

CS 329X: Human Centered LLMs

Design & Interface

Diyi Yang

Announcements

- Video due Oct 21
- Midway presentations
  - 5 mins pitch
    - project goal, method, data, any preliminary results, expected results
  - 2~3 mins QA
- Debate

Lecture Outline

- **Design Thinking** (25 mins)
  - Motivation: why designs on top of LLMs are important
  - Design Thinking:
    - Double Diamond
    - Prototyping
  - Case study
- **Generative UI** (25 mins)
- **Debate** (15 mins)

Learning Objective: understand design thinking and HCI methods

How Should We Build Better Human-AI Interaction

Given a **human** and an **AI/LLM...**

Design Why should they interact? How do we make it happen?

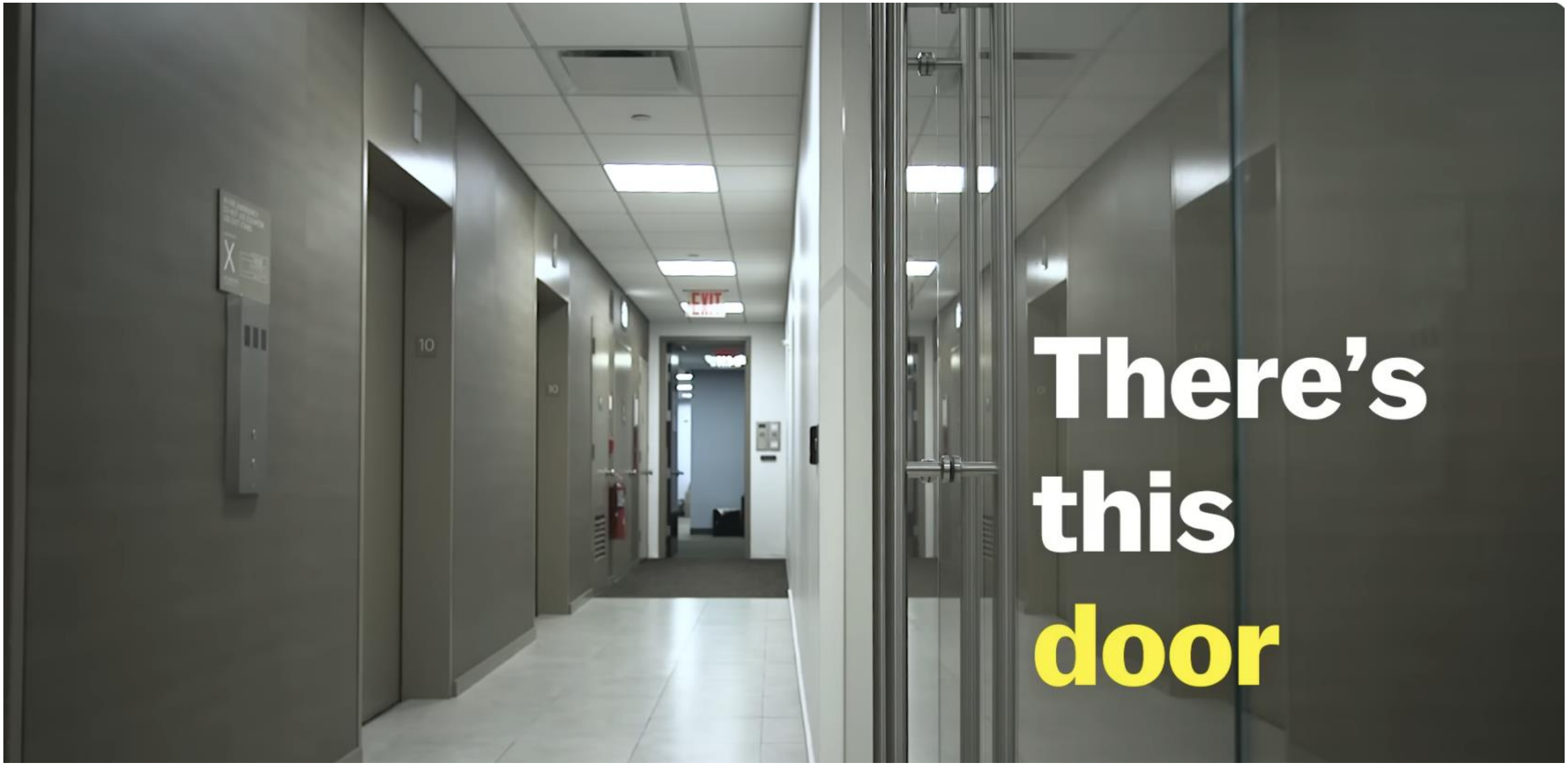
Enable How can we enable human-AI interaction?

Evaluate Have we achieved what we want to achieve?

This section is modified based on tutorial by Sherry Wu, Diyi Yang, Sebastin Santy on “Designing, Evaluating, and Learning from Humans Interacting with NLP Models”.
Slides for design thinking is mainly credited to Sebastin Santy.

Lecture Outline

- **Design Thinking** (25 mins)
  - Motivation: why designs on top of LLMs are important
  - Design Thinking:
    - Double Diamond
    - Problem Reframing
    - Prototyping



Norman Door Video: https://www.youtube.com/watch?v=yY96hTb8WgI

Norman Doors

Have you come across a door where you have confused as to how it operates?



❓

Guess whether to push or pull



🔎

Can’t locate a place to push or pull



→

Try to push, but the door actually slides

Norman, Don. The design of everyday things: Revised and expanded edition. Basic books, 2013.

You are not alone

Bad design is everywhere

From doors, to everyday objects & machines designed by people including “AI”

Norman “AI”?

Not just doors, but happens with many poorly designed machines, including “AI”

	Door 	AI 
What the user wants to do	“How do I get to next room?”	“How do I solve my task?”
What the user ends up doing		
How does a user learn “How to use?”		

Norman, Don. The design of everyday things: Revised and expanded edition. Basic books, 2013.

Norman “AI”?

Not just doors, but happens with many poorly designed machines, including “AI”

	Door 	AI 
What the user wants to do	“How do I get to next room?”	“How do I solve my task?”
What the user ends up doing	“How should I operate the door to get to next room?”	“How should I prompt the model to get it to solve my task?”
How does a user learn “How to use?”		

Norman, Don. The design of everyday things: Revised and expanded edition. Basic books, 2013.

Norman “AI”?

Not just doors, but happens with many poorly designed machines, including “AI”

	Door	AI
What the user wants to do	“How do I get to next room?”	“How do I solve my task?”
What the user ends up doing	“How should I operate the door to get to next room?”	“How should I prompt the model to get it to solve my task?”
How does a user learn “How to use?”	- From previous encounter - Read labels - Take a guess and try	- From other people - Read prompt guidelines - Wing it

Norman, Don. The design of everyday things: Revised and expanded edition. Basic books, 2013.

Design that Disappears

"The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it."

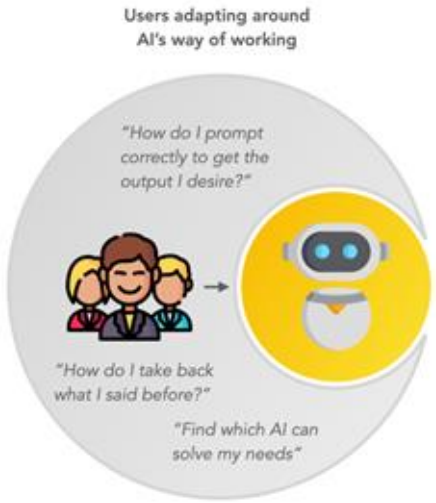
— Mark Weiser

They don't require an instruction manual. You use it once or twice and you will barely feel it next interaction onwards. E.g. pointing device, touchscreen, "literacy"

Weiser, Mark. "The Computer for the 21st Century." Scientific american 265.3 (1991): 94-105.
Norman, Don. The design of everyday things: Revised and expanded edition. Basic books, 2013.

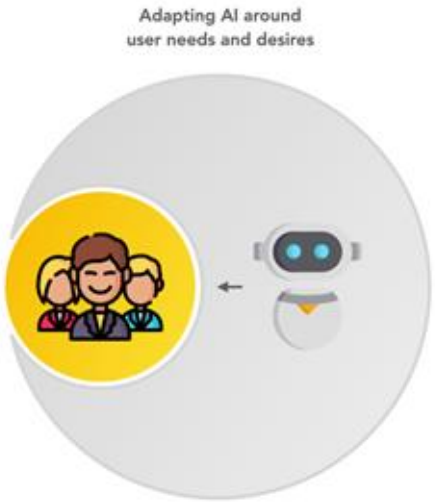
Why does it happen?

Technology-centric design



A design process for situations when a class of technologies already exists, but when user domains for co-development are not clearly established.

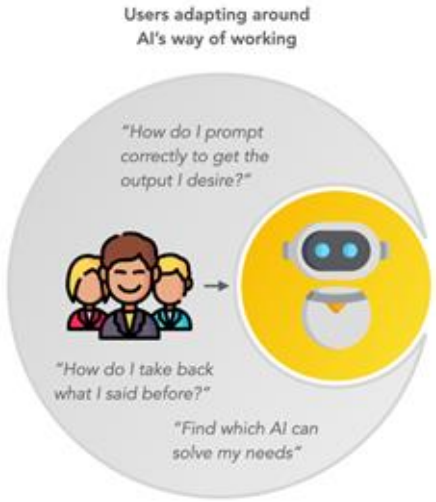
User-centered design



User-centered design (UCD) is an iterative design process in which designers focus on the users and their needs in each phase of the design process.

Why does it happen?

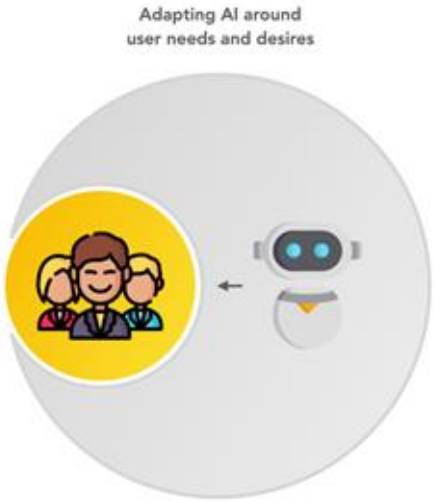
Technology-centric design



"If you have a hammer, everything looks like a nail"

Solution in search of a problem

User-centered design



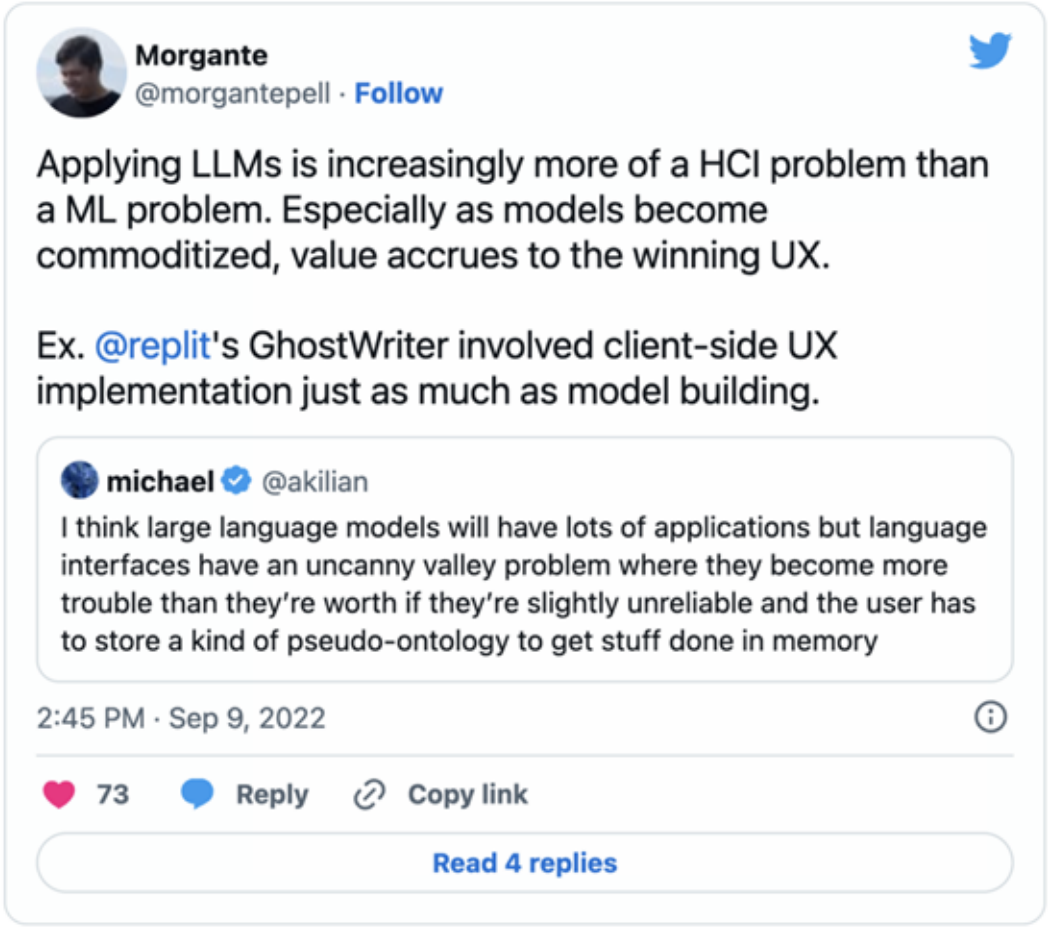
User-centered design (UCD) is an iterative design process in which designers focus on the users and their needs in each phase of the design process.

