CS 329X: Human Centered LLMs
HCI & Design Thinking

Diyi Yang

Announcements

- Midway presentations
  - 3 mins pitch
    - One-page: project goal, method, data, any preliminary results, expected results
  - 2~3 mins QA
  - **Judges:** TAs + 3 external judges
  - Presentation scores will be averaged across TA + judge

Ryan Louie
LLM + Mental Health

Vishakh Padmakumar
LLMs + Creativity

Caleb Ziems
LLMs + Social Intelligence

Example Project Slide

Team Member 1, Team Member 2, …

Research Problem / Project Goal

Idea
- Bullet point
- Bullet point

Impact /vision
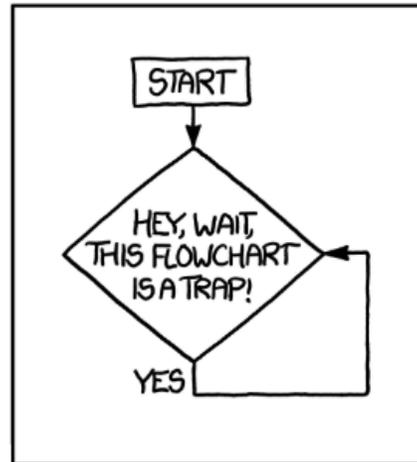
What impact your project will introduce

Methods / Data / Model

Idea
- Bullet point
- Bullet point

Idea
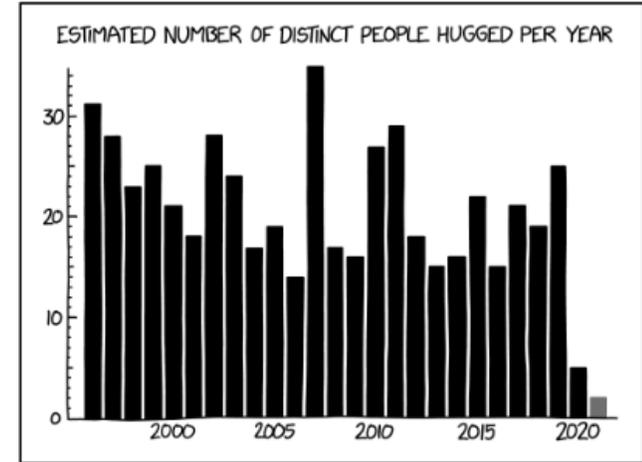- Bullet point
- Bullet point
- Bullet point

Caption: x

Preliminary Results / Analysis (if any)

Idea
- Bullet point
- Bullet point

Idea
- Bullet point
- Bullet point
- Bullet point

ESTIMATED NUMBER OF DISTINCT PEOPLE HUGGED PER YEAR

Caption: y

Next Steps / Challenges / Questions / Discussion

Idea
- Bullet point
- Bullet point

Example Project Slides (1-page, 3 mins)

- Keep these five blocks!

- But feel free to make any changes to the style (e.g., format, font, and visualization)!

- Make a copy of Page 3 and create one for your project!

- **Room assignments and slides to upload your projects will be released today after class!**

Outline

- **Design Thinking (40 mins)**

- **Case Study on Knoll by Dora Zhao (20 mins)**

- **Small-Group Discussion** (20 mins)

Learning Objective: understand design thinking and HCI methods

The Rise of "Human-AI Interaction"

Definition: how humans and AI systems interact

Humans: AI researchers, model developers, domain experts, end users...

AIs: LLMs, VLMs, translator, recommender system, autonomous driving system...

Interact:

Humans collaborate with AI,
Humans get assistance from AI,
Humans analyze AI,
AI helps human,
& many other forms

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

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	"<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?"	- 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.
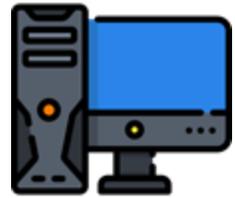
Design that Disappears

