommendations	6
Navigation (Mobile)	Route planning	5
Music Recommenders (Mobile)	Recommendations	5
Activity Trackers (Device)	Walking detection and step count	5
Autocomplete (Mobile)	Autocomplete	5
Social Networks (Mobile)	Feed filtering	5
Email (Web)	Importance filtering	5
Voice Assistants (Device)	Creating a reminder with a due date	5
Photo Organizers (Mobile)	Album suggestions	4
Web Search (Web)	Search	4



Phase 4: Expert Evaluation of Revisions

Phase 1: Consolidating guidelines Set appropriate expectations. Set accurate expectations to give people a clear idea of what the experience is and isn't capable of doing.
Phase 2: Internal evaluation Set appropriate expectations.
Phase 3: User study G1: Make capabilities clear. Help the user understand what the AI system is capable of doing. G2: Set expectations of quality. Help the user understand what level of performance the AI system is capable of delivering.
Phase 4: Expert evaluation of revisions G1: Make clear what the system can do. Help the user understand what the AI system is capable of doing. G2: Make clear how well the system can do what it can do. Help the user understand how often the AI system may make mistakes.

Conclusion

- A major theme of these guidelines focus on transparency, which as we have seen is not always applicable
- Focused on user experience