



CS 329X: Human Centered LLMs
HCI & Design Thinking

Diyi Yang

Outline

- **Design Thinking**
- **Natural Language As the New Interface**
- **Small-Group Discussion** (20 mins)

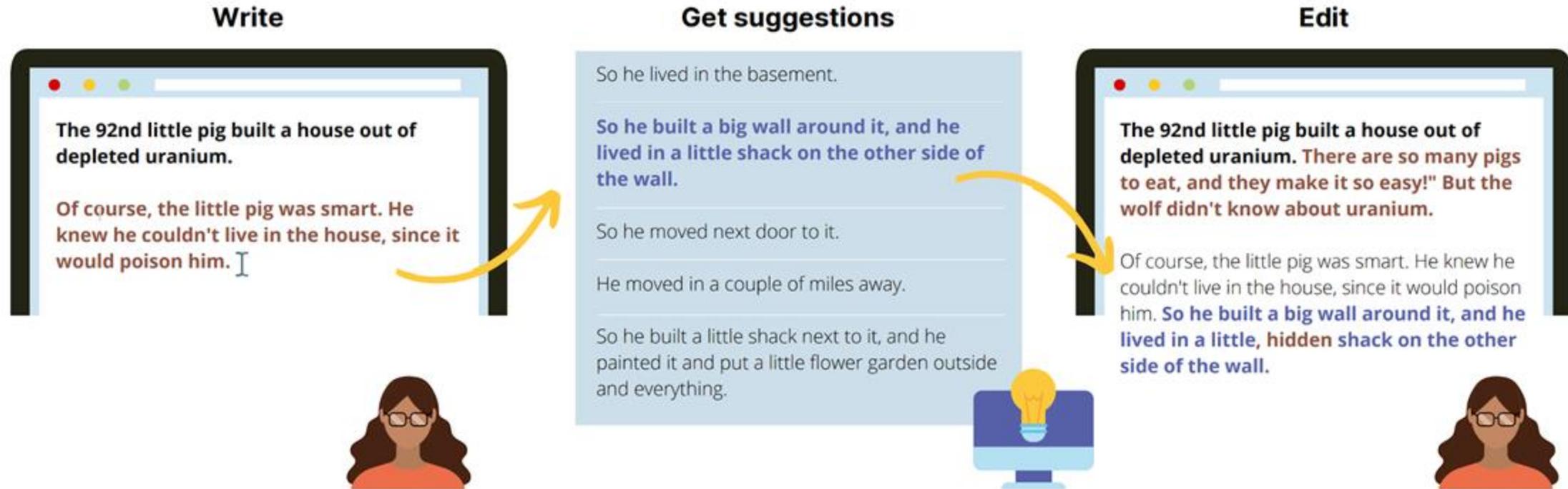
Learning Objective: understand design thinking and HCI methods

Human-AI Interaction

- **Humans:** AI researchers, model developers, domain experts, end users.
- **AIs:** dialog system, translator, recommender system, autonomous driving system, LLMs, agents...
- **Interact:**
 - Humans collaborate with AI,
 - Humans get assistance from AI,
 - Humans analyze AI,
 - AI helps human,
 - & many other forms

Human-AI Collaboration

The cooperative and coordinated interaction between humans (**mostly non-AI experts**) and AI to solve complex problems or achieve certain goals.



Humans Get Assistance from LLM-empowered Apps

Humans are still mostly end users and domain experts.

The big difference is AI is not a partner, but a tool (and part of “AI-infused applications”)

Examples:

- LLM-empowered tools for doctors, teachers, software engineers



Carnegie Mellon's AI system, Coscientist, autonomously conducts chemistry experiments, significantly advancing scientific research. By using large language models and automating the experimental process, it offers a new level of efficiency and accessibility in scientific research, with an emphasis on safety and ethical use. Credit: SciTechDaily.com

<https://scitechdaily.com/automating-scientific-discovery-carnegie-mellons-ai-coscientist-transforms-lab-work/>

How to Enable 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-LLM interaction?

Evaluate Have we achieved what we want to achieve?

Overview

- **Design Thinking**
 - Motivation: why designs on top of LLMs are important
 - Design Thinking:
 - Double Diamond
 - Problem Reframing
 - Prototyping
 - Interview and Think Aloud Studies

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.

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

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 “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	“ <u>How should I operate the door</u> to get to next room?”	“ <u>How should I prompt the model</u> to get it to solve my task?”
How does a user learn “How to use?”		

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	“ <u>How should I operate the door</u> to get to next room?”	“ <u>How should I prompt the model</u> to get it to solve my task?”
How does a user learn “How to use?”	<ul style="list-style-type: none">- From previous encounter- Read labels- Take a guess and try	<ul style="list-style-type: none">- From other people- Read prompt guidelines- Wing it

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

“The father of ubiquitous computing”

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.

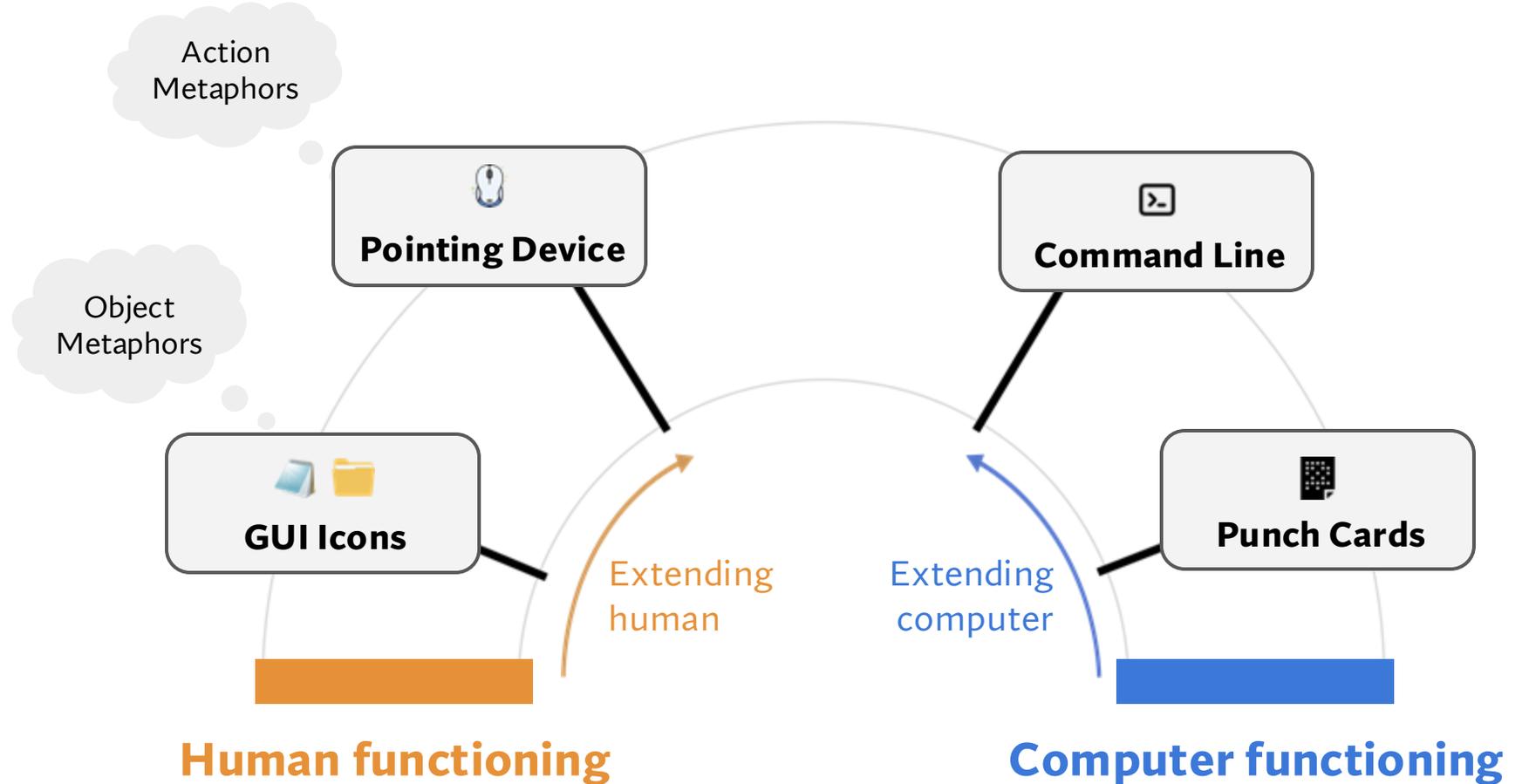
Norman, Don. *The design of everyday things: Revised and expanded edition*. Basic books, 2013.