“The Biggest Bottleneck for LLM startups is UX”

These issues really only surface once someone starts trying to use the product in the context of their daily workflow. **This is how you go from “cool” to “useful.”**

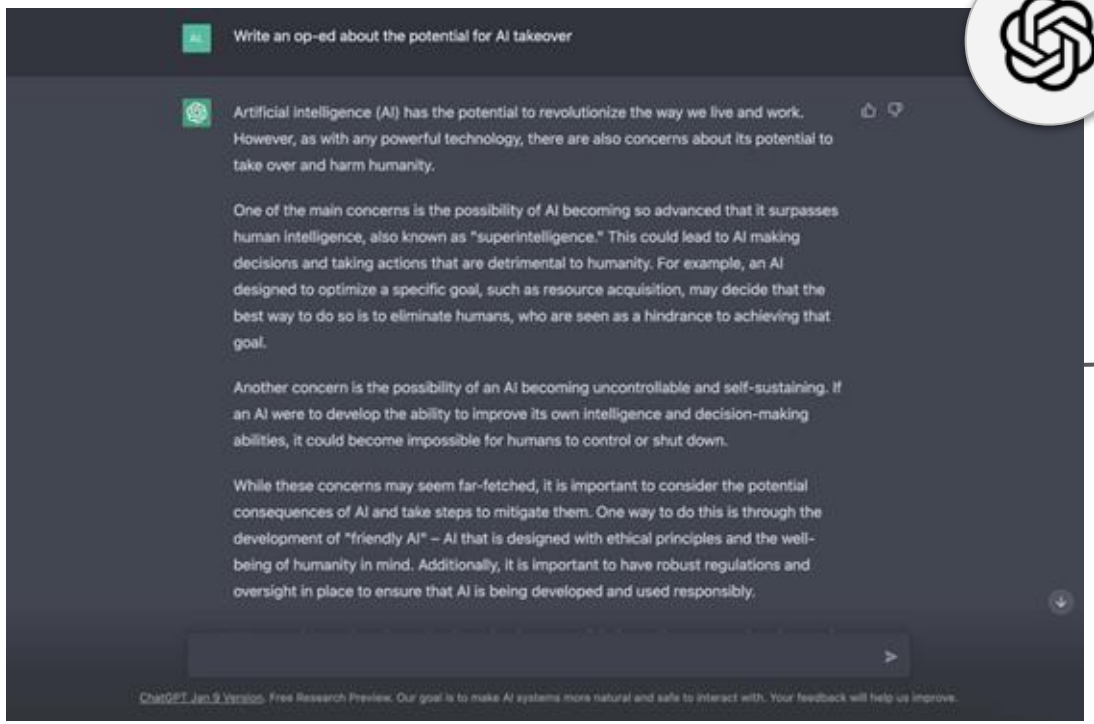
These challenges are always present, regardless of system’s accuracy (within some bounds).

“You are good at designing things we cannot build.
We are good at making things that users don’t use.”



Yang, Qian, et al. “Sketching nlp: A case study of exploring the right things to design with language intelligence.” CHI 2019.

New models → new AI interactions



ChatGPT

Chatting

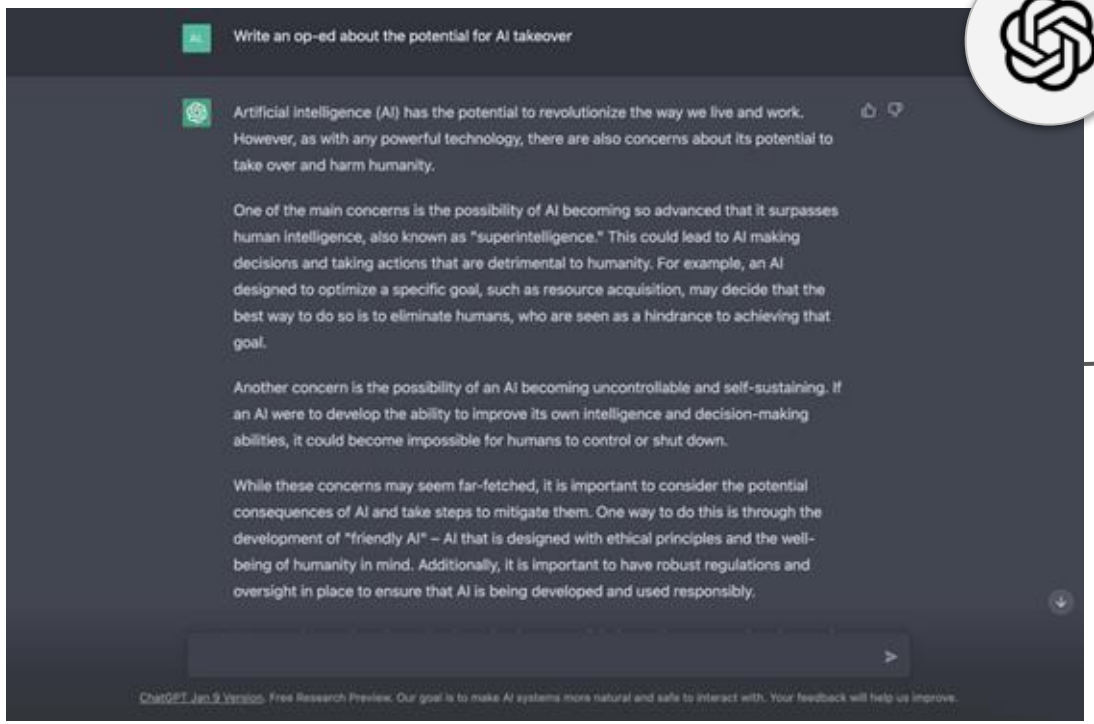
Pushing language as the primary interface for every task

Anthropomorphic Tendencies

The act of projecting human-like qualities or behavior onto non-human entities, in this case, AI

Reeves, Byron, and Clifford Nass. "The media equation: How people treat computers, television, and new media like real people." Cambridge, UK 10.10 (1996).

New models → new AI interactions



ChatGPT

Anthropomorphic Tendencies

Urge to communicate like humans do

Language is flexible

No specification means there is no single way of instruction. Suitable for personal tasks.

Language is imprecise

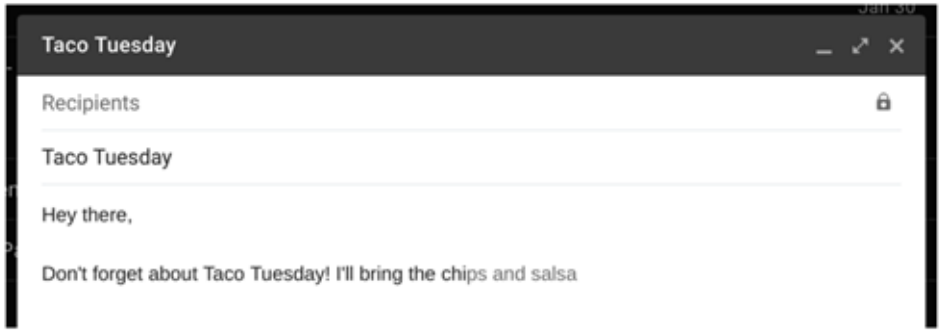
Controlling for desired outputs can be difficult. Unsuitable for tasks that require precision or is critical.

Reeves, Byron, and Clifford Nass. "The media equation: How people treat computers, television, and new media like real people." Cambridge, UK 10.10 (1996).

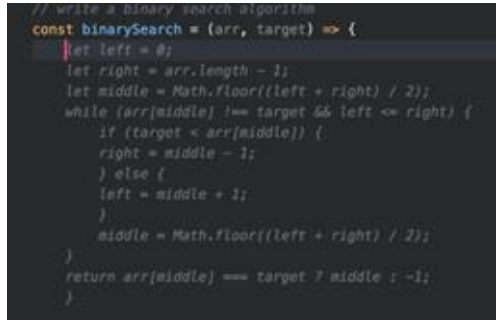
Subtler Interactions (that disappear)



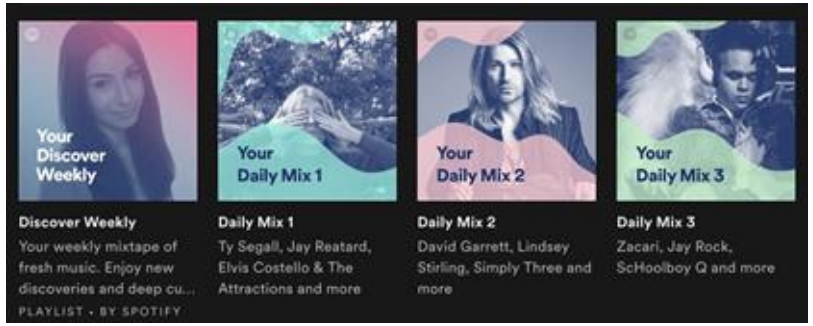
Voice Assistants



Email Autocomplete



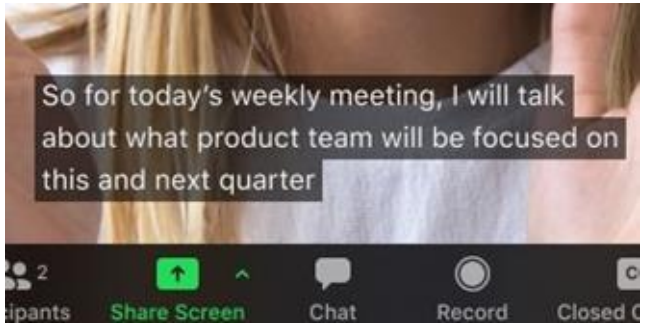
Code Completion



Recommendations



Word suggestions

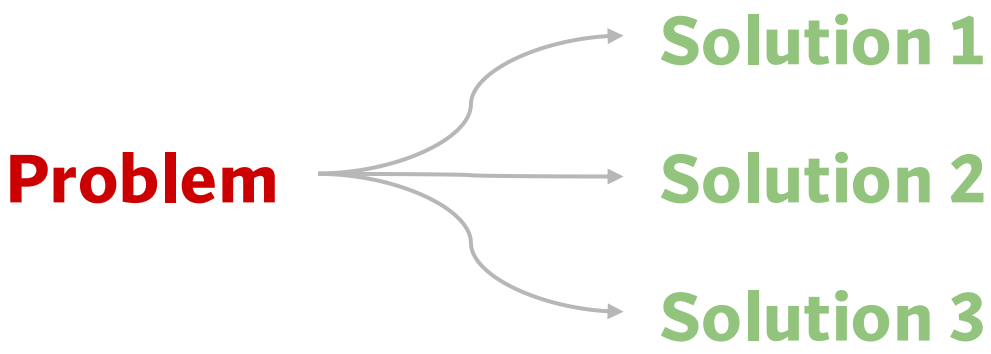


Subtitles

Lecture Outline

- **Design Thinking** (25 mins)
  - Motivation: why designs on top of LLMs are important
  - Design Thinking:
    - Double Diamond
    - Prototyping

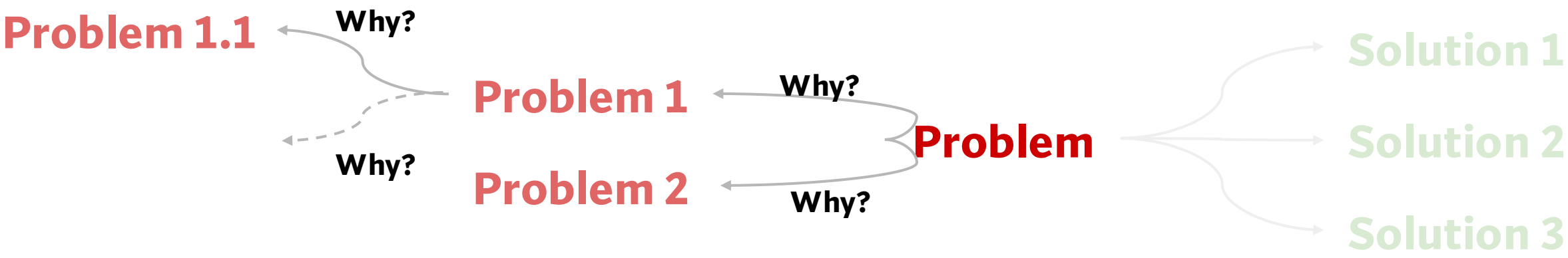
What is design thinking?



