13: AI Application Design Process

Juho Kim

Human-Al Interaction KAIST Fall 2023 | hai.kixlab.org

Administrative Notes

- Assignment #3: "Evaluating Long-Tail AI Tasks"
 - Shorter than #1 and #2
 - To be released later this week
 - There will be another short, final assignment on multimodal interaction.
- Milestone #2: Prototype

• Due: Nov 22 (Wed)

Previously on CS492G...

DATA

The Norwich line steamboat train, from New-Loodon for Boston, this Intraining ran of the track even miles north of New-Joon morning morning Verlooks

Cobi	search	smitch by session / paper / autor traine														na -						
	ESTA	Schrödung weaked Managing Social Madia Number: charge in # of conflicts																				
* Conflicts @	Unsch	Unscheduled Sessions 🚯 Unscheduled Papers 🚯																panicipants last month				
High severity (20) authors with papers in apposing sessions (7) papers of mutual interests in opposing sessions (20)		unused assiston 9	cross- tear work	Performing on Ditage	Musical performan	During Seconds	Invited Talk	unused assessor 2	*		122	Many se of sbook										
Medium severity (46)	Room/ Time	But	Bordeau	a 2525	352AB	Hacane	241	HDA	21	351	242A	2428	243	253	343	252A	301	362/363	221/221M	· ~ · (1000		and the second s
session (42) topics of interest to a persona in opposing sessions (4)	Mon 11:0 12:29	 Managing Social Media 	Lifetime Research Autor2	Cull Al Game Changers	Multitouch and Gestu	at Uter Interfaces	Reflection and Evaluation	Interaction in the Wild	Learning	Crowdsour Pergie Power	Erhancing Access		Uber Interface Design and	Six Steps to Successful UIX in an	Rapid Design Lat - A Toxi h	Body, Whyt 5 Videotape	Designing Interactive Secure	Humani Computer Interaction	Birds of a heather - session 1		CLEAR COLO CO	
Preferences																				You've got email: Discover what	Amazon, Apple, Facebook, Google:	COVID-19 dilemmas around the
popers that are good in the same	Mon 1410 15:20	C- INH Massive Online	and fransieto	Gare	Displays	Community Community	Co-Design with Users	Brain Sensing and Analysis	Evaluation Mathods 1	Reyboards and Hotevs	Technologies for Life 1		Practical Statistica fo Uber	Agle User Experience and UCD	Papid Design Lal	Speech- to based	+	The Plote of Engineering Work in CP	t Bixts of a 5 feather - 5 session 2	Tall us your calcian about some mal-world	Can you tell the difference?	Test how your percentes to maral diamona
View Options		+5 *	42 **	*2 **	2 +4	+2 1.2	44	+1 #-2	+1 #+1	+1 =-4	+8 +-7 +-1		+2 ***	+2 1+2	+2 **	+2 1.2	+2 *-1	+6 **	+2 **2	email designs and you will learn what looks will tend to make you click! This test takes	This study takes 5 minutes.	compare to others. This study takes around 5 minutes.
Conflicts Poderscore Session Type	Mon 16:0 17:35	 Smart Yools, Smart We +4 	Tables av Roors sk 4 +7	Progress of +5	theraction +6 1+1	+2 1-2	d Innovating User- Centered +5	Design for Class srooms 1 +9	Exploring Games +10 ^{#+1}	Large and Public Displays +6 +7	Creating and Authoring +6 #+6		Practical Statistics In Uber +5 1+5	Agle User Experience and UCD +5 1+5	Repid Design Lat -A Tool Is +5	Speech based interaction +5 \$+5	+	Enhancing the Research +5 1-5	Birds of a feather - session 3 +5	around 7 minutes.		
Duration			2 84	<u>.</u>	8-6	#-t		8-4	8-3	8-1	8-1		-		-		-			Participate now!	Participate now!	Participate now!
Awards Honorable Mentions	10.20	around Devices	Classroom 2	- an Activ	Video az	for Life 2	n Expenence	in Mones	Creativity	Studies	the Hume		Experience Evaluation	Decision Making for	Crash Dummes	Rocus Media Data	Engagement in Health	A new Perspective for the	mather - session 4			
Session Types		+2 :	+4		+1	0		-1	+1 ***	+7	+2 ***					-1 **	+2 ***	+2				8-9 1000
> Personas	Tue 11.00 12:20	Erergy	Pull-Body Interaction	UX Managem Current an	Serising Touch e	Methods	Colonard of Practice	Video Communicat	Exerganes and Bayon	Pointing and Fitts Lave	d Impairment and Rehybritatio		User Experience Evaluation	Choice and Declaran Making for	Capitolive Crash Durwinus	Acatumy Social Media Data	Research- Practice Interaction	Digital Art Challenging Perspective	Birds of a feather - er session 5			
> Communities	Tax 16:00	-1 E	-2	+1	2 +2 1-3		-2	-1 g.t.	+3 8+3	+4 #+4	-1		-1 3-1 Partice	1.11	-1 -1	-1 ad	0 #-1	+1 +1	-1 and	Test your spatial reasoning!	Do you make assumptions about	How well can Al understand your
* History	15.20	Eceryate	re the Horse	Doctor Listering 1	Esperance	es impact Award	How We Work	Programming	Design	Design	Settings		Statistics N User	For Experts	in the second	n für Nevigation	(Non-Voual Interaction)	Perspectiv	et feather - menator 6	root your optical foldoring:	people without knowing it?	speech?
You unscheduled paper. The Many		0	2 +2 **	2	+2 ***	+1 #+1	+2 =+1	+2 1-2	+12	+1 ***	-1 1.1		+1 #+1	+1 #+1	+1 .	+1 #*1	+1 8+1		+1 #+1	Test how well you can visualize objects and images in space! The test typically takes 15 minutes to complete.	How well do you find patterns? Do you make assumptions about people without knowing it?	Find out how well AI can understand your
Faces of Facebook: Ex., from Managing Social Media	Tue 16:00 17:00	Public Displays	Ethics in HC2	Gamilicati @ Work	an Collaborati Orealizer	Ni Design Research	Studying Digital Antifacto	Embodied Interaction 2	Reading and Writing	E-2 Developing the World	8-3 Communicat Health		Practical Statistics In	Expert Reviews -	Make This stroducte	Card Sorting	HCI with Sports	Managing UK Teams	Birds of a heather - session 7			speech by reading out loud a few words and sentences! The test takes around 7 minutes.
		+1 .	a +1 **	1	+3 **	+2 1-2	+3 **	+1 1-1	+3 8.4	+1 #+1	+4 =+3		-		-	- Carlos		+1 #+1				