How do we achieve this?

By extending humans instead of extending technology, reducing “frictions of learning”

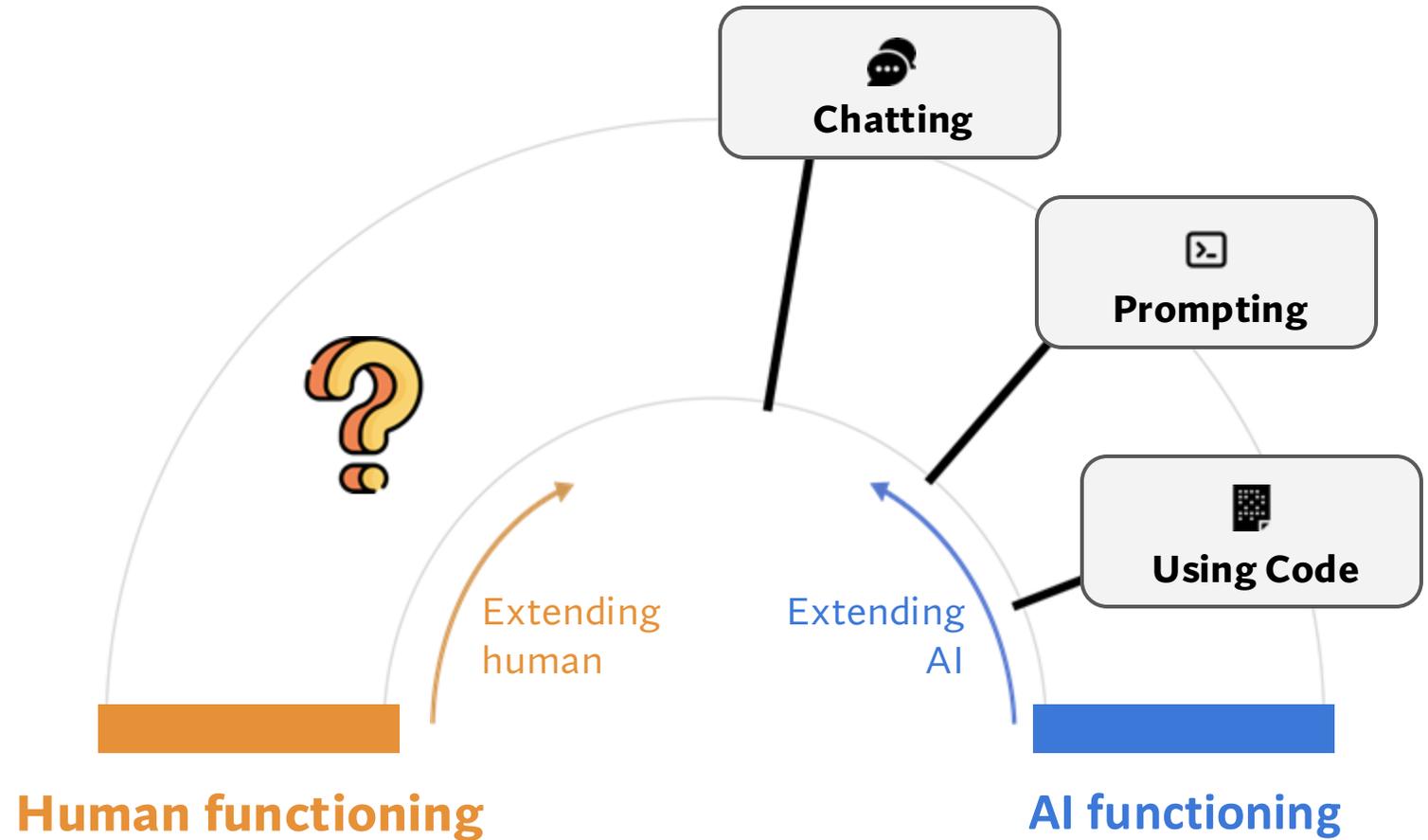
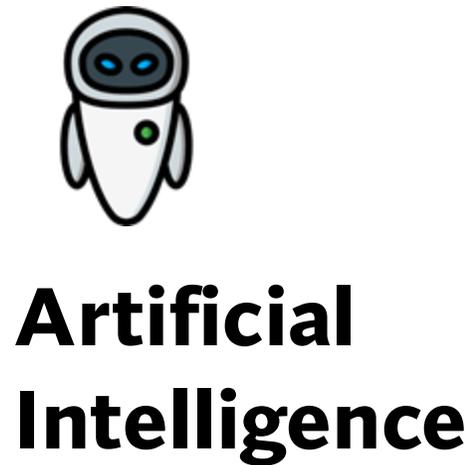


Personal Computing



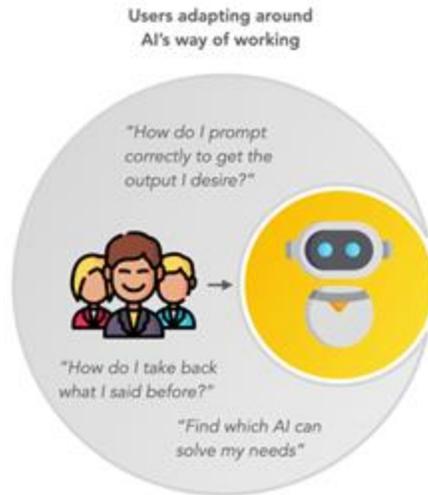
How do we achieve this?

In AI, we've been only thinking about extending technology



Why does it happen?

Technology-centric design



User-centered design

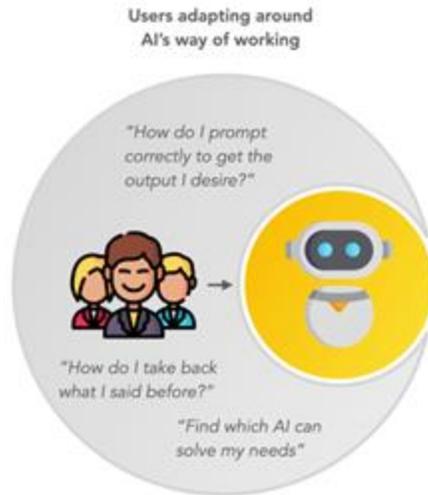


A design process for situations when a class of technologies already exists, but when user domains for co-development are not clearly established.

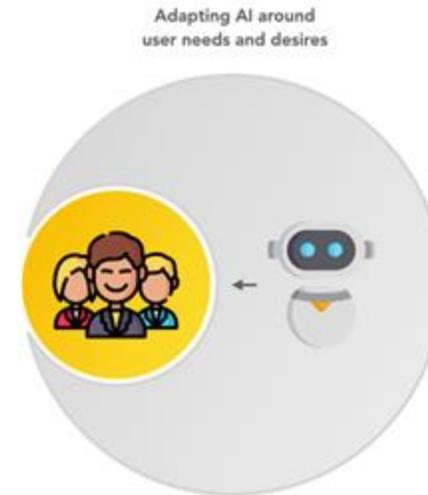
User-centered design (UCD) is an iterative design process in which designers focus on the users and their needs in each phase of the design process.

Why does it happen?

Technology-centric design



User-centered design



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

Solution in search of a problem

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.