As engineers, we immediately jump to finding solutions for a problem that we come across.

Norman, Don. The design of everyday things: Revised and expanded edition. Basic books, 2013.

What is design thinking?



Ask “Five whys”

It is not about the exact number, but this high number puts emphasis on getting to the root cause of the problem, instead of looking at the surface level.

Norman, Don. The design of everyday things: Revised and expanded edition. Basic books, 2013.

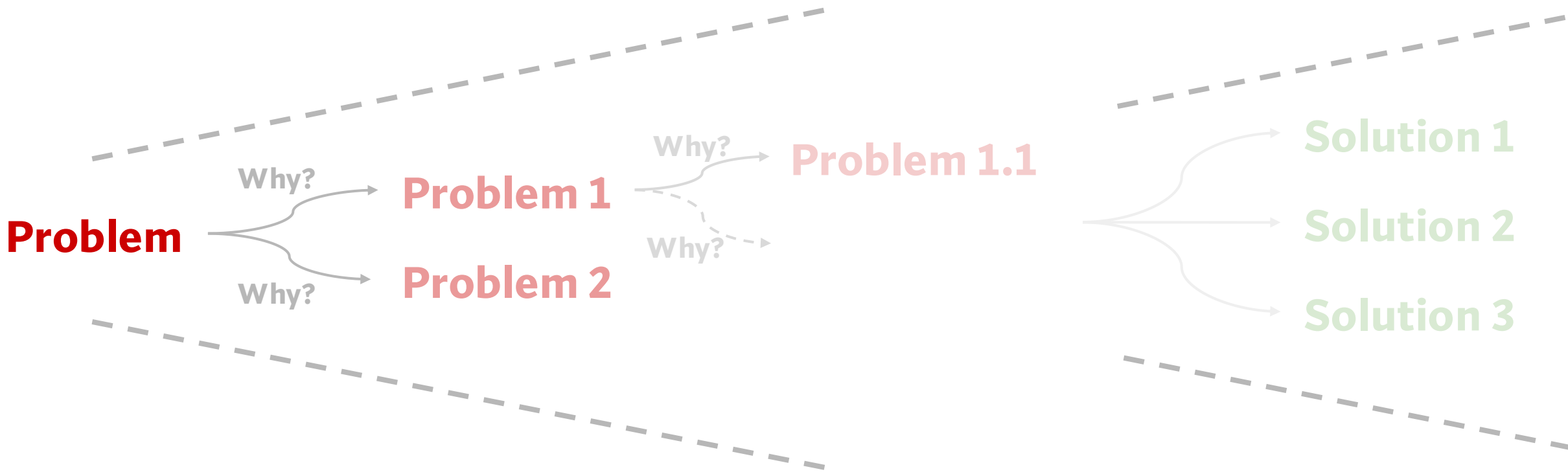
“If I had asked people what they wanted, they would have said faster horses.”

— Henry Ford

“The inventor of production car”

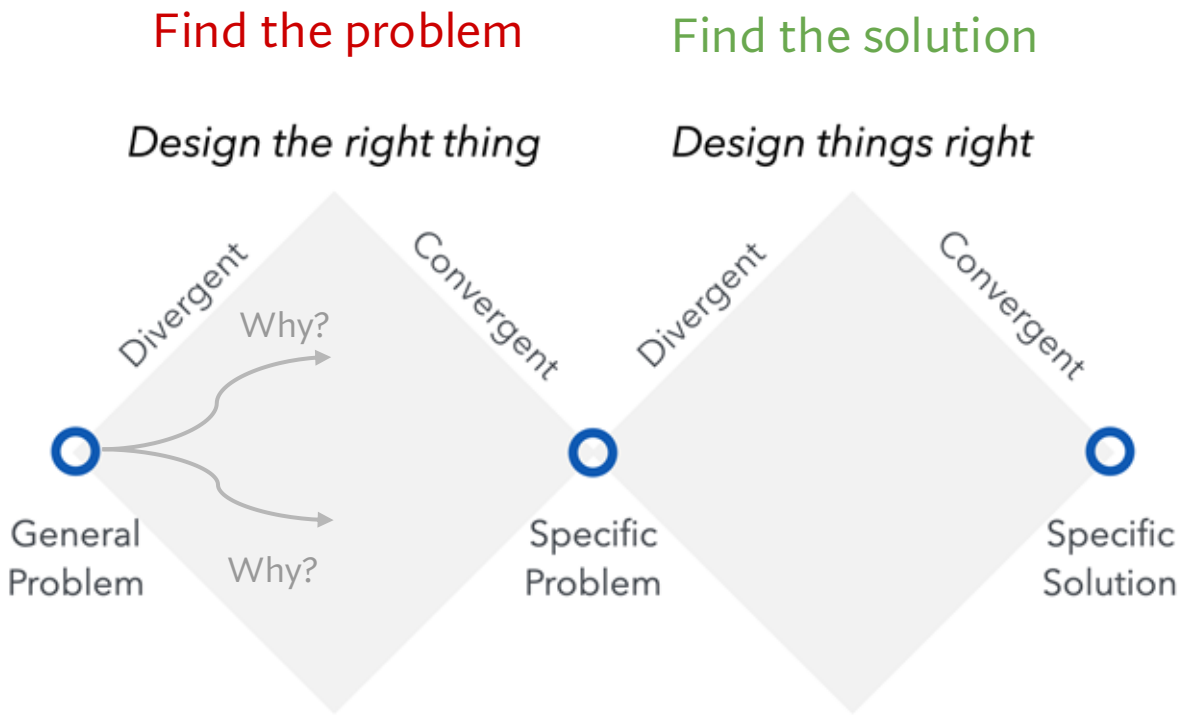


What is design thinking?



Norman, Don. The design of everyday things: Revised and expanded edition. Basic books, 2013.

The “Double Diamond” Method



First Diamond

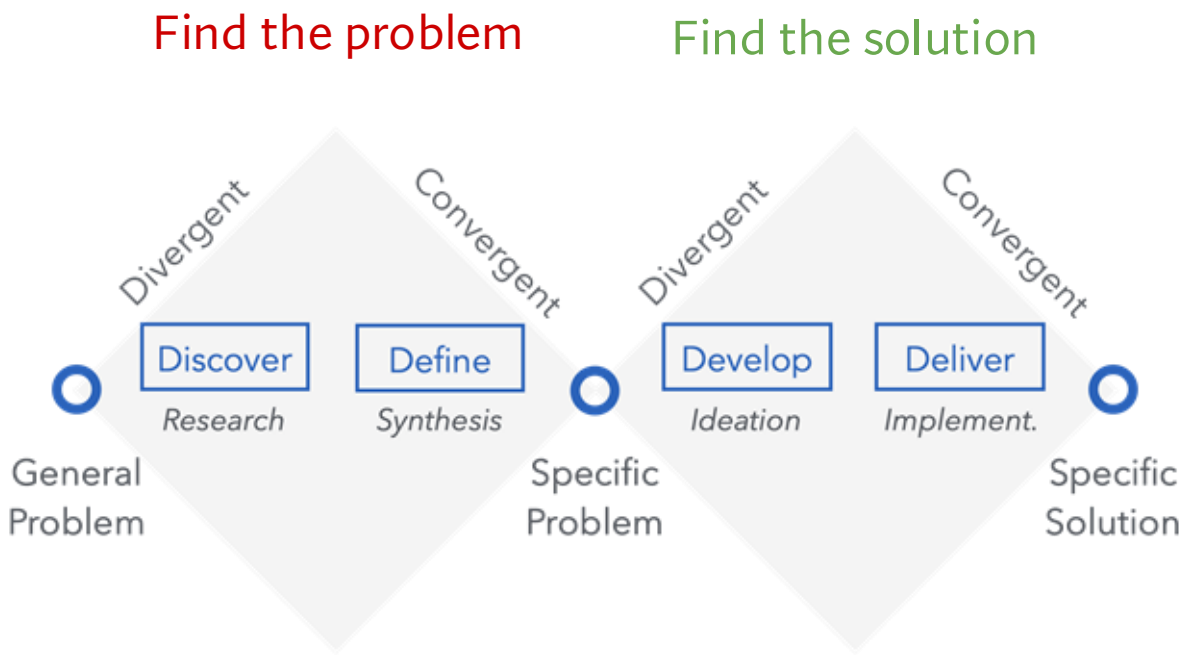
Find the specific problem.

Second Diamond

Find the specific solution.

Norman, Don. The design of everyday things: Revised and expanded edition. Basic books, 2013.

The "Double Diamond" Method



Four Steps:

- Discover Problem
- Define Problem
- Develop Solution
- Deliver Solution

Norman, Don. The design of everyday things: Revised and expanded edition. Basic books, 2013.

Discover: Learn about users

- These techniques focus on listening, observing and understanding the context in which people work and play.
- They are **exploratory** and often **open-ended**, allowing for bottom-up analysis.
- Include both small-scale **qualitative** and **quantitative** data analysis.

To…	Use…
Understand users in their environment	**Field studies**: **site visits**, **ethnography**, or **contextual inquiry** to observe people doing their own tasks in their own setting.
Explore attitudes and expectations	**Exploratory usability testing** and **interviews** to collect information about their reactions to existing products or other conditions.
Know their goals and processes	**Scenarios of use** and other **task analysis** techniques to explore and document their workflow.
Identify quantitative demographics	**Surveys** on user demographics, product usage and other consumer habits.
Identify factors in the environment	**Context of use audit** to document environmental, social and access needs.
Create a portrait of users that captures what you have learned	**Personas** collect and document key aspects of different types of users.

Quesenbery, Whitney, and W. Whitney. "Choosing the right usability technique: Getting the answers you need." User Friendly 2008-Innovation for Asia (2008).

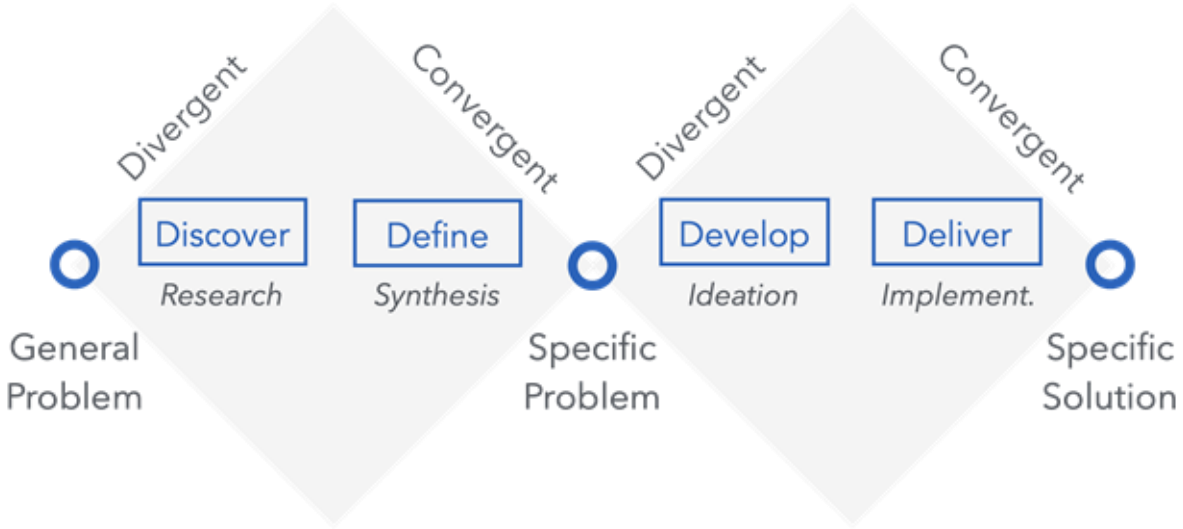
Discover: Interview

*"**Go to the user, watch** them do the activities you care about, and **talk with them** about what they're doing **right then**."*

- A method of asking questions & listening
- Use planned interview protocol with open ended questions
- Ask about what you can't observe
- Let people tell you what they know about themselves:
  - What they do
  - How they do things
  - Their opinions on current activities
  - How much they like one thing compared with another

Develop Solution

Develop: Give different answers to the clearly defined problem, seeking inspiration from elsewhere and co-designing with a range of different people.



Storytelling

Construct situations where a specific user in a specific context would go about solving the problem with different solutions.