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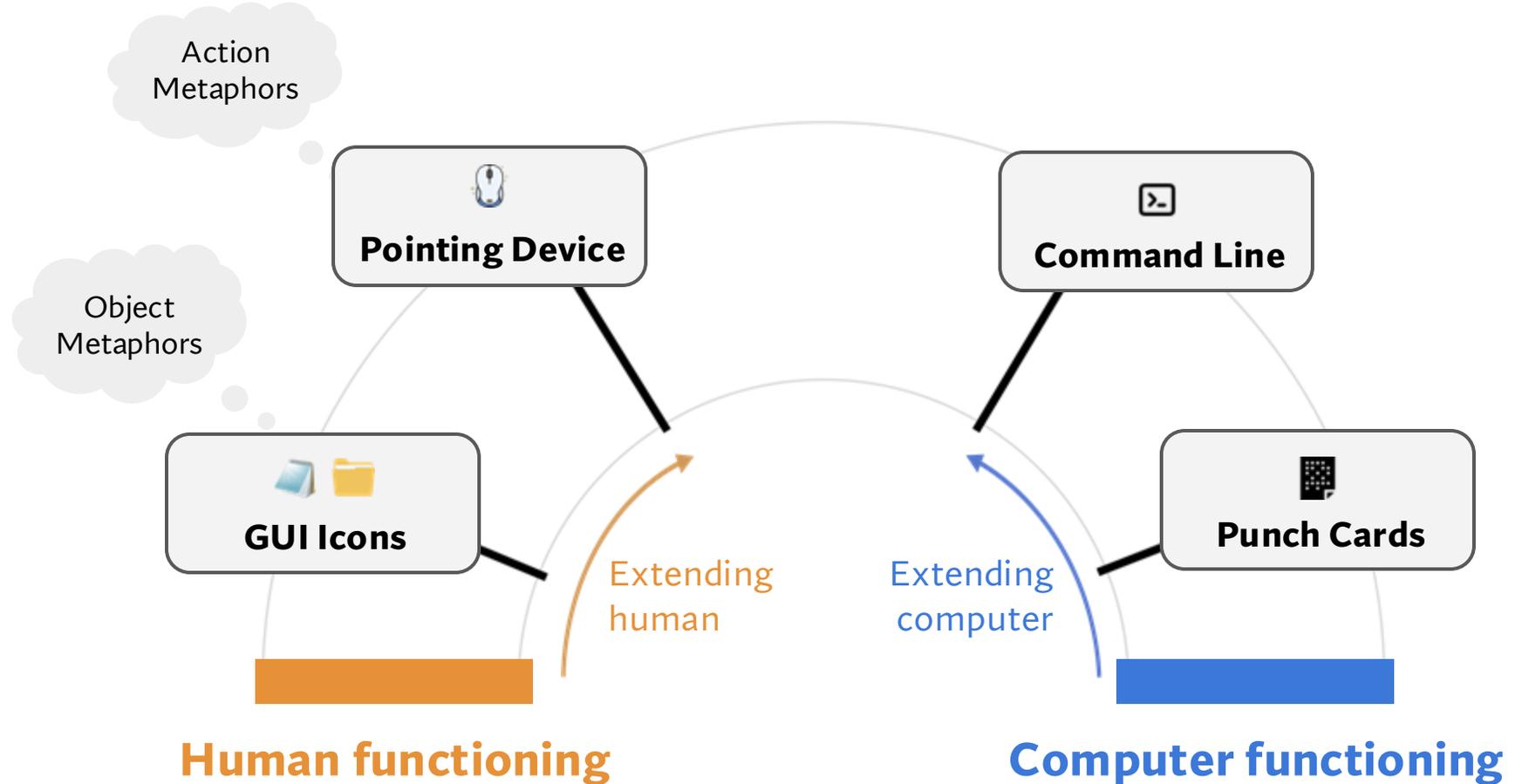
— Mark Weiser

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.

How do we achieve this?

By extending humans instead of extending technology, reducing "frictions of learning"