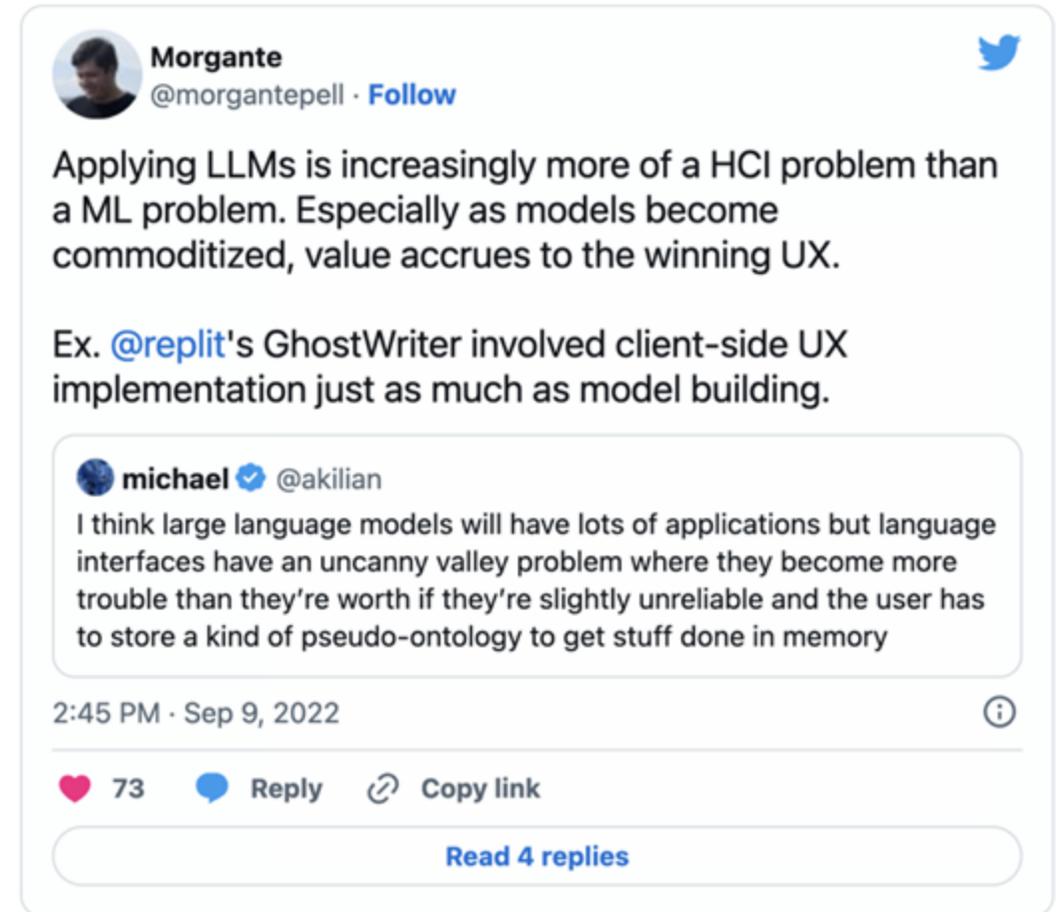
Blog Post: The Biggest Bottleneck for LLM startups is UX

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.”*

The biggest bottleneck for large language model startups is UX
Yang, Qian, et al. “Sketching nlp: A case study of exploring the right things to design with language intelligence.” CHI 2019.



The image shows a screenshot of a Twitter thread. The main tweet is from Morgante (@morgantepell) and discusses the HCI nature of LLMs. A reply from michael (@akilian) discusses the uncanny valley problem of language interfaces. The tweet includes engagement metrics like 73 likes and a 'Read 4 replies' button.

Morgante @morgantepell · Follow

Applying LLMs is increasingly more of a HCI problem than a ML problem. Especially as models become commoditized, value accrues to the winning UX.

Ex. @replit's GhostWriter involved client-side UX implementation just as much as model building.

michael @akilian

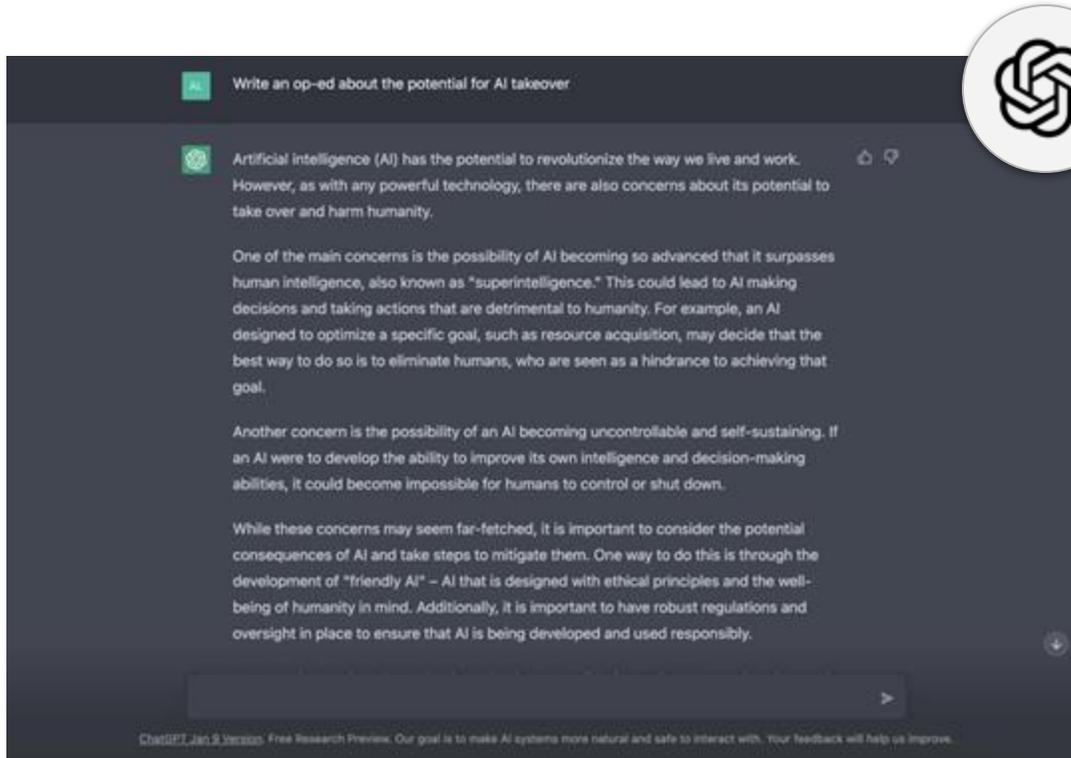
I think large language models will have lots of applications but language interfaces have an uncanny valley problem where they become more trouble than they're worth if they're slightly unreliable and the user has to store a kind of pseudo-ontology to get stuff done in memory