INTERACTION

MODEL

Today's Learning Objectives

After today's class, you should be able to...

- Understand distinct stages of desinging an Al-powered application.
- Identify major design / engineering considerations in each of the design stages.

Reflection on the last in-class activity

Future of Crowd Work

Crowdsourcing is useful in all stages of Al design.

DATA	MODEL	INTERACTION
• Data generation	 Evaluating and debugging models 	 Hybrid intelligence systems Behavioral studies

Future of Crowd Work

"Can we foresee a future crowd workplace in which we would want our children to participate?"

- Worker considerations
 - Motivation, feedback, pay
- Requester considerations
 - Coordination, task decomposition, quality control

Kittur, Aniket, et al. "The future of crowd work." *Proceedings of the 2013 conference on Computer supported cooperative work.* 2013.



"These people doing 'ghost work' make the internet seem smart."

- Generating training data
- Flagging bad content
- Manually fixing Al errors

~8% of Americans have contributed to "ghost economy".

Future of Crowd Work: Design Goals

- Create career ladders
 - motivation, job design, reputation, hierarchy.
- Improve task design through better communication
 - QA, job design, task assignment, realtime crowd work, synchronous collaboration, platform.
- Facilitate learning
 - quality assurance, Als guiding crowds (and vice versa), task assignment, reputation and credentials, platform.

Al Design Process

What's unique in designing Al-powered systems?

• Data matters

 Big data preferred, Cost & time associated in collecting and handling data, Bias and fairness issues

Model matters

 Probabilistic, Hard to predict, Possibly a black box, Transparency, Interpretability, Resources

Interaction matters

 Mental model, Explanation & feedback, Intelligent UI, Helping users in the task, Human-AI collaboration

Our Three-Stage Framework

DATA

MODEL

INTERACTION

- Collecting, cleaning, processing, labeling, verifying data
- Crowdsouring, Logs, User input, ...