Personal Computing

Action Metaphors

Object Metaphors

Pointing Device

Command Line

GUI Icons

Punch Cards

Extending human
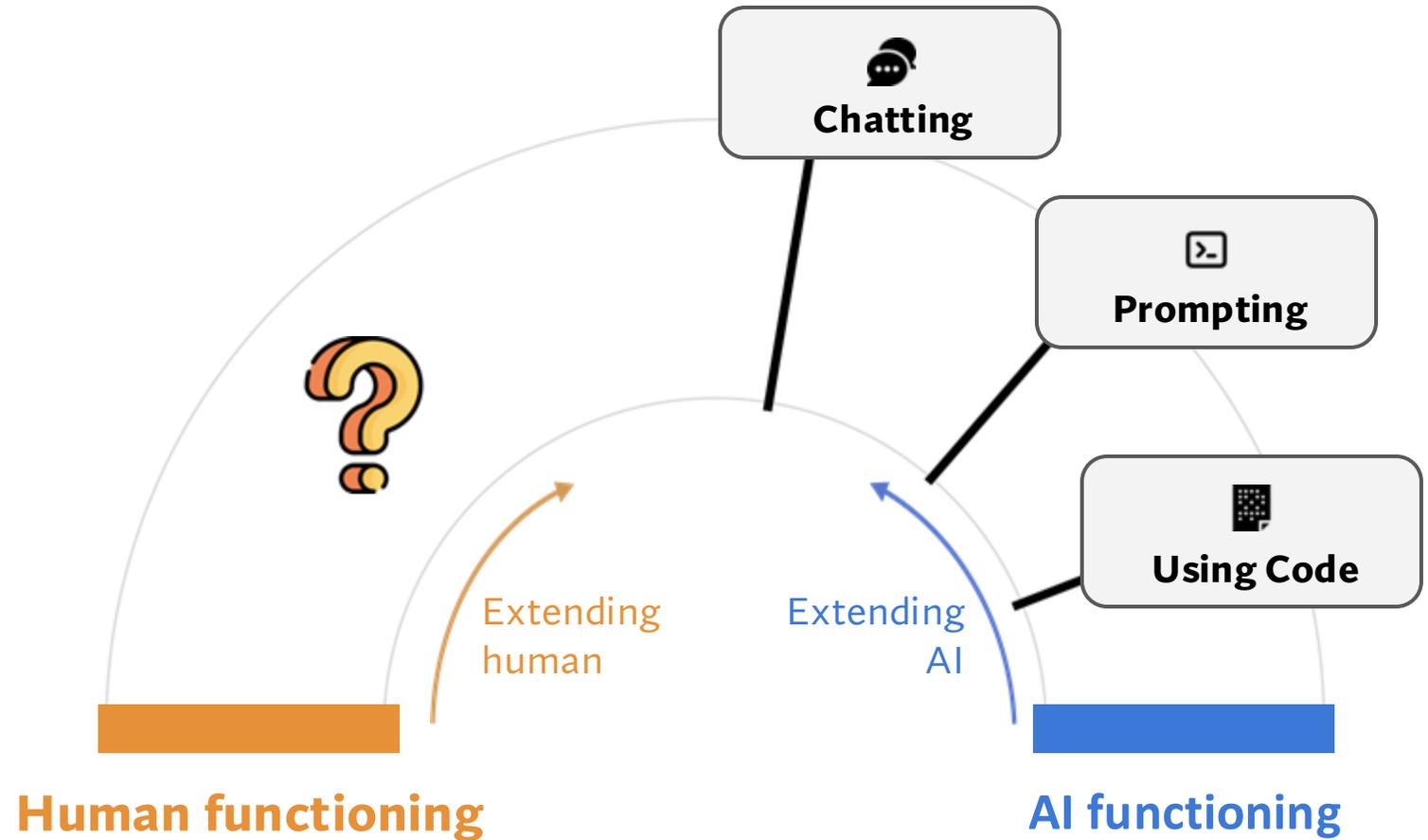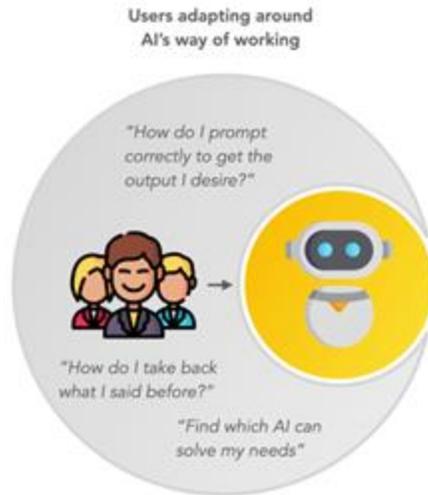
Extending computer

Human functioning

Computer functioning

How do we achieve this?

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

Artificial Intelligence

Chatting

Prompting

Using Code

Extending human

Extending AI

Human functioning
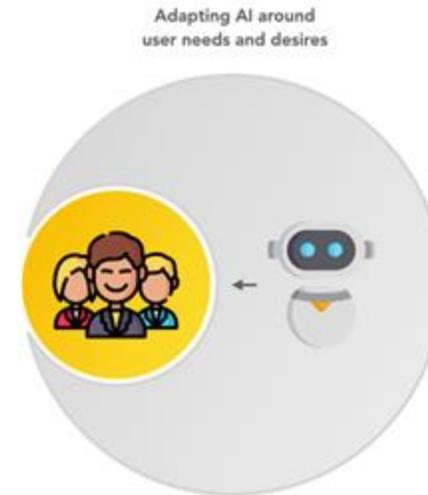
AI functioning
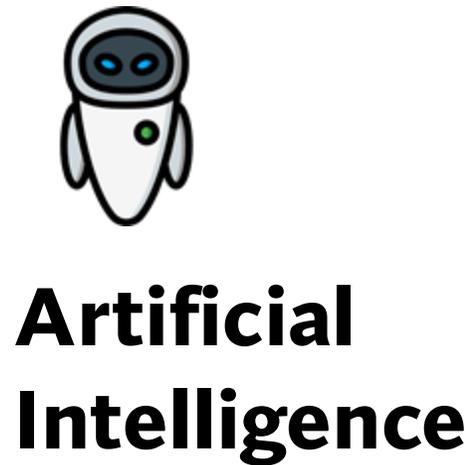
Why does it happen?

Technology-centric design
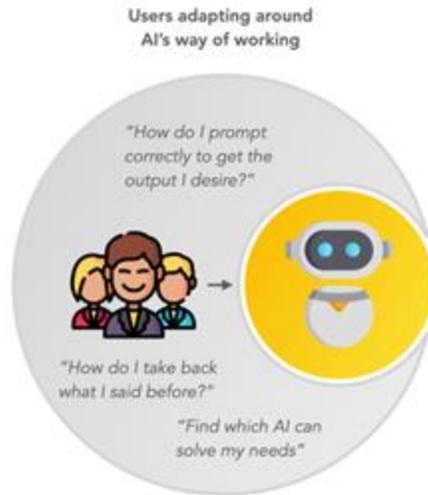
User-centered design

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.

Bly, Sara, and Elizabeth F. Churchill. "Design through matchmaking: technology in search of users." interactions 6.2 (1999): 23-31.

Why does it happen?

Technology-centric design

Users adapting around AI's way of working

"How do I prompt correctly to get the output I desire?"

"How do I take back what I said before?"

"Find which AI can solve my needs"

> "If you have a hammer, everything looks like a nail"
>
> **Solution in search of a problem**

User-centered design

Adapting AI around user needs and desires

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.

Bly, Sara, and Elizabeth F. Churchill. "Design through matchmaking: technology in search of users." interactions 6.2 (1999): 23-31.

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.

New models → new AI interactions

ChatGPT

Chatting

Pushing language as the primary interface for every task

Anthropomorphic Tendencies

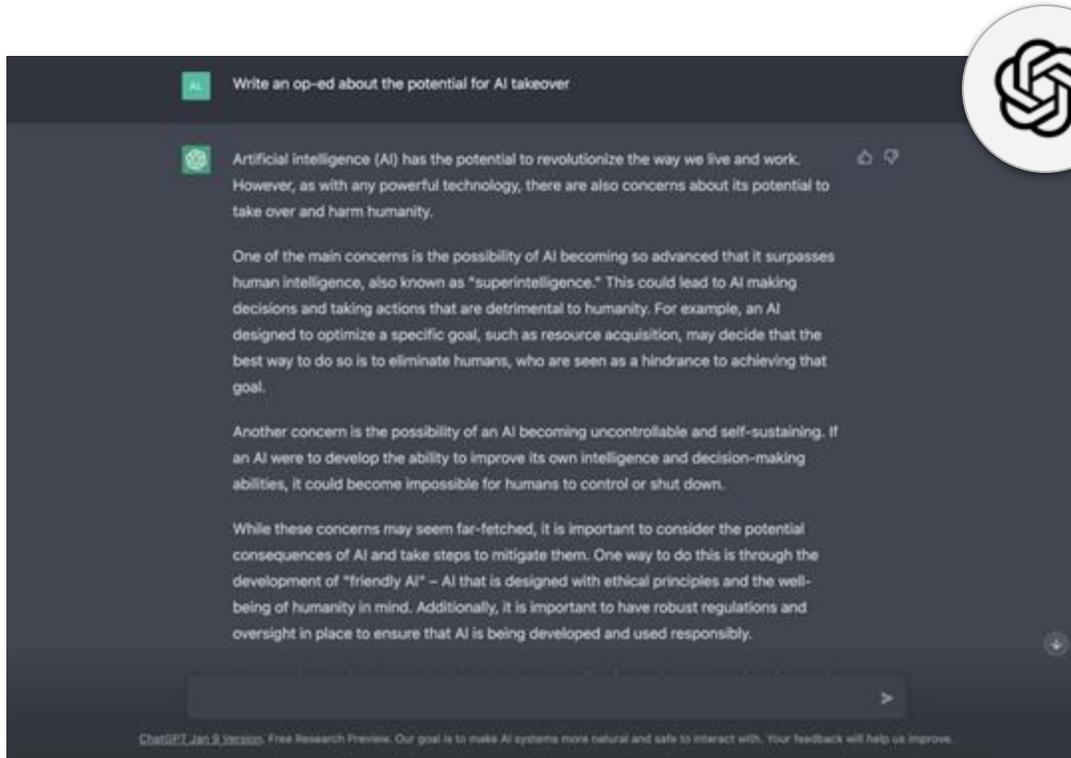The act of projecting human-like qualities or behavior onto non-human entities, in this case, AI

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
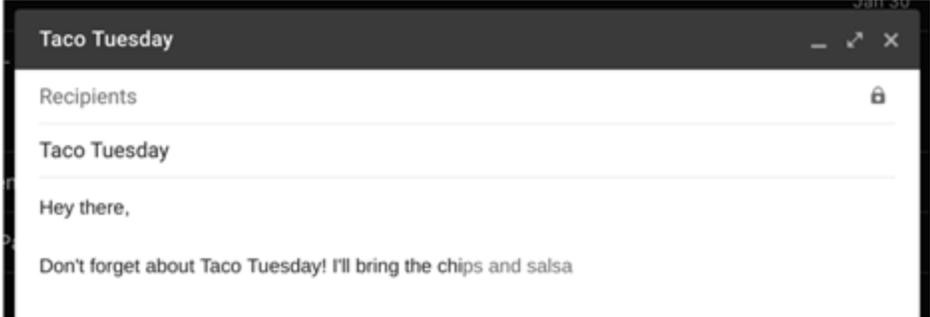
Email Autocomplete
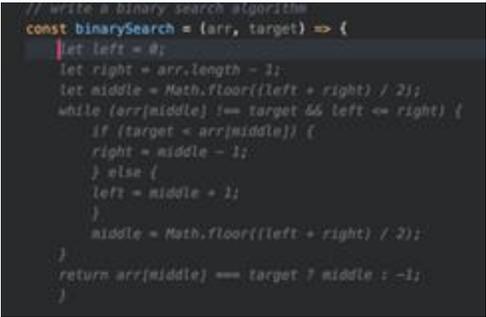
Code Completion

Recommendations
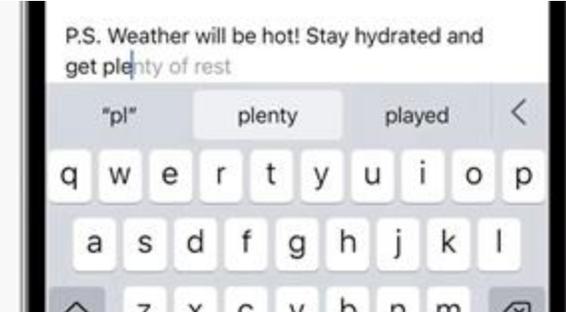
Word suggestions

Subtitles

What is design thinking?

Problem → **Solution 1**
             → **Solution 2**
             → **Solution 3**
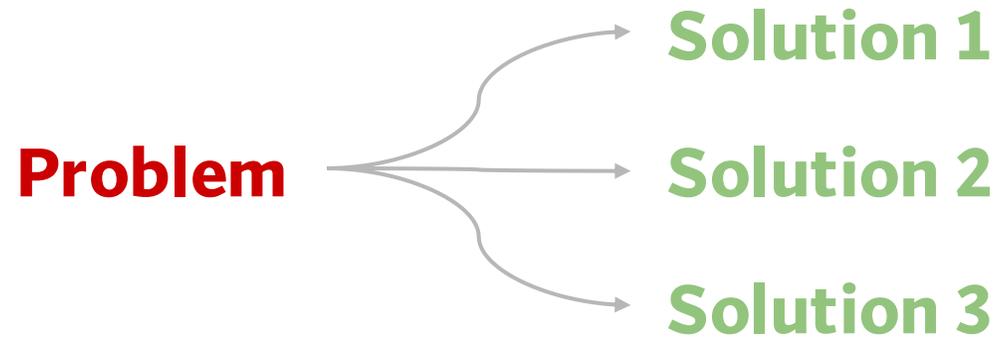
As engineers, we immediately jump to finding solutions for a problem that we come across.

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

What is design thinking?

Problem 1.1 ← **Why?**

Problem 1 ← **Why?**

Why?

Problem 2 ← **Why?**

Problem → Solution 1

Solution 2

Solution 3
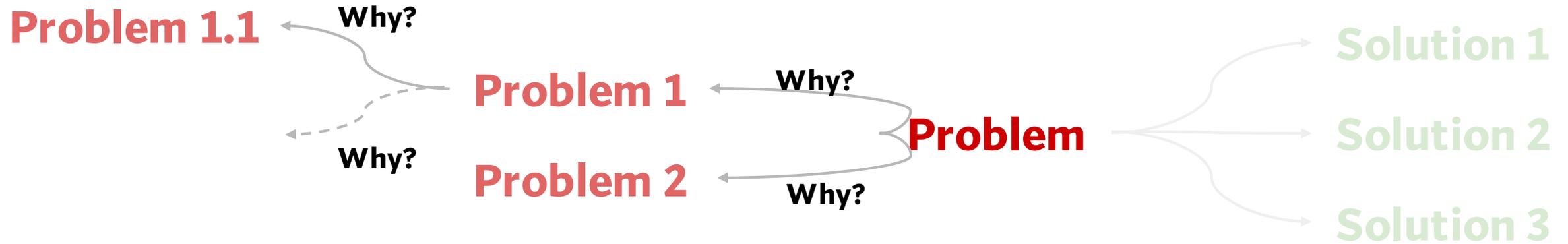
Ask "Five whys"

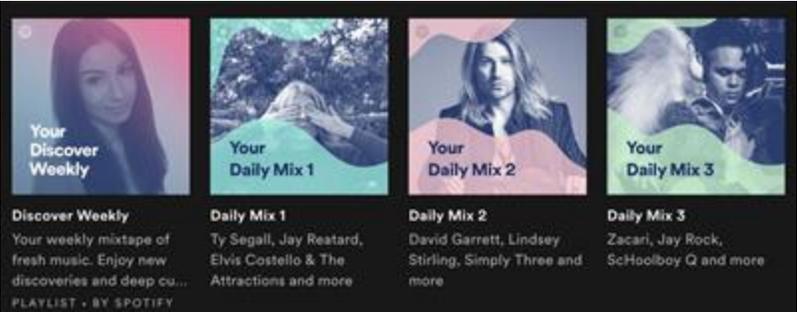It is not about the exact number, but this high number puts emphasis on getting to the root cause of the problem, instead of looking at the surface level.

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"

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?
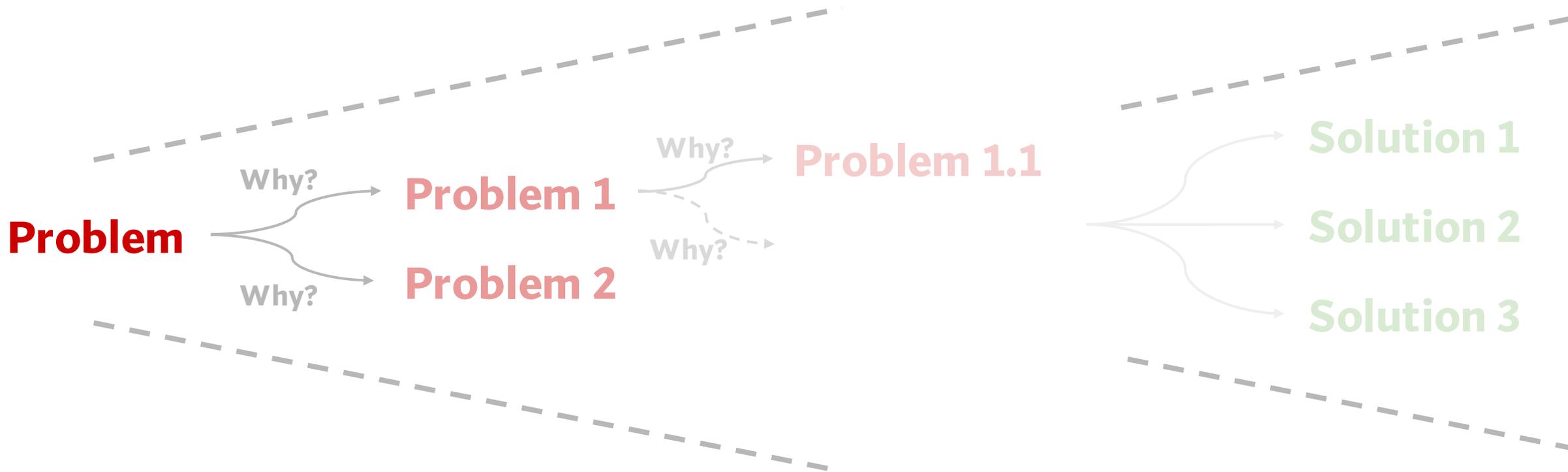
Problem → Why? → **Problem 1** → Why? → **Problem 1.1** → **Solution 1**

Why? → **Problem 2** Why? → **Solution 2**

Solution 3

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

How to incorporate design thinking?

The "Double Diamond" Method

Find the problem   Find the solution

Design the right thing   Design things right

Divergent   Why?   Convergent   Divergent   Convergent

Why?

General Problem   Specific Problem   Specific Solution
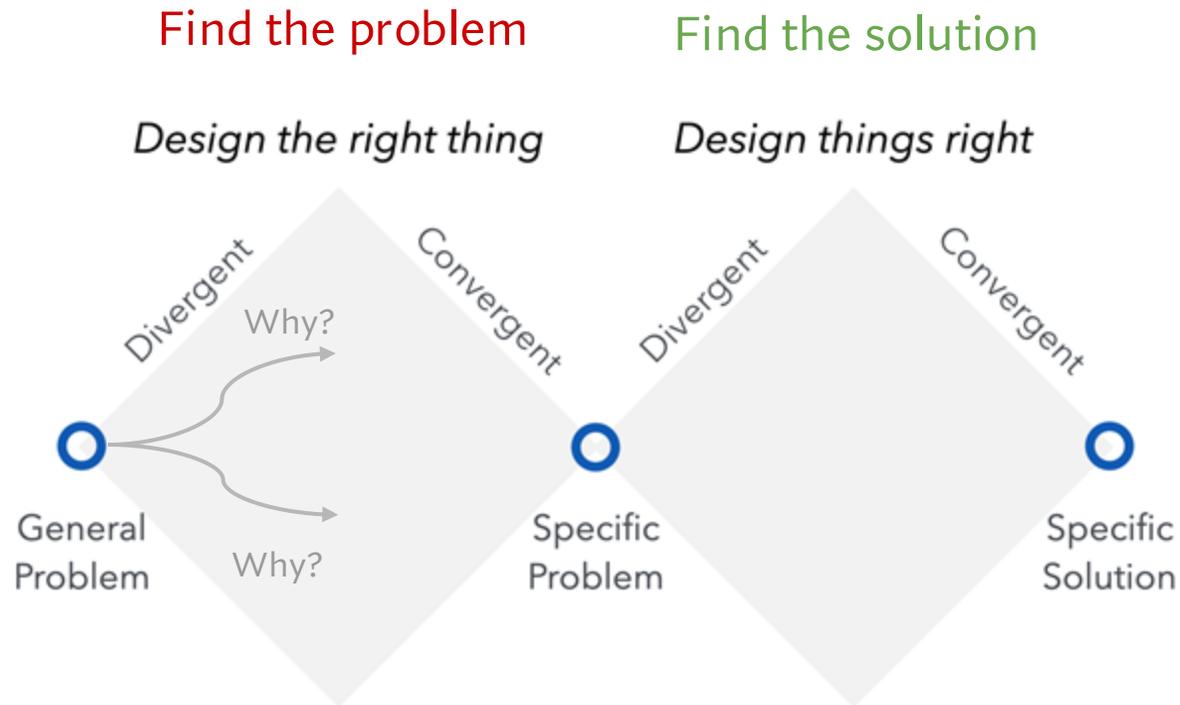
First Diamond