2:45 PM · Sep 9, 2022

73 Likes Reply Copy link

Read 4 replies

New models → new AI interactions



Pushing language as the primary interface for every task

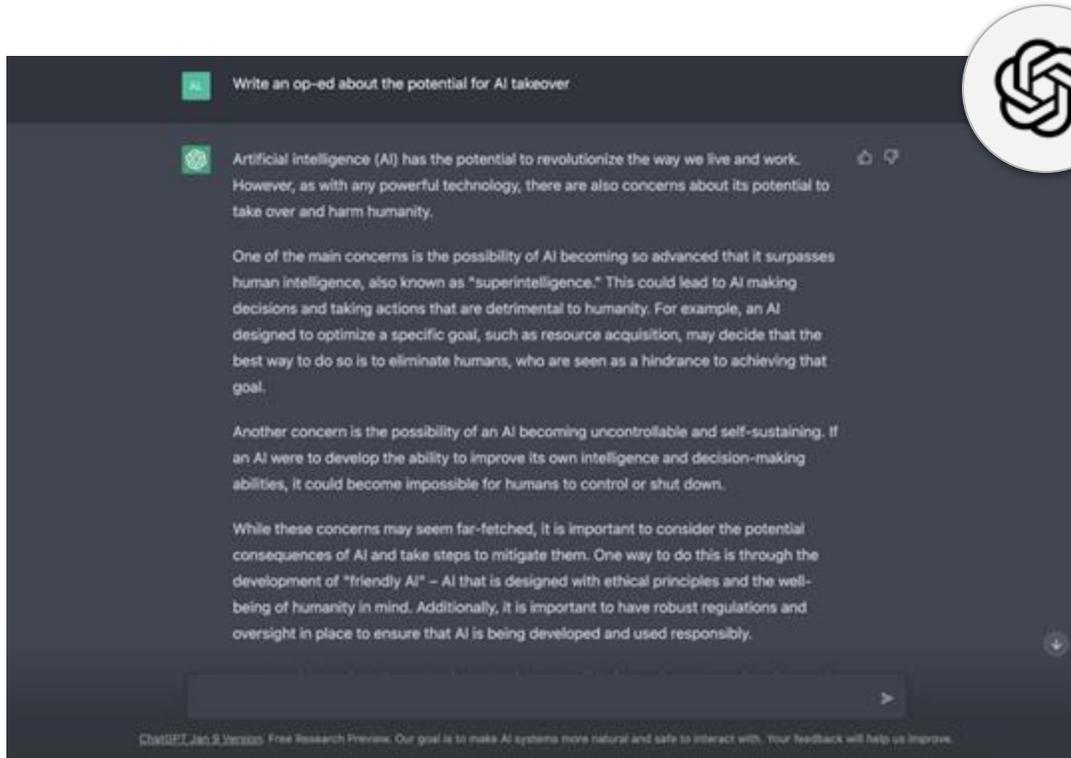


Anthropomorphic Tendencies

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

ChatGPT

New models → new AI interactions



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

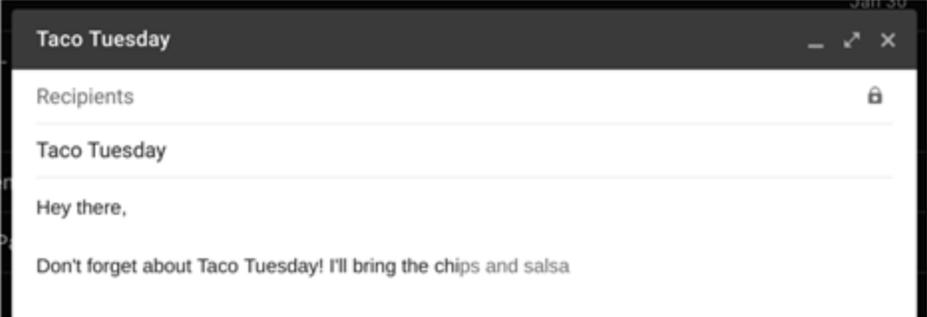
Controlling for desired outputs can be difficult. Unsuitable for tasks that require precision or is critical.

ChatGPT

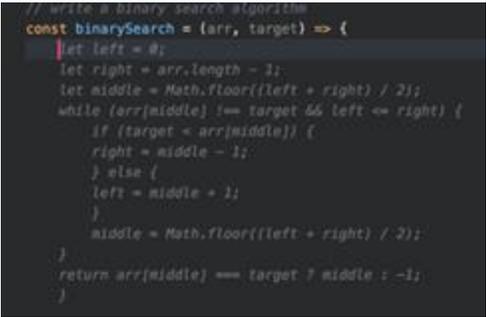
Subtler Interactions (that disappear)



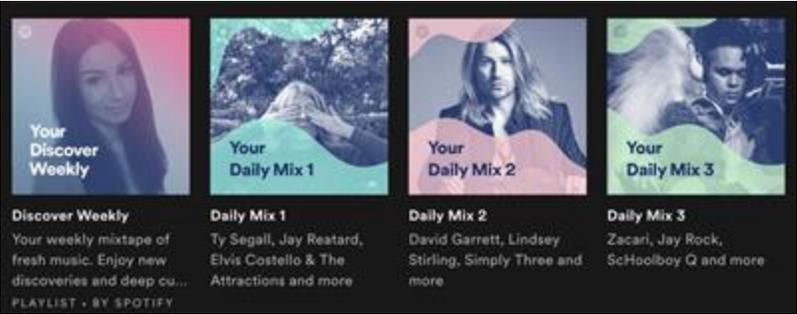
Voice Assistants



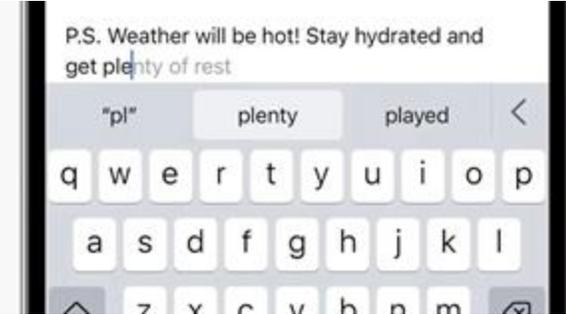
Email Autocomplete



Code Completion



Recommendations

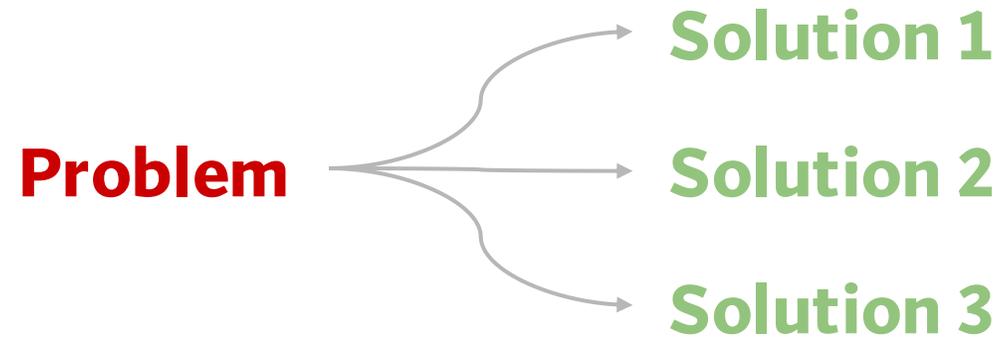


Word suggestions



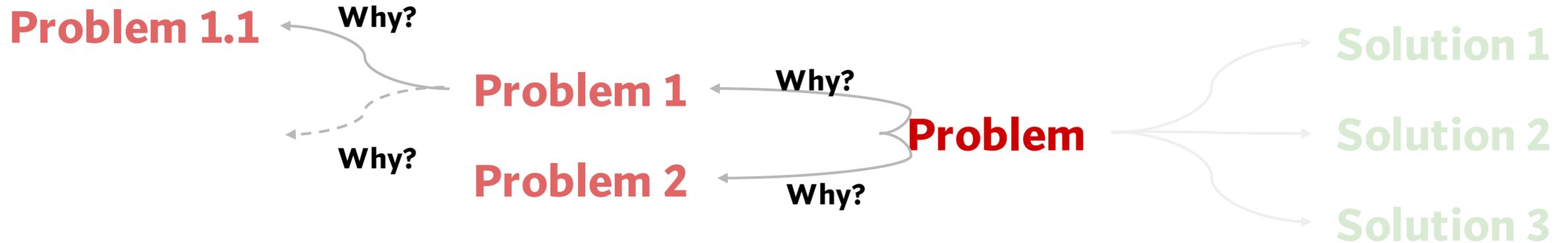
Subtitles

What is design thinking?



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

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.