- Train + Test + Parameter tuning
- Apply to a new problem for prediction.
- Picking the right model is challenging.

- Model results (and explanations) are shown to the user through UI.
- User feedback can be fed back to the system.

Steps in applying ML to practice

7 Steps of Machine Learning

- Gathering Data
- Preparing that Data
- Choosing a Model
- Training
- Evaluation
- Hyperparameter Tuning
- Prediction



https://www.youtube.com/watch?v=nKW8Ndu7Mjw&vl=en

Al Design Process

- Gather data
- Prepare the data
- Choose a model
- Train the model
- Evaluate the model
- Tune hyperparameters
- Predict

DATA

MODEL

(Human-Centered) Al Design Process

- Set a goal & identify user needs.
- Gather data
- Prepare the data
- Choose a model
- Train the model
- Evaluate the model
- Tune hyperparameters
- Predict
- Design human-Al interaction
- Apply to UI
- Evaluate & Improve

PLANNING

DATA

MODEL

INTERACTION

Step 1. Set a goal & identify user needs.

- User-centered design process is helpful here.
- Without this stage, it's likely to yield technology that kind of works but nobody uses.
- Netflix: "The user either finds something of interest [within the first 60 or 90 seconds] or the risk of the user abandoning our service increases substantially,"



Gomez-Uribe, Carlos A., and Neil Hunt. "The netflix recommender system: Algorithms, business value, and innovation." ACM Transactions on Management Information Systems (TMIS) 6.4 (2015): 1-19. http://www.spring2innovation.com/2019/04/design-thinking-vs-user-centred-design/

Step 2. Gather data

- You need to know the required quantity and quality levels of the data to be collected.
- Features of the data need to be identified.
- User privacy / logging overhead / cost of collection
- Netflix
 - your interactions with our service (such as your viewing history and how you rated other titles)
 - other members with similar tastes and preferences on our service
 - information about the titles, such as their genre, categories, actors, release year, etc.
 - \circ the time of day you watch
 - \circ \quad the devices you are watching Netflix on
 - how long you watch

Step 3. Prepare the data

- Processing, Labeling, Filtering
- Exploratory Data Analysis: get a sense of your data
- Segment into training/test data
- Crowdsourcing / dedicated data workers
- Netflix: 1,300 clip groups based on preference, 2,000 preference groups
 - "Netflix has hired real life humans to categorize every bit of TV shows and movies and apply tags to each of them in order to create hyperspecific micro genres such as "Visually-striking nostalgic dramas" or "Understated romantic road trip movies".

https://uxplanet.org/netflix-binging-on-the-algorithm-a3a74a6c1f59

Step 4. Choose a model

- There are tons of models available.
- Common to test multiple models and compare their performance.
- Data matters: type, size, features... of the data affect which model suits better.

Step 5. Train the model

- Iteratively improving the model performance through trial and error.
- Make predictions with a set of weights and biases, compare against the ground truth, and make improvements.



https://www.youtube.com/watch?v=nKW8Ndu7Mjw&vl=en

Step 6. Evaluate the model

- Apply the trained model to unseen data (test data) and see how well it performs.
- Overfitting: model too optimized for the training data
- Occam's razor: "entities should not be multiplied unnecessarily which is interpreted as requiring that the simplest of competing theories be preferred to the more complex" [Merriam-Webster]
- Model: abstract, simple, general solution to explain a complex phenomenon.



Step 7. Tune hyperparameters

- Make further performance improvements by changing various factors that comprise the model.
 - e.g., number of iterations, learning rate
- Experimental process: Often relies on experience and best practices

Step 8. Predict

- Now apply the model to a real target scenario so that it is used to make a (hopefully accurate) prediction.
 - e.g., Netflix recommends clips a user is likely to watch.
- This is where AI truly shines (if done well): relying on model prediction vs manual judgment or hard-coded rules
 ⇒ more efficient and scalable
- Practical issues
 - Content: does the model fare well against new content?
 - Users: does the model fare well against new users?