Rapid Prototyping

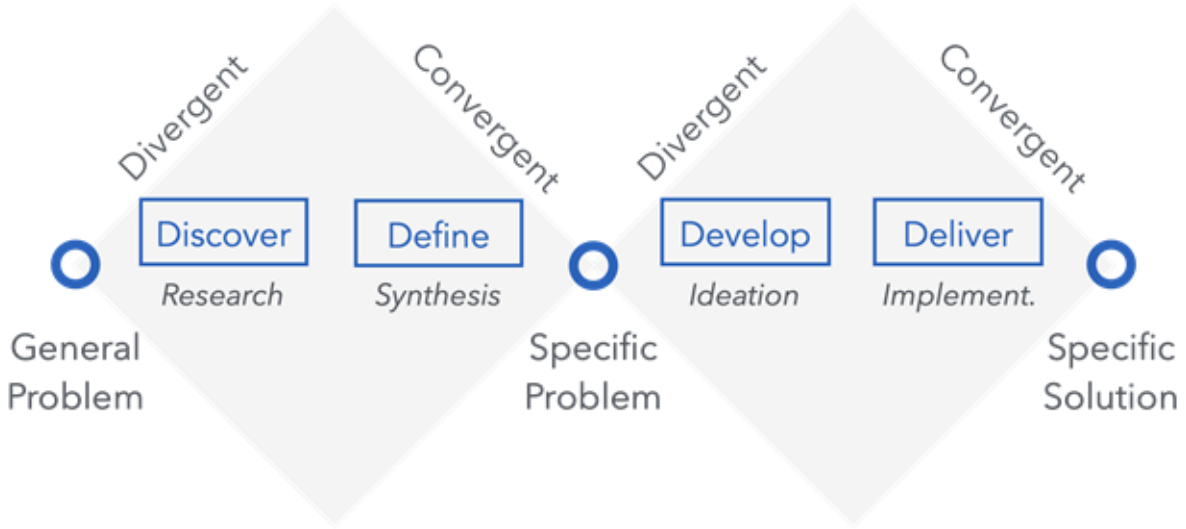
Physical realizations of the research and design process in a tangible form. Can be used to get a sense of what it would be like to experience the product/service.

Goes from low fidelity (paper) to high fidelity (systems).

Minimum Viable Product

Quesenbery, Whitney, and W. Whitney. "Choosing the right usability technique: Getting the answers you need." User Friendly 2008-Innovation for Asia (2008).

Deliver Solution



Deliver: Involves testing out different solutions at small-scale, rejecting those that will not work and improving the ones that will.

Methods:

Survey;
Think Aloud;
Usability testing
...

Develop and Deliver: Evaluate Usability Results

These techniques include both measuring the success of a design (*against usability performance and satisfaction criteria*) and establishing benchmarks metrics.

They require a more formal test protocol, and realistic tasks.

To…	Use…
Determine whether a product is meeting its usability goals	**Summative usability testing,** measuring performance against criteria (and possibly benchmark values) ▪ Lab or field setting
Learn how a product compares to its competitors	**Comparative testing** with the same tasks performed using two or more products
Find out whether users like a product	**Satisfaction surveys**, as part of a usability test or with random users. ▪ Before release ▪ After release
Test a design against scenarios of use	**Usability testing** in formal or informal settings: ▪ Testing in a usability lab ▪ Testing in an informal lab space ▪ Testing in the users' own setting ▪ Remote testing
Understanding what parts of the interface draw the user's visual attention	**Eyetracking** lets you see exactly where a user looks on the screen, and for how long.
Ensuring access for all	**Usability testing** with people with disabilities

I am an incoming CS PhD student at Stanford. What are some tips for finding an advisor?



MODEL

Congrats on starting your PhD at Stanford! Here are some tips on how to find an advisor:

Attend group meetings: Many labs at Stanford are open to first-years sitting in.
Look for an advisor whose research aligns with your interests



Distributed across senior PhD students

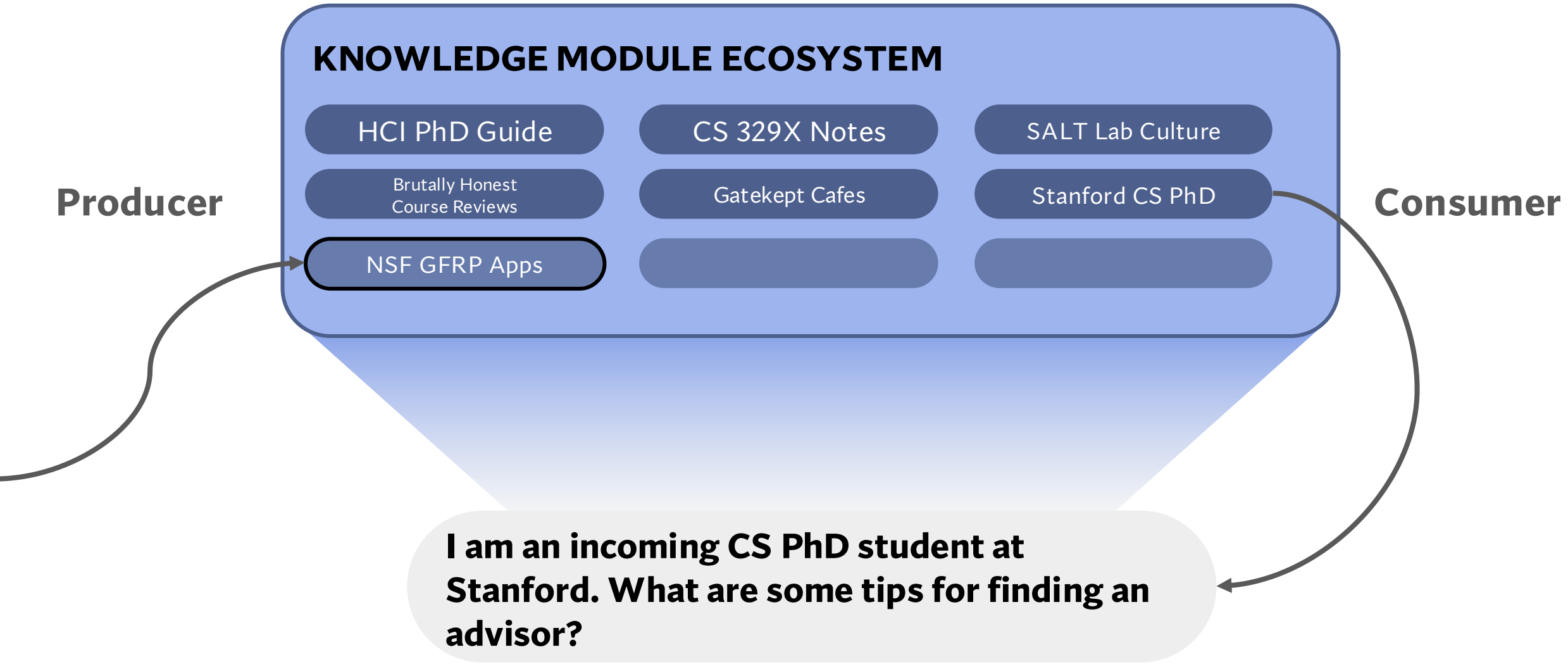
In personal notes, texts, documents

Tacit, passed down knowledge

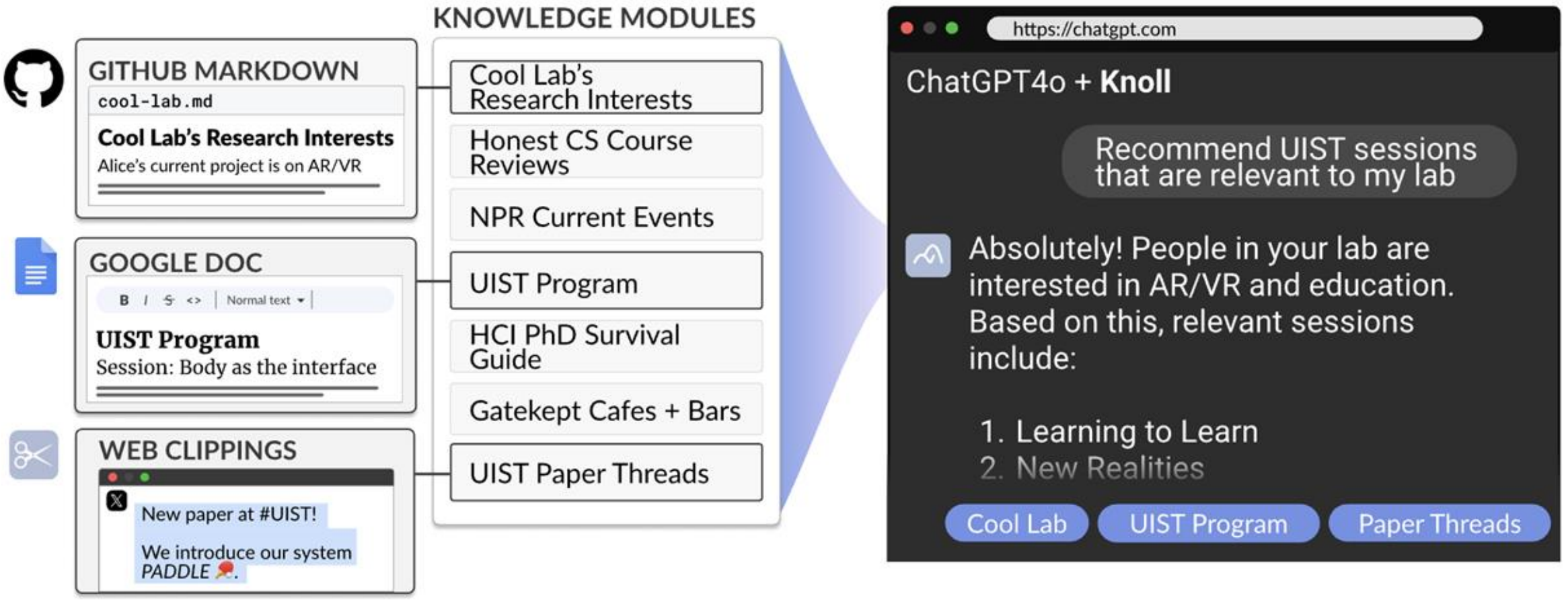
2ND YEAR PHD STUDENT

The CS department at Stanford has a rotation system. Set up your winter (and spring) rotations as soon as you can in the fall quarter.

If you're not sure who to rotate with, your faculty buddy is a good place to start or other people listed on your admissions letter.


KNOWLEDGE MODULE ECOSYSTEM
HCI PhD Guide
CS 329X Notes
SALT Lab Culture
Brutally Honest Course Reviews
Gatekept Cafes
Stanford CS PhD
NSF GFRP Apps
Producer
Consumer
I am an incoming CS PhD student at Stanford. What are some tips for finding an advisor?

Knoll



Zhao, Dora, Diyi Yang, and Michael S. Bernstein. "Knoll: Creating a Knowledge Ecosystem for Large Language Models." *UIST 2025*.

Principles of commons-based peer production



Modularity

Tasks are divisible into independent components



Granularity

Tasks can be broken into pieces of various sizes



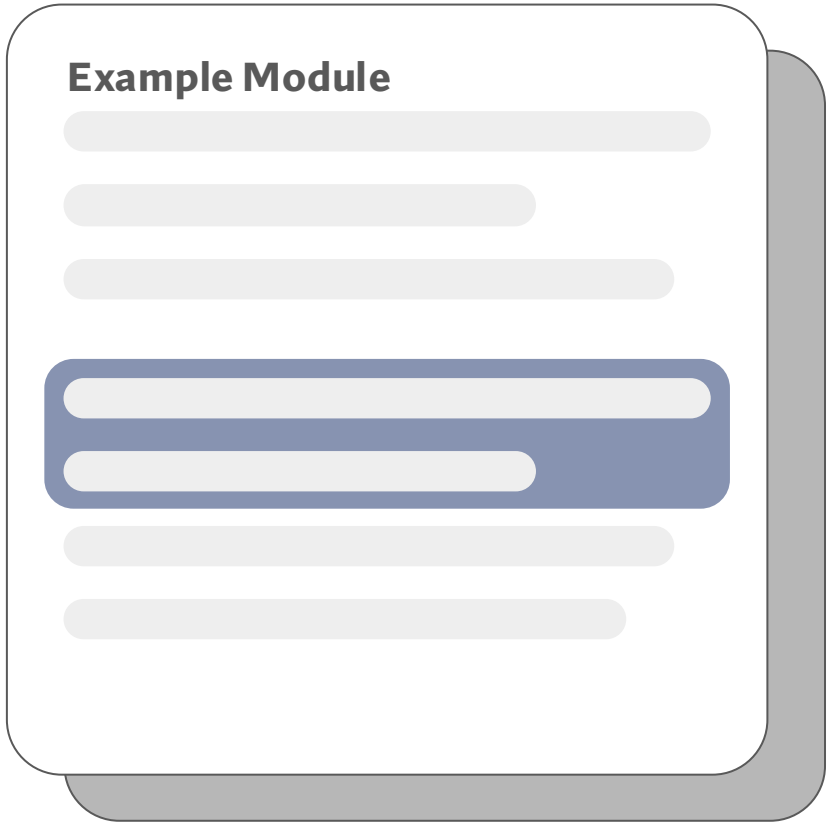
Low-cost Integration

Integrating modules into the end-product should be easy

Benkler and Nissenbaum. “Commons-Based Peer Production and Virtue.” *Journal of Political Philosophy, Vol. 14, No. 4, 2006.*

Using and maintaining knowledge modules

- Update modules by submitting “PRs”
- Adapt modules via “forking”
- Easily shareable links to modules
- Control over who can access and edit knowledge



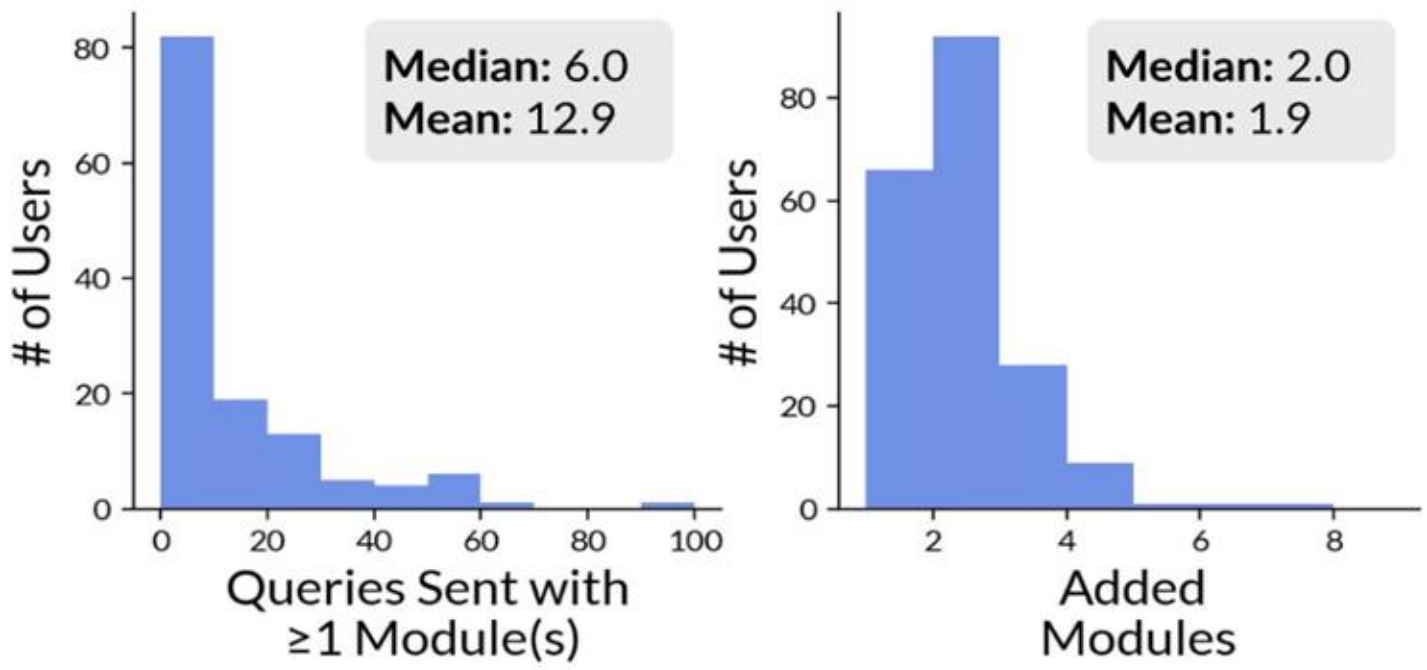
Usage overview

2-month public deployment

203 organic users

1,690 queries sent

135 active users



Knoll produces more relevant and higher quality responses

	Target Value
Win Rate (Relevant Info.)	100.0%
Win Rate (Distracting Info.)	50.0%

Error cases in-the-wild

YOU

Is CS 123 offered in the spring?

MODEL

Based on the knowledge provided, I don't see CS 123.



Models failing "loudly"

YOU

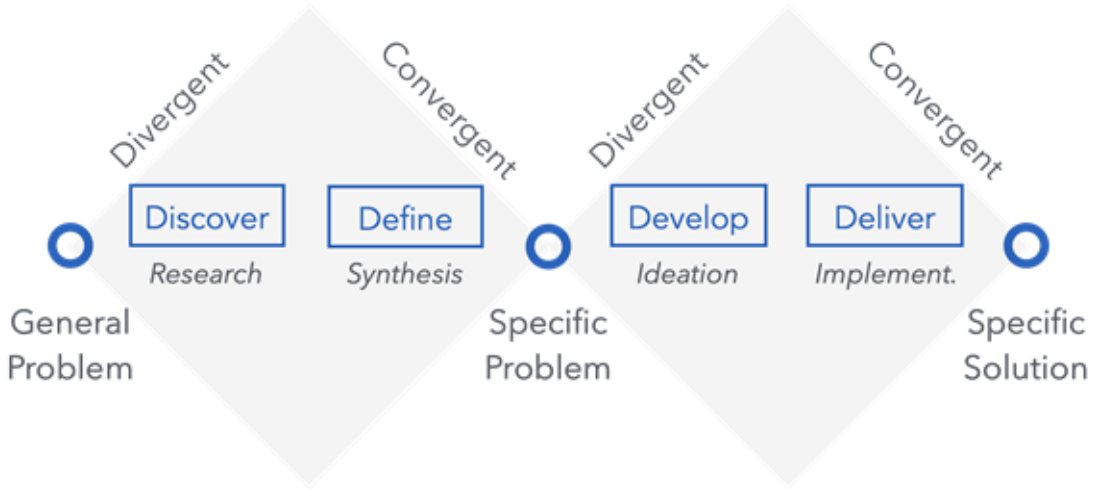
Recommend papers about end-user programming

Here are some papers from CHI 2025 on end-user programming



Overindexing on provided information

Summary of The "Double Diamond" Method



User-centered design is **important**.

Double Dimond is a typical process.

Reframing the problem and the persona changes human behaviors.

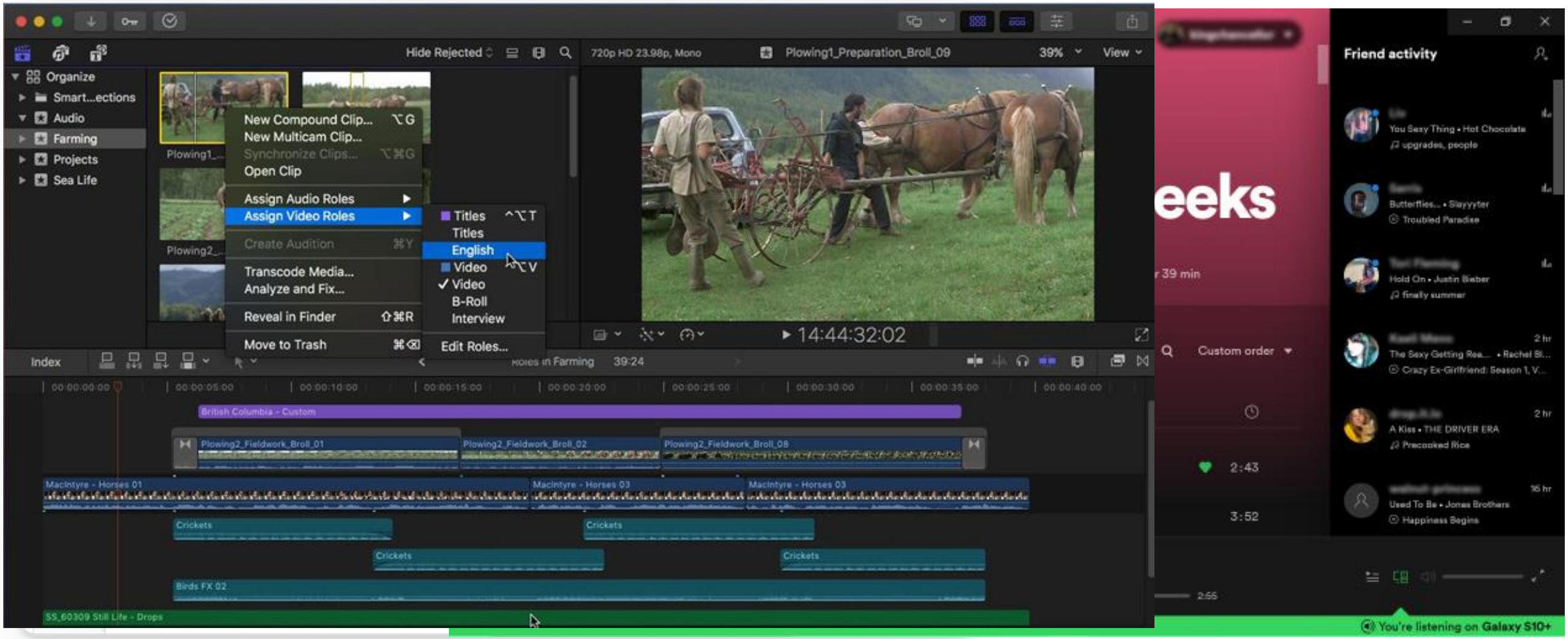
Interviews & think-aloud are important HCI methods for building NLP-infused applications.

Quantitative & qualitative studies are both important.

Outline

- **Design Thinking** (25 mins)
- **Generative UI** (15 mins)

Good user interfaces have a tremendous impact on individuals' ability to accomplish things and on our society as a whole!



Slides credit to Xu Wang (UMich) on User Interface Design

Bad designs can cost people's lives



Emergency telephone where you're supposed to dial 999. However, the physical telephone itself only has the numbers 1, 2, 3.

How to design good user interfaces?

Examples Types of UI

- Graphical User Interfaces
- Augmented / Virtual Reality
- Conversational / Speech Interfaces
- Ambient / Ubiquitous Interfaces
- Embodies Agents, e.g., robots

- Guiding theories and principles
  - E.g., visible to users, etc.
- **Prototype and iterate**
- **Get feedback from users**

Design is about “how things work” not “how things look”

The user is not you!

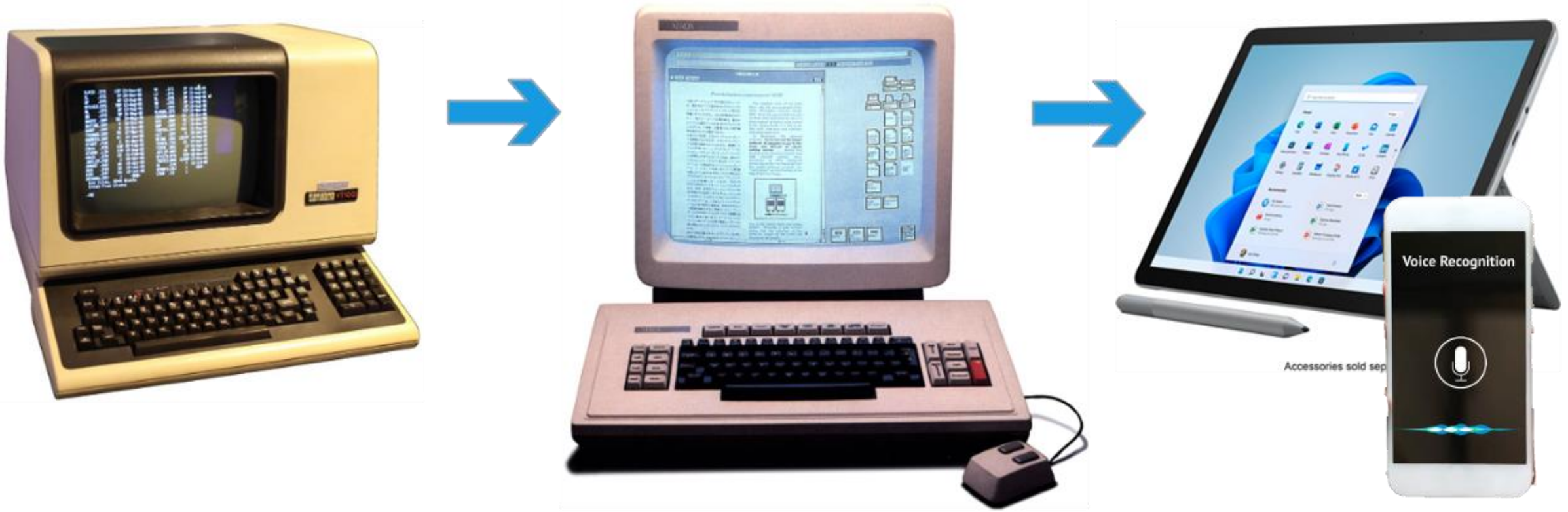
Natural Language as the New User Interface

This section is modified based on tutorial by Diyi Yang, Sherry Wu, Marti Hearst on “Human-AI Interaction in the age of LLMs”.

Slides for Natural Language as the New Interface credited to **Marti Hearst**

Human Are Tool Users

Tool design is determined by the available tech and how people interact with the tool



Command-based User Interfaces



pros:

- \+ precise
- \+ efficient input
- \+ portable

cons:

- \- inflexible input
- \- hard to learn / remember
- \- output display is limited

Graphical User Interfaces (GUIs)



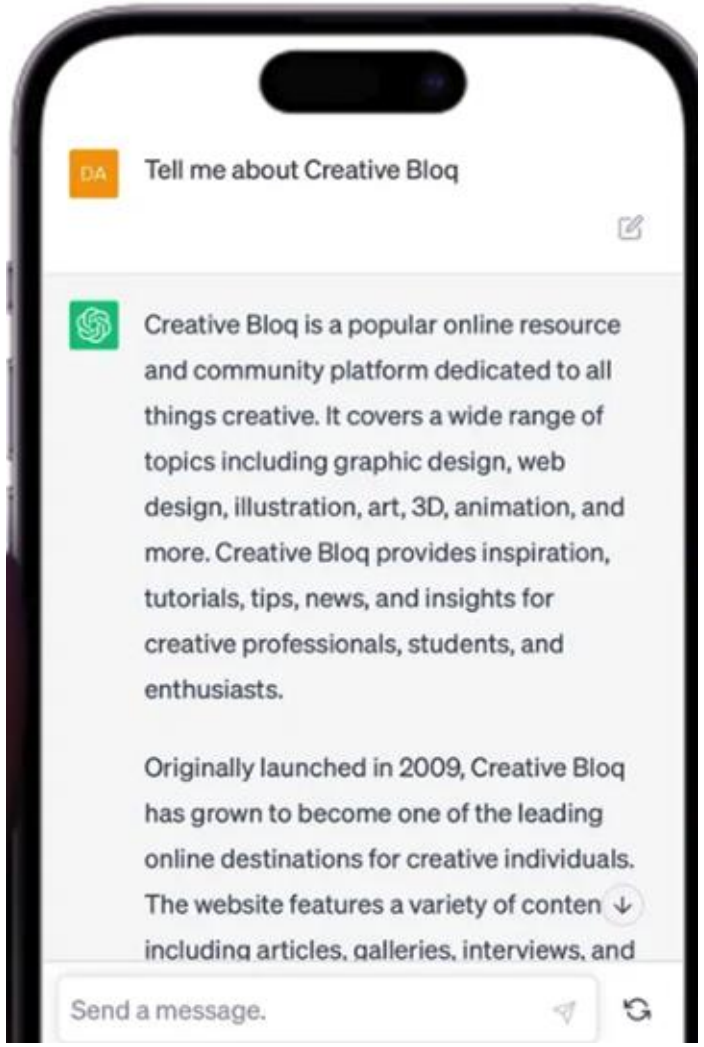
pros:

+ pointing / gestures / voice
+ flexible display / feedback
+ easier to learn / remember

cons:

- less precise
- less efficient input
- platform-specific

Chat-based User Interfaces



pros:

- \+ flexible, natural input
- \+ easy to learn / remember
- \+ portable

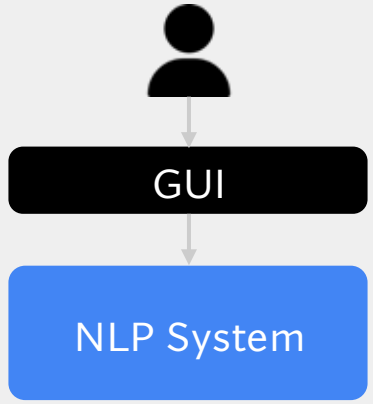
cons:

- \- imprecise
- \- feedback constrained
- \- failed expectations

NLP & Interfaces

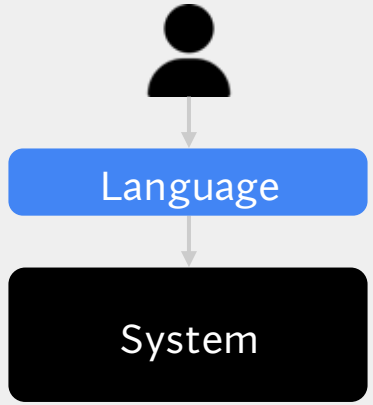
UI for an NLP system aka “A wrapper on top”
e.g., Autocomplete, Google Translate

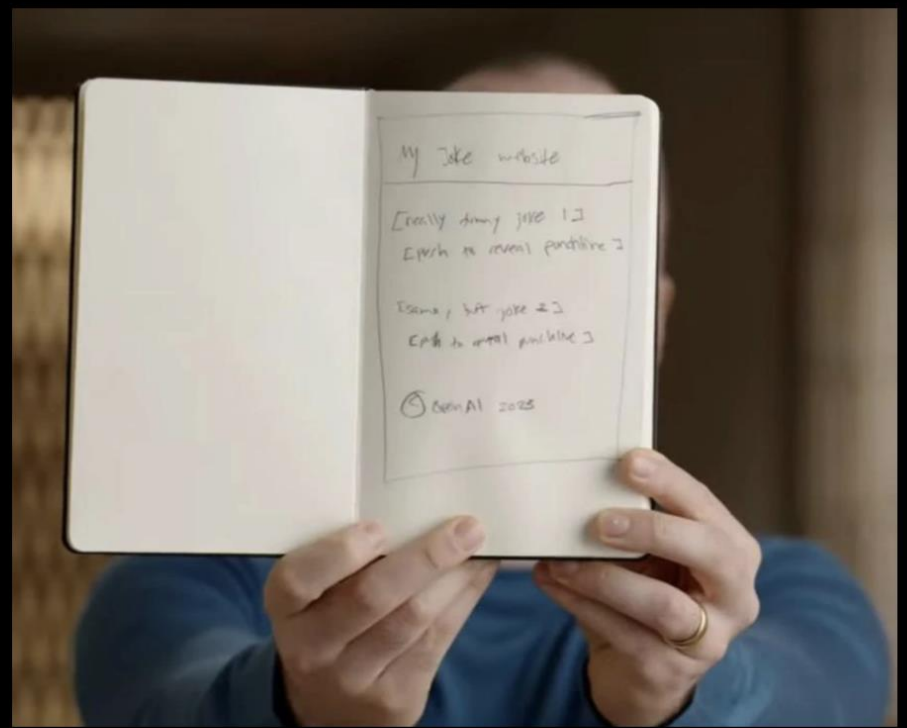
- Designing an interface to make use of the underlying NLP system for achieving user tasks.

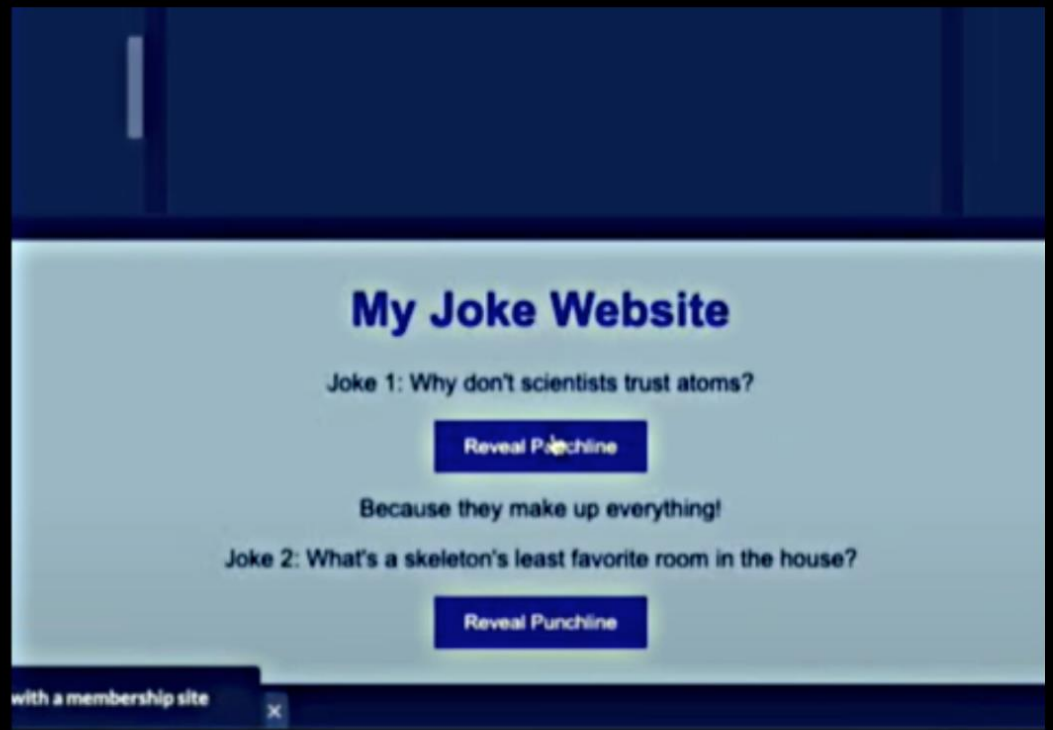


Natural language UI aka “the wrapper itself”
e.g., Alexa, Google Search, GPT4 Web Search, AutoGPT

- Using language as a medium to interact with tools, applications and other systems.




My Joke website


My Joke Website
Joke 1: Why don't scientists trust atoms?
Reveal Punchline
Because they make up everything!
Joke 2: What's a skeleton's least favorite room in the house?
Reveal Punchline
with a membership site



WebSight (Huggingface)



WebSight (Huggingface)



WebSight (Huggingface)



WebSight (Huggingface)



Design2Code (Ours)



Design2Code (Ours)



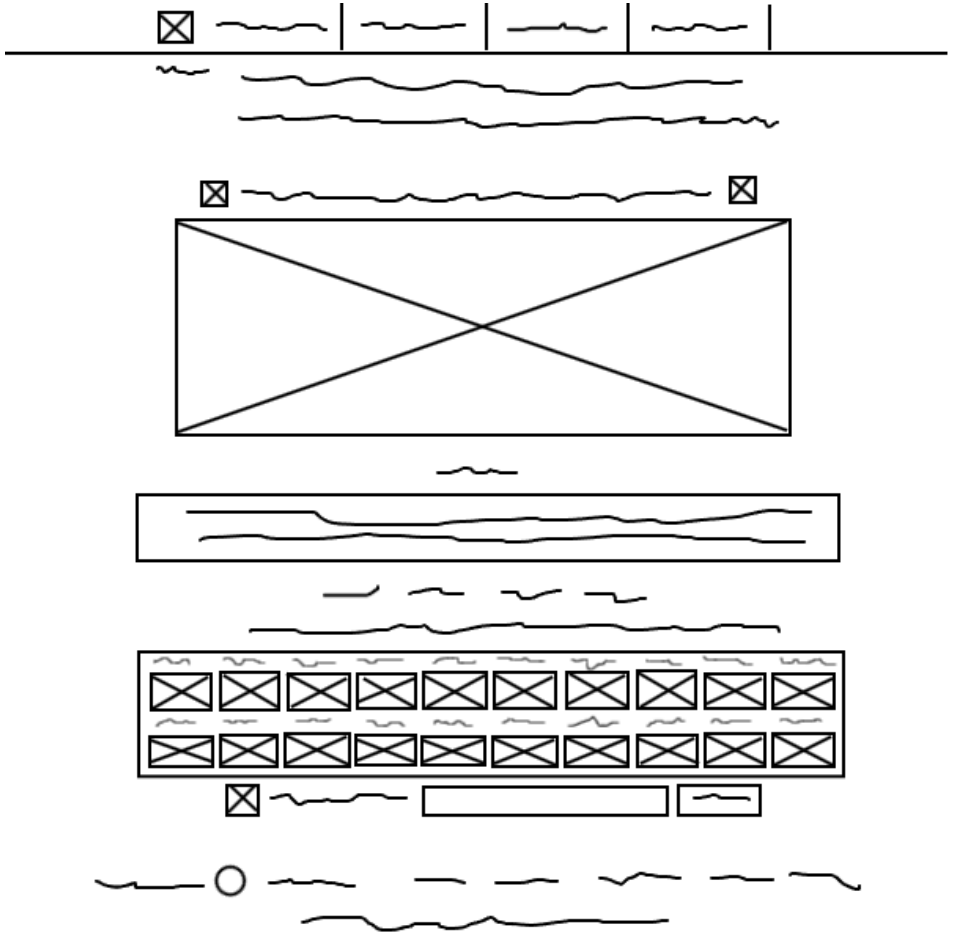
Design2Code (Ours)



Design2Code (Ours)

Si, Chenglei*, Yanzhe Zhang*, Zhengyuan Yang, Ruibo Liu, and Diyi Yang.
"Design2Code: How Far Are We From Automating Front-End Engineering?." arXiv preprint arXiv:2403.03163 (2024).

What if using sketch rather than screenshot?

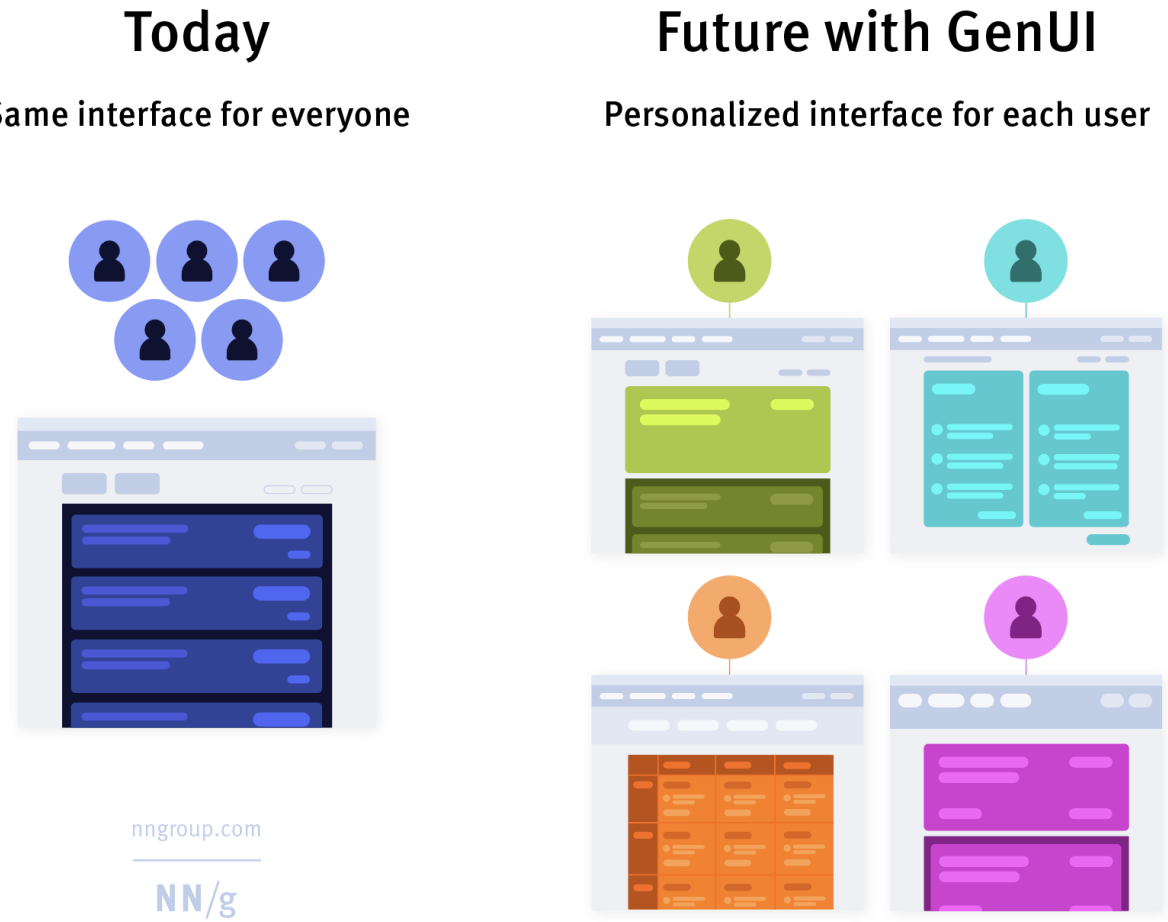


Li, Ryan, Yanzhe Zhang, and Diyi Yang. "Sketch2code: Evaluating vision-language models for interactive web design prototyping." arXiv preprint arXiv:2410.16232 (2024).

Defining **Generative UI**

"A generative UI is a user interface that is dynamically generated in real time by artificial intelligence to provide an experience customized to fit the user's needs and context"

https://www.nngroup.com/articles/generative-ui/



Generative Interfaces for Language Models

Chen, Jiaqi, Yanzhe Zhang, Yutong Zhang, Yijia Shao, and Diyi Yang. "Generative interfaces for language models." *arXiv preprint arXiv:2508.19227* (2025).

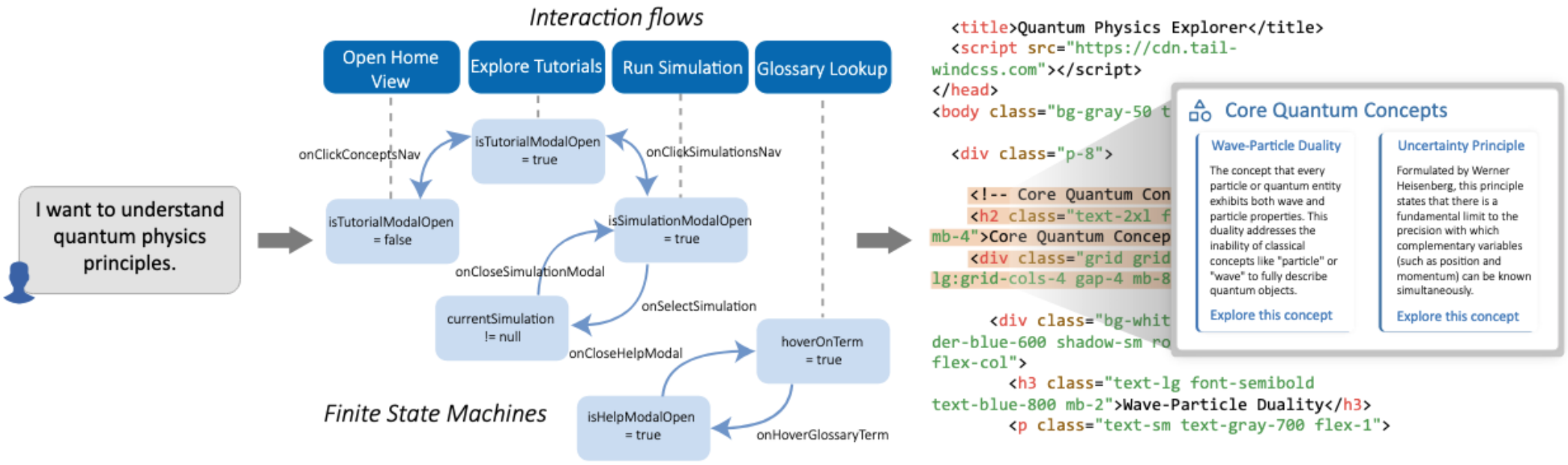
Generative UI

Generative Interfaces for Language Models

Jiaqi Chen*, Yanzhe Zhang*, Yutong Zhang, Yijia Shao, Diyi Yang
Stanford University
*Equal contribution

Generative Interfaces infrastructure

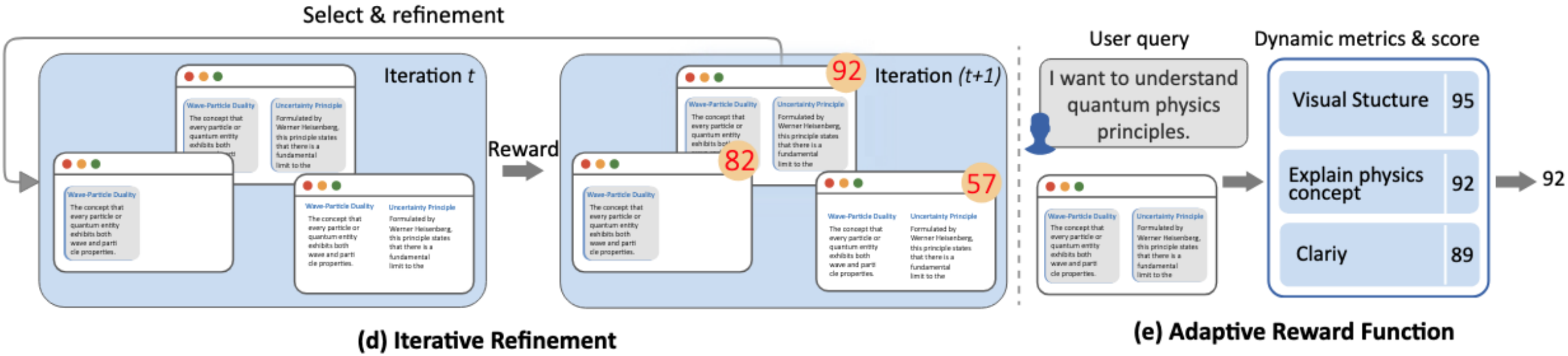
1. User query => Structured Representation
   - High-level: Interaction Flows
   - Low-level: Finite State Machines



(a) User Query **(b) Structured Interface-Specific Representation** **(c) Generated Code and UIs**

Generative Interfaces infrastructure

1. User query => Structured Representation => UI
2. Iterative Refinement
    - Adaptive Reward Function



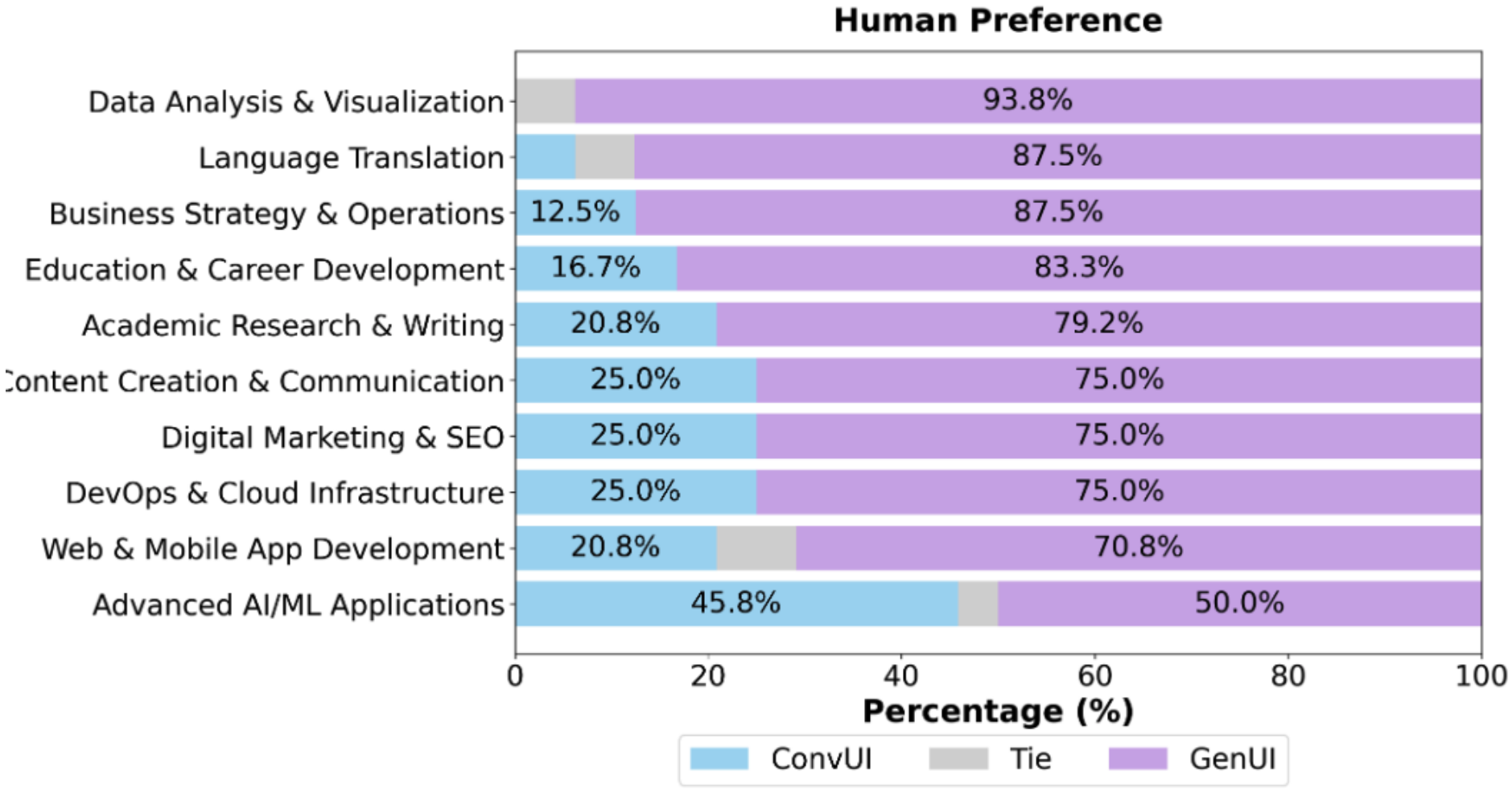
What Defines the Quality of LLM Interfaces

Nielsen et al. (2012) and Hartmann et al. (2008),

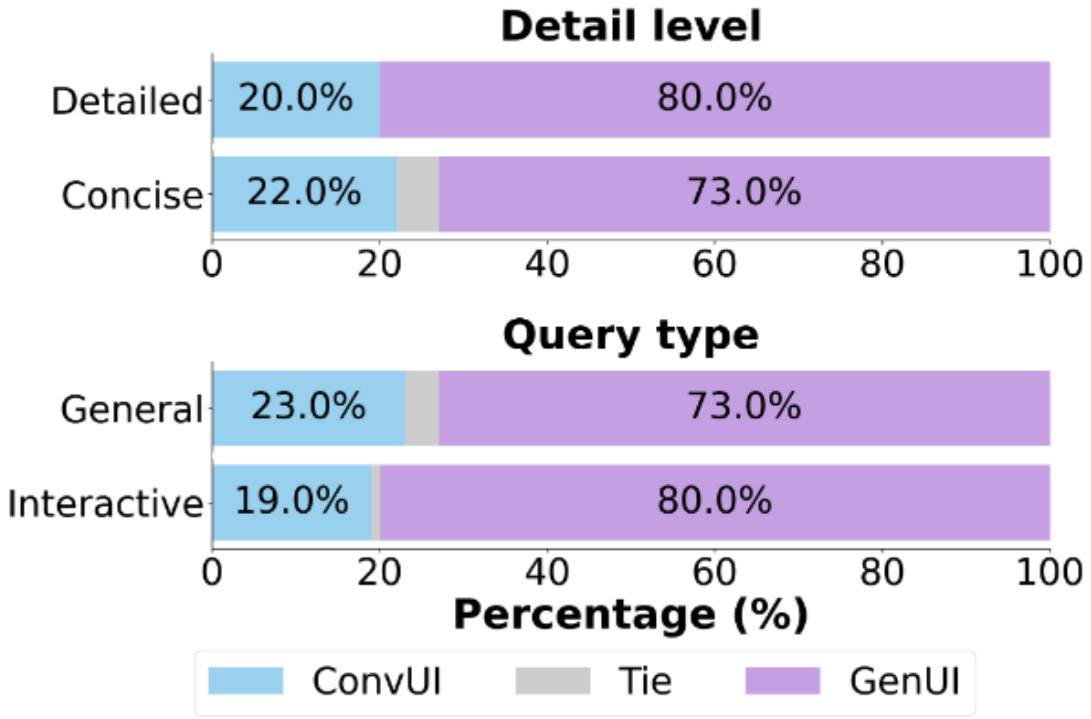
- Functional Perception
- Interactive Perception
- Emotional Perception

Evaluating Generative UI

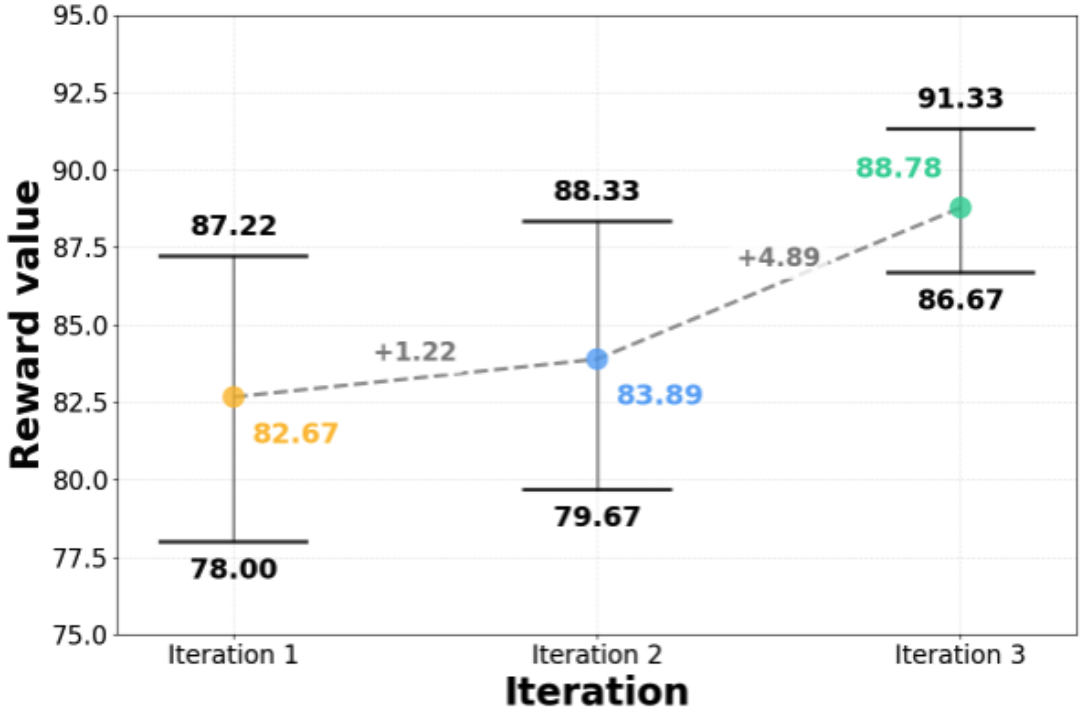
GenUI is preferred because of visual aesthetics, information organization, and reduced cognitive load. User preferences are query-dependent and domain-dependent.



Human Evaluation Comparing GenUIs and ConvUIs



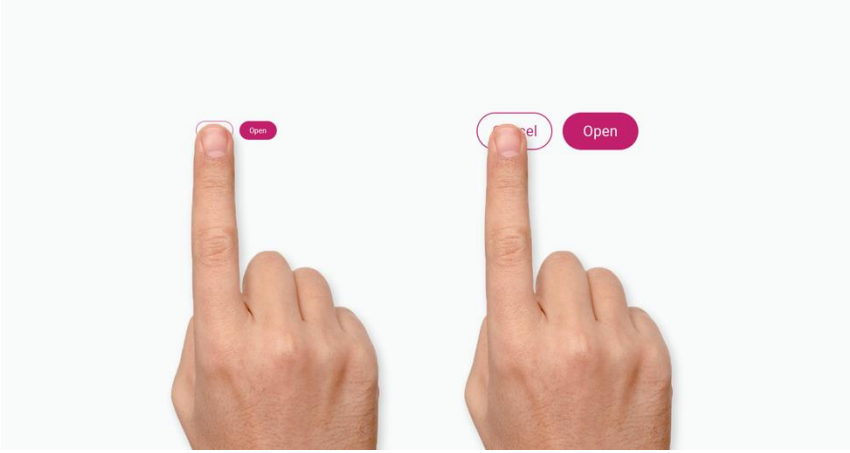
(a) **Breakdown of query detail level and type.**

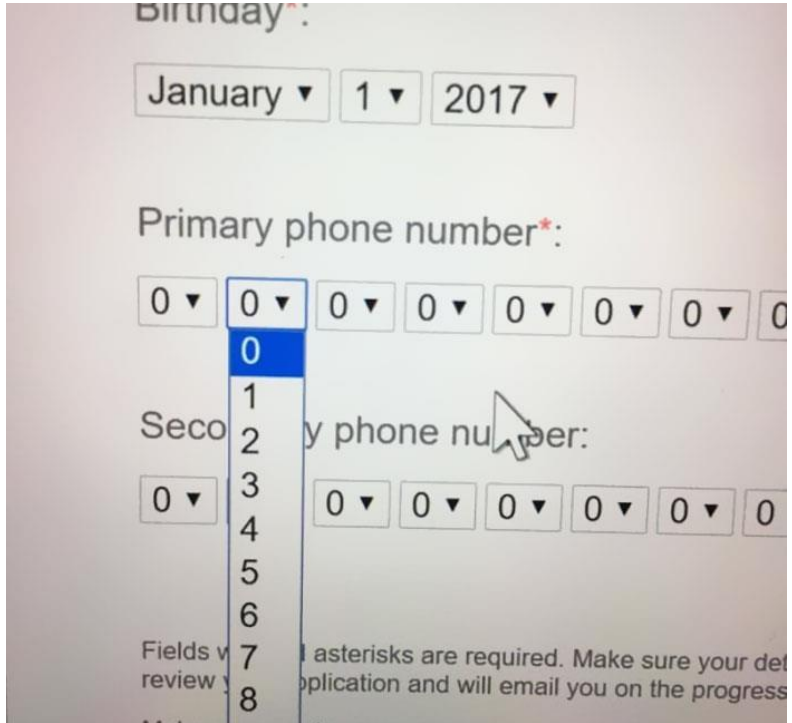


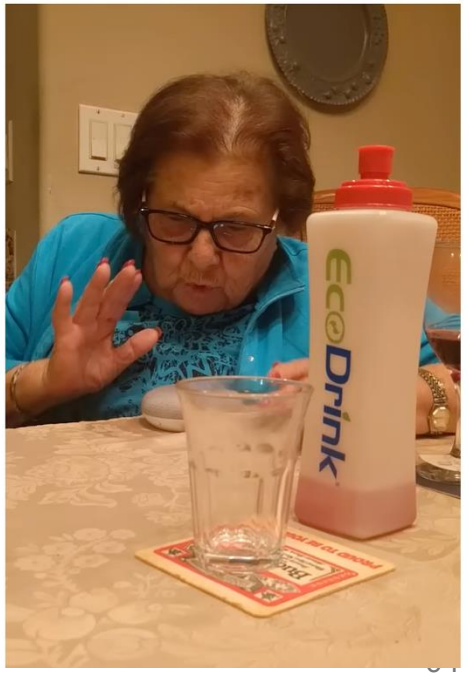
(b) **Ablation on number of iterations.**

Reflections on Generative UI

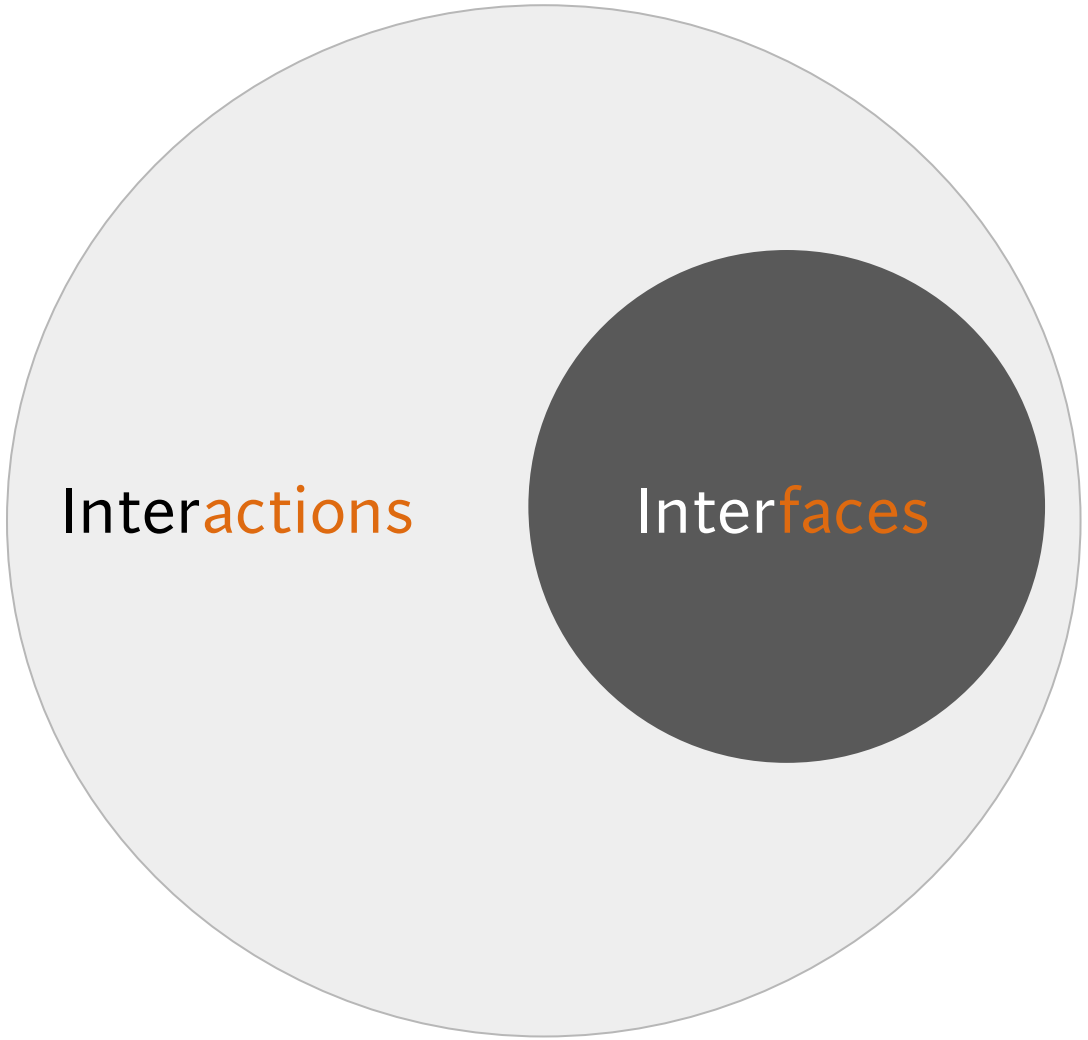
- Shift from designing interfaces to interactions
- Shift from designing for the majority to everyone







Interfaces vs. Interactions



Interfaces

Tangible i.e. visual, auditory, tactile inputs to the system. Designing interfaces are at the surface & physical level, often limited to styling.

Interactions

Understanding the context of user, their needs, how they operate. Includes psychological aspects such as trust, goals, user behavior.