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

— Henry Ford

“The inventor of production car”



Getting to the root cause of the problem

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

— Henry Ford

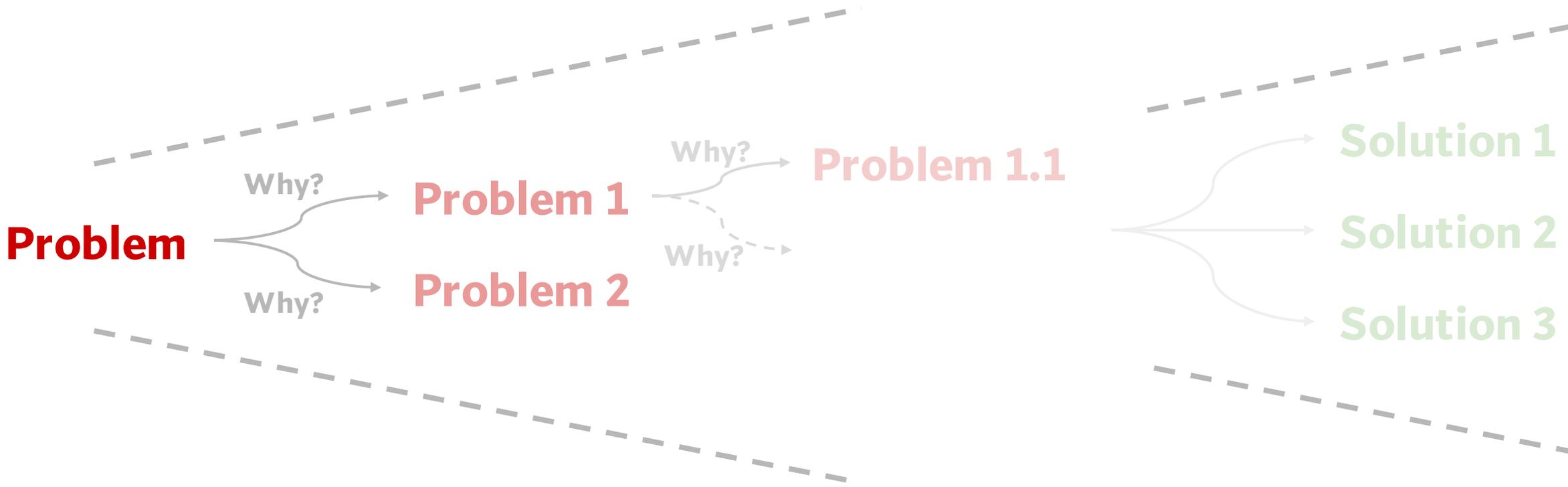
“The inventor of production car”

“Within a design context, framing is often seen as the key creative step that allows an original solution to be produced.

Designers report on the need to get to **‘the problem behind the problem’** (as initially presented by the client), and about creating a ‘fresh perspective.’ ”

— Bec Paton and Kees Dorst

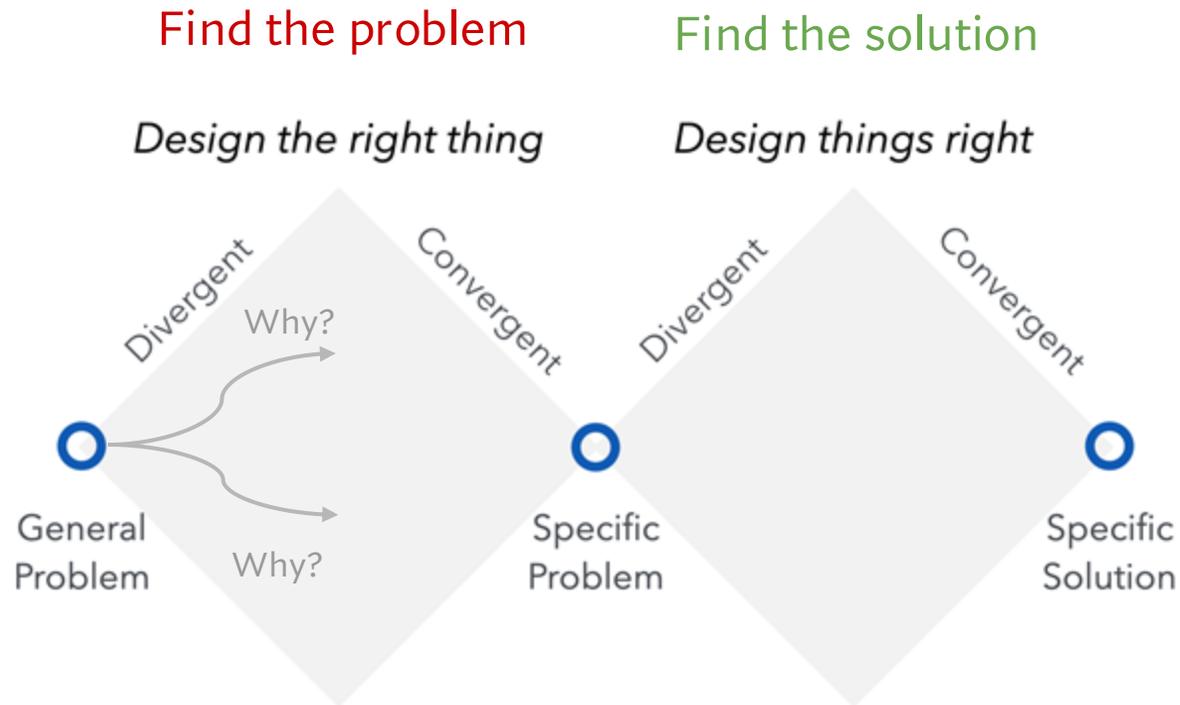
What is design thinking?



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

How to incorporate design thinking?

The “Double Diamond” Method



First Diamond

Find the specific problem.

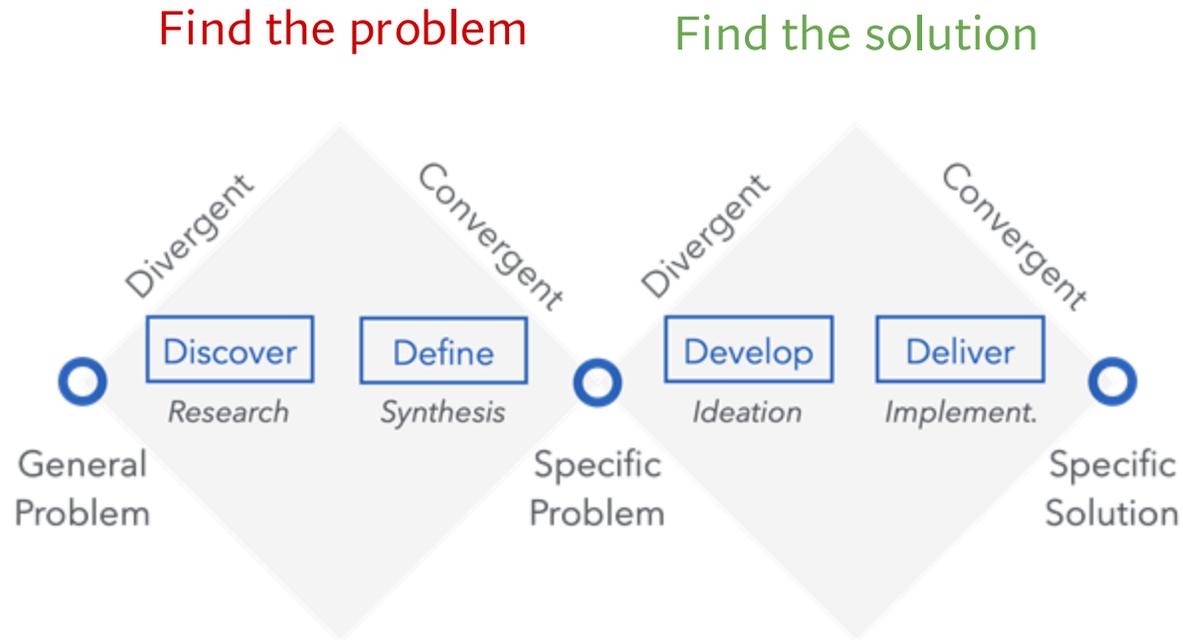
Second Diamond

Find the specific solution.

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

How to incorporate design thinking?

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 Problem

Discover: Understand the issue rather than merely assuming it. It involves researching, speaking to and spending time with people who are affected by the issues.



Field Study

Site visits, Ethnography to observe people doing their own tasks in their own setting.

Interview and Surveys

To collect information on their reactions to existing products and conditions

Environmental Factors

Understand the context, and its needs.

Market Research

Stakeholder Interviews, check raised tickets, traffic and sales analysis, competitive audits

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.

Discover: Learn about business environment

- These techniques focus on **what is happening in the business or personal domain.**
- They are a snapshot of the competitive environment, trends surrounding the product and actual use of the product.