Step 9. Design human-Al interaction

- Now come back to the user and think how the model can be used to truly address the user's needs.
- UX of AI matters.
 - User motivation
 - Mental model
 - Risk
 - Trust
 - FAccT
 - Interpretability
 - 0 ...
 - "How much Al is okay?"

Step 10. Apply to UI

• Combine the human-Al interaction model and the Al model, and integrate them into a UI.



https://netflixtechblog.com/selecting-the-best-artwork-for-videos-through-a-b-testing-f6155c4595f6 https://uxplanet.org/netflix-binging-on-the-algorithm-a3a74a6c1f59

Step 10. Apply to UI

- Thinking about how AI can enhance UX is important.
- Thinking how AI can damage UX is (even more) important.



https://netflixtechblog.com/selecting-the-best-artwork-for-videos-through-a-b-testing-f6155c4595f6 https://uxplanet.org/netflix-binging-on-the-algorithm-a3a74a6c1f59

Step 11. Evaluate and iterate

- Use various metrics to measure the success through a live deployment: beyond just the model performance.
- Often combined with A/B testing for more informed decision making.
- Metrics in practice: click-through rate, time to completion, diversity of options presented, explainability, ...
- Establish a process for iterative improvement.

Step 11. Evaluate and iterate

Quantifying Netflix Benefits

A good choice leads to a complete viewing A poor choice leads to abandonment, and risk of cancel

10% "better" choices \rightarrow +500M/month good outcomes If 1% of those avoids a cancellation \rightarrow \$500M/year

Our measurement thresholds:

0.1% retention improvement (\$5..50M/year)

0.1% more viewing per time period

(Human-Centered) AI Design Process: THOUGHTS?

- Set a goal & identify user needs.
- Gather data
- Prepare the data
- Choose a model
- Train the model
- Evaluate the model
- Tune hyperparameters
- Predict
- Design human-Al interaction
- Apply to UI
- Evaluate & Improve

PLANNING

DATA

MODEL

INTERACTION

What parts of it has changed with genAl/LLM?

- Why do you think these changes occurred?
- What do these changes mean?

- Set a goal & identify user needs.
- Gather data
- Prepare the data
- Choose a model
- Train the model
- Evaluate the model
- Tune hyperparameters
- Predict
- Design human-Al interaction
- Apply to UI
- Evaluate & Improve

. PLAN-NING **INTER-ACTION** DATA MODEL

- UX methods, needfinding
 - Model performance to interaction performance
 - Mental model, Explanation & feedback, Intelligent UI, Human-AI collaboration
 - Data pipeline
 - Collecting, labeling, processing cost
 - Bias, fairness
 - Model performance
 - Computational cost
 - Interpretability, transparency

MODEL-CENTRIC DESIGN PROCESS

Here's a cool model. Let's build a better performing model.

Here's a cool model. What can we build with it?



ITERATIVE USER-CENTERED DESIGN PROCESS



Image Credit: Stanford d.school

ACTIVITY: KAIST course recommendation system

- As designers of such a system, what are some guiding questions in each of the stages in the design process we should ask ourselves?
- Teams of 3-4, 10 mins



Issues in the AI design process

- There exists a large technical debt in ML systems.
- Data handling consumes too much time / cost.
- Communicating about data / model is difficult.
- Not much attention is given to human-Al interaction.



Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

Sculley, David, et al. "Hidden technical debt in machine learning systems." Advances in neural information processing systems. 2015.

Data work consumes most time.



What data scientists spend the most time doing

- Building training sets: 3%
- Cleaning and organizing data: 60%
- Collecting data sets; 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%

https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/#713197216f63

Model Cards

- Short document about a trained ML model, with its intended use and performance characteristics.
 - Goal: To help users decide whether and how to apply the model to their context.
- A structured communication medium (like a spec sheet for hardware devices and electrical components) to be shared across different stakeholders.
- Also, data sheets for data

Gebru, Timnit, et al. "Datasheets for datasets." arXiv preprint arXiv:1803.09010 (2018). Mitchell, Margaret, et al. "Model cards for model reporting." Proceedings of the conference on fairness, accountability, and transparency. 2019. https://modelcards.withgoogle.com/about

ACTIVITY: Let's crowdsource model card generation.

- Let's fill in the missing information to complete a model card for an image cropping Al.
- Groups of 3-4, 20 mins

www.yellkey.com/special