To...	Use...
Learn about a new business environment	Stakeholder interviews to collect input from different areas of the business domain.
Find trends or gaps in a business process	Review problem reports from technical or customer support for usability problems or unmet needs
Understand usage patterns on a web site	Traffic analysis of web site logs, looking for patterns in use, navigation, referrals and related sites or pages
Understand the competition	Competitive audits or comparative usability test with competitive products, or other products and sites that are part of the business domain

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

Discover: Interview

	Structured	Semi-structured	Unstructured	Focus group
Pre-defined questions?	✓	✓	✗	✓
Open-ended questions?	✗	✓	✓	✓
Fixed order of questions?	✓	✗	✗	✗
Fixed number of questions?	✓	✗	✗	✗
Can ask additional questions?	✗	✓	✓	✓

- Semi-structured is **most common**.
- Allows for **exploratory** studies.
- Provides comparable, reliable data, and the flexibility to ask follow-up questions.

Semi-Structured Interviews: Thematic Analysis

- Identify common themes from transcriptions – topics, ideas and patterns of meaning that come up repeatedly
- Define codebook, multiple coders, compute annotator agreement

Interview extract

Personally, I'm not sure. I think the climate is changing, sure, but I don't know why or how. People say you should trust the experts, but who's to say they don't have their own reasons for pushing this narrative? I'm not saying they're wrong, I'm just saying there's reasons not to 100% trust them. The facts keep changing – it used to be called global warming.

Codes

- Uncertainty
- Acknowledgement of climate change
- Distrust of experts
- Changing terminology

How many participants to interview?

Depends on Goals, Context, Resources/Timing.

As many as you

need for finding new things out (data saturation)

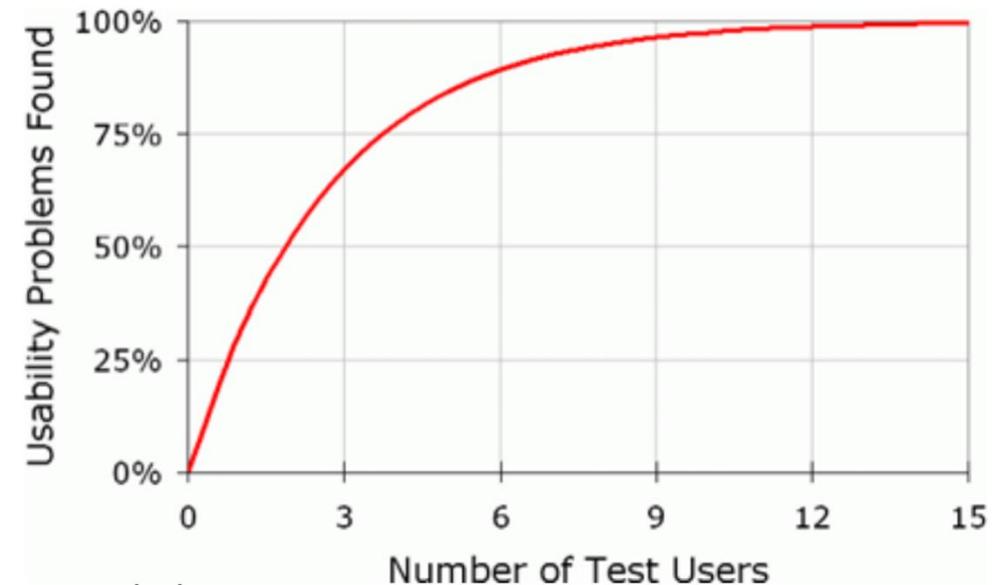
can afford (time, incentives, etc.)

have time to analyze (2x+ per participant)

Magic: 12 is a good number (minimum of five)

Make sure to choose representative users

Or stop when findings start to converge



Define Problem

Define: The insight gathered from the discovery phase can help to define the challenge in a different way.



Perspective Framing

Participatory design to develop a consensus view of the overall process.

Task & Information Analysis

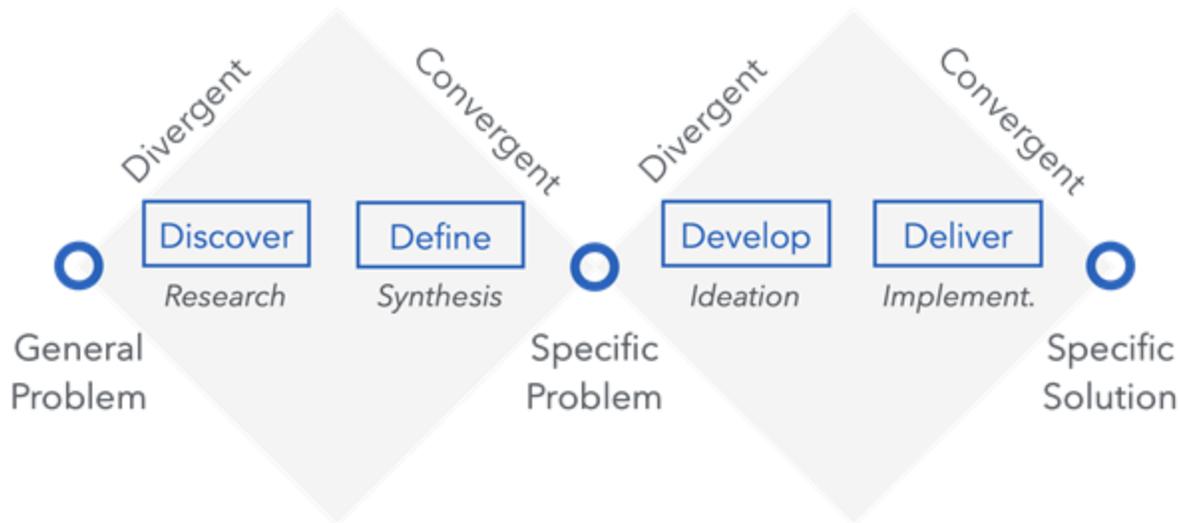
Learning about relationships between tasks and information; Creating logical groups from the users' point of view.

Affinity Diagrams

To group and explore the structure of information.

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.

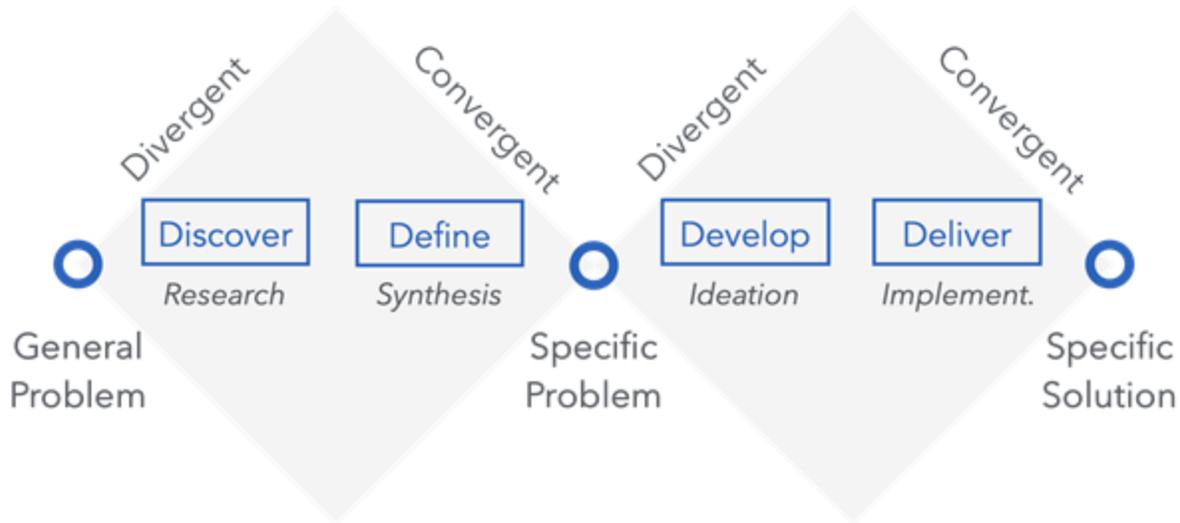
Rapid Prototyping

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

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) <ul style="list-style-type: none">▪ 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. <ul style="list-style-type: none">▪ Before release▪ After release
Test a design against scenarios of use	Usability testing in formal or informal settings: <ul style="list-style-type: none">▪ 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

What is a “Think Aloud”

- A research method used to gain insight into a person’s thought processes as they perform a task or solve a problem. The participant is asked to verbalize their thoughts as they perform the task, which allows the researcher to understand how the participant approaches the task.

"Thinking aloud may be the single most valuable usability engineering method."

What is a “Think Aloud”

“I’m going to ask you to ____ and while you are doing that, can you tell me whatever you are thinking. Whatever comes into your mind while you are working on that. Okay?”

Protocol

Give participants specific tasks to accomplish (but not HOW to do it)

Have them speak aloud as they complete the tasks

Keep interruptions to a minimum

Ask for open-ended questions & clarification after the task is complete

Learning effect - if you make tasks, watch for biasing test due to order

Typically used to test the usability of a website, app or object

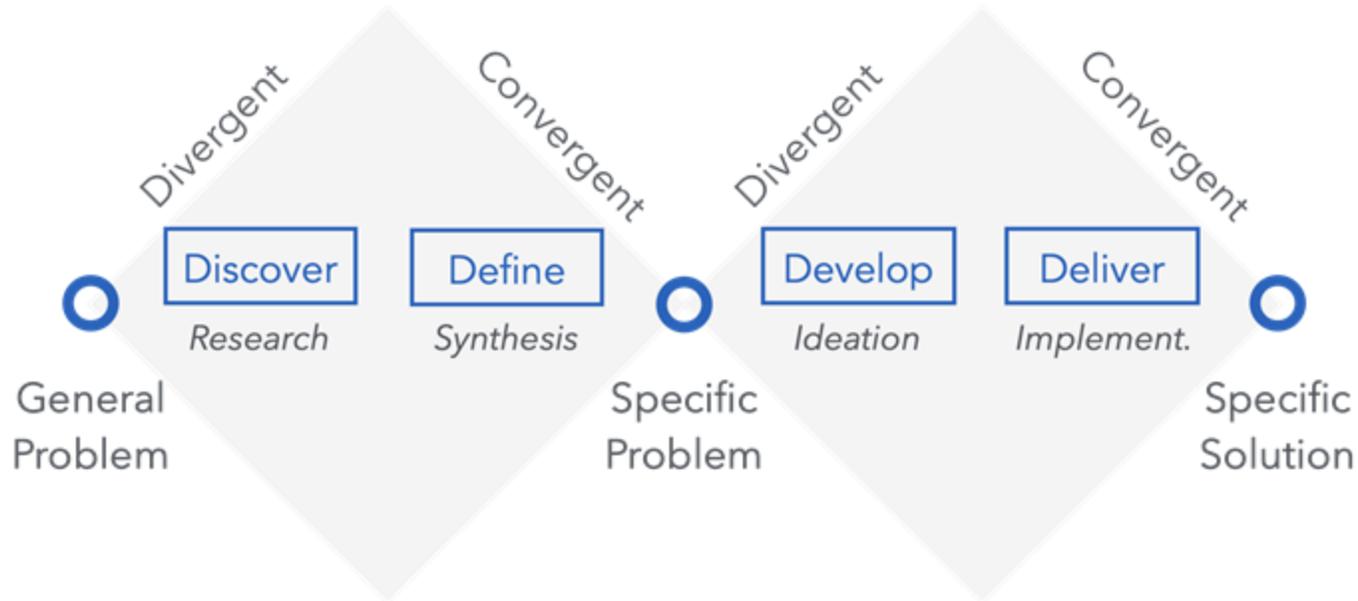
Think-Aloud User Studies

	Quantitative	Qualitative
Definition	Gather numerical data to be analyzed using statistical methods	Gathering descriptive, non-numerical data to be analyzed through interpretation and contextualization
Data source	surveys, questionnaires, experiments	interviews, observations, and document analysis
Presentation	tables, graphs, and statistics	quotes and narratives that reflect the participants' experiences and perspectives
Goal	establish cause-and-effect relationships between variables	gain a deeper understanding of social phenomena, meanings, and processes

Conduct the Think Aloud: Test/Pilot the Study

- Discover problems with study or concept being tested
- Estimate time needed for test
- Refine test script and tasks
- Verify typical tasks (something users actually do?)
- Practice before going live
- New ideas for follow-on questions or things to observe

The “Double Diamond” Method



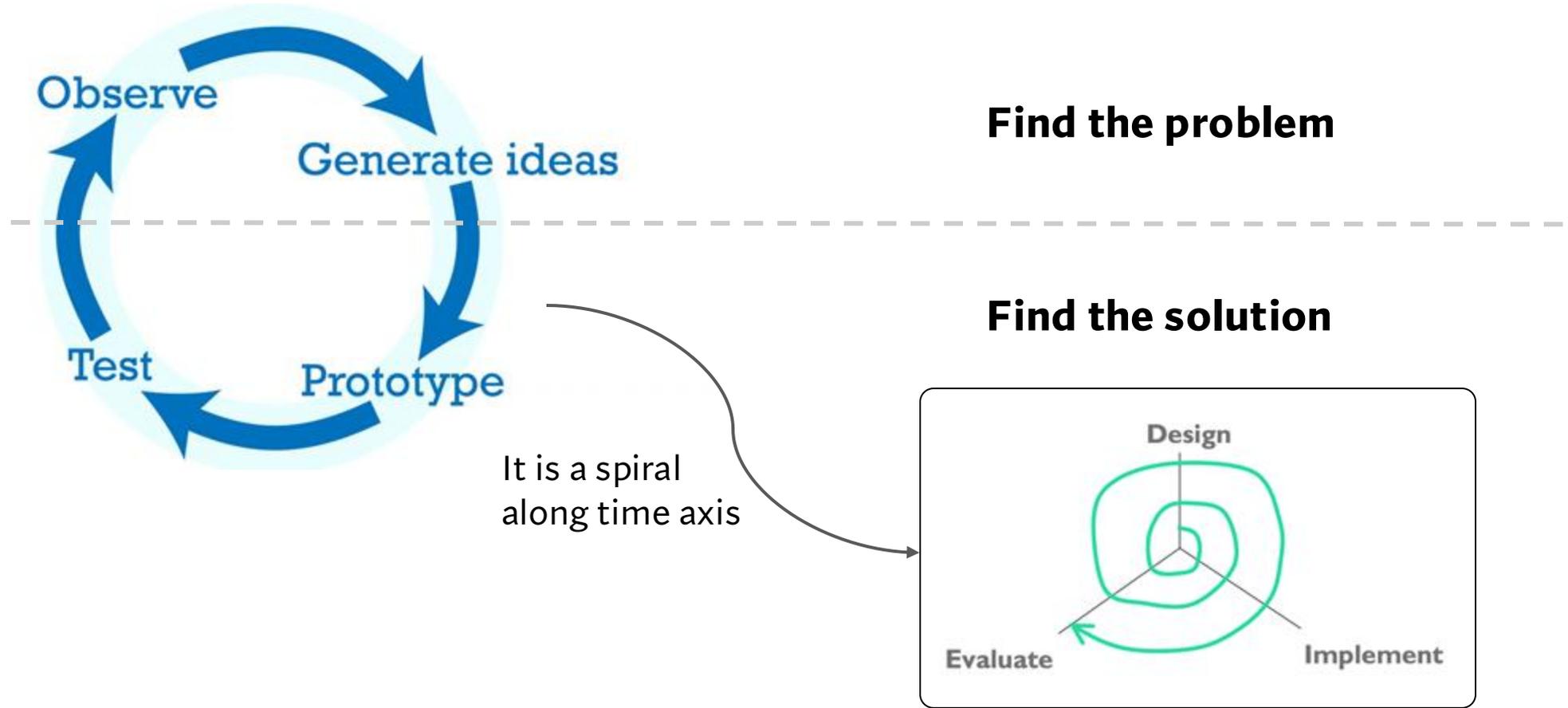
Four Steps:

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

But this is not once and for all

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

Iterative Design



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

Example Task: Make writing faster for humans

June 28, 2018

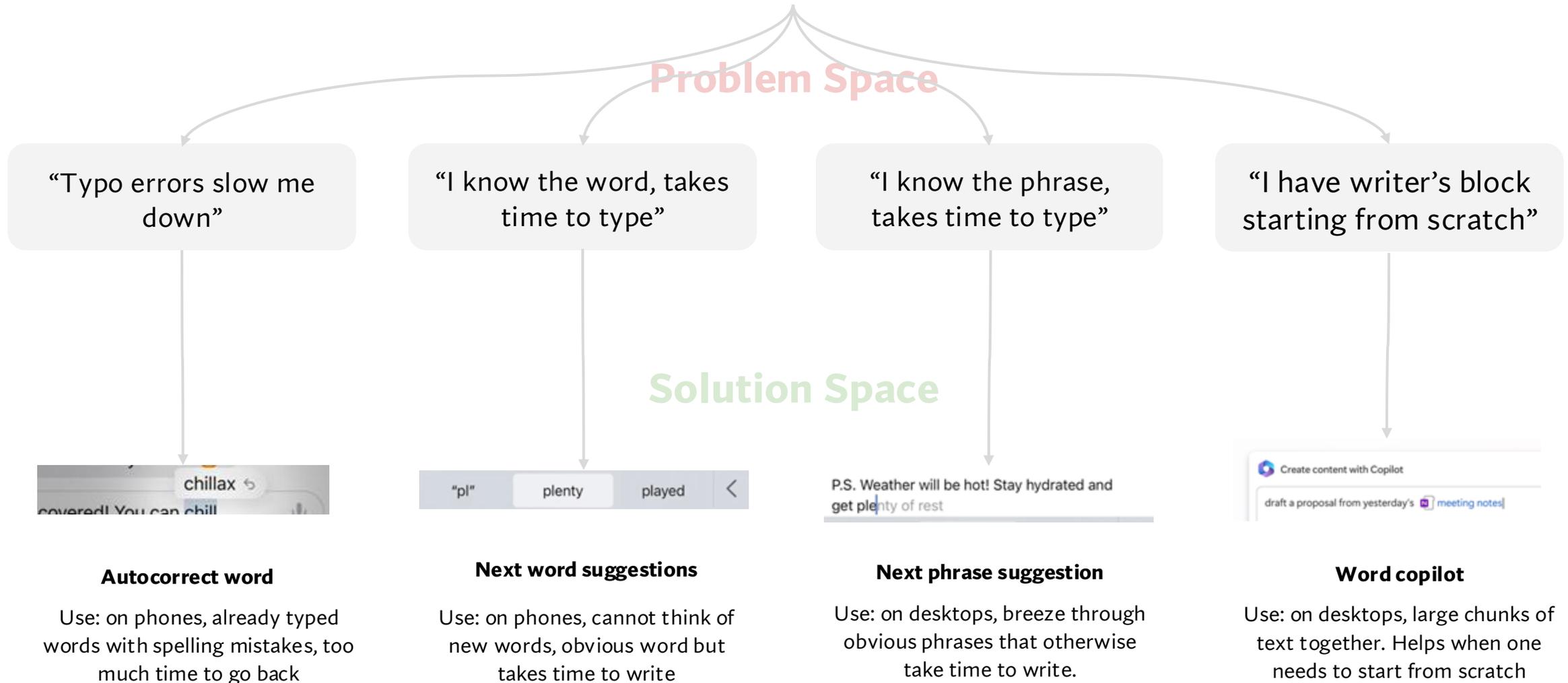
Justice Moore
IT Security and Solutions Manager
Alliope Systems, Research Triangle Park
456 Davis Drive
Durham, NC 27603

Dear Mr. Moore:

Thank you for your generous donation on behalf of Alliope Systems for a **gift certificate for an 18-hole round of golf at Weston Heights Country Club**. It will be considered a big-ticket item to be bid upon during the silent auction at Lakestone Montessori's 7th annual [Tee off for Technology Charity Golf Tournament](#). Your donation is a wonderful addition to our auction, and I predict that it will be a very popular item!

As you may know, with the recent budget cutbacks in education funding, this annual fundraiser is one of our most important events of the school year. Your donation will help us to reach this year's fundraising goal of \$10,000 to help support and increase technology resources at Lakestone Montessori. In addition to replacing aging computer equipment, we would like to purchase 10 Apple iPads to help support the use of technology in expanding and exploring lesson content.

Example Task: Make writing faster for humans



Develop Solution: Prototype

Wizard-of-Oz

Fake features so that the user thinks that the responses are computer-driven when they are actually human-controlled.

Challenge for NLP: AI errors are hard to simulate.

Mimic simple functionality

Ensemble multiple tools, LMs, simple models and expectations.

Challenge for NLP: cannot simulate SOTA model capabilities

LM Prototypes / Scaffold

Simulate users that may use the system

Use LMs to build prototypes

Develop Solution: Prototype Persona

Algorithmic persona: human roles that users assign to the algorithm to explain the algorithm's goals, behaviors, and characteristics.



Agent

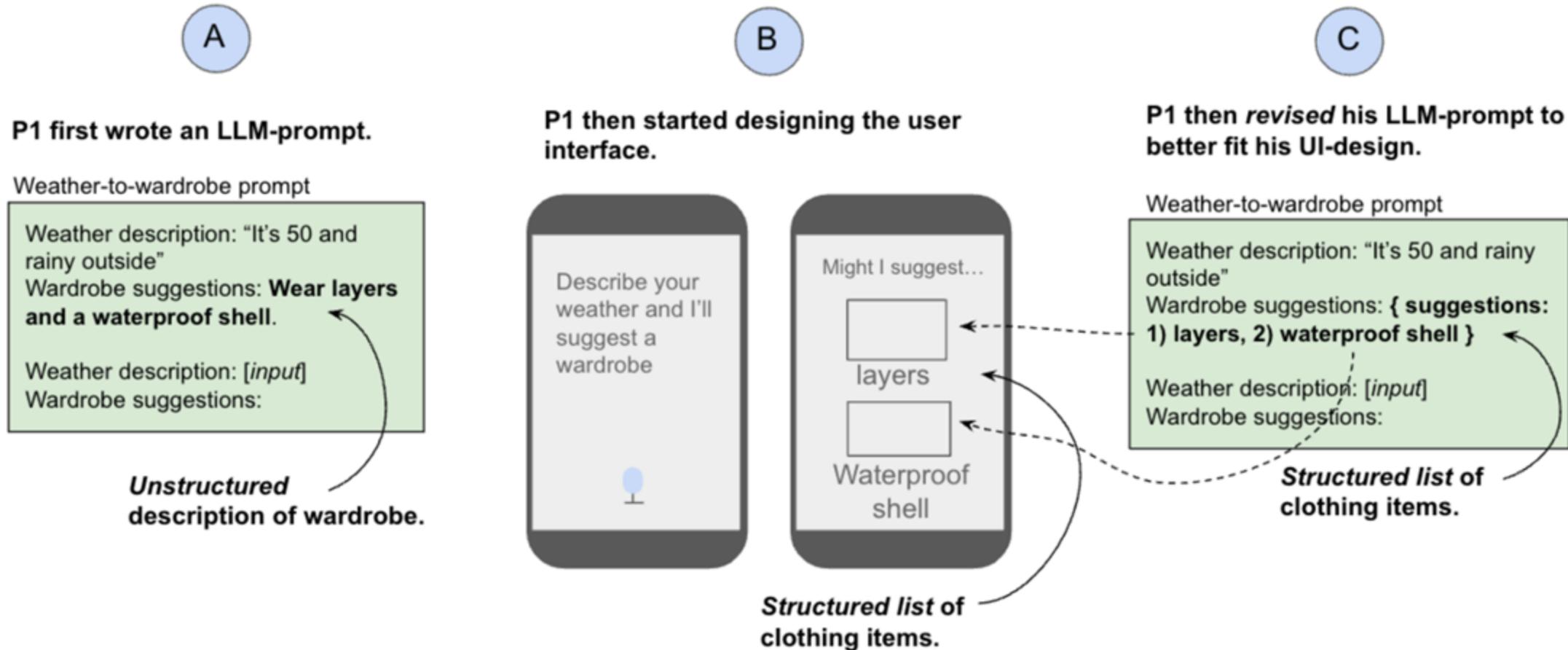


Gatekeeper



Drugdealer

Develop Solution: LM as a Prototyper / Scaffolding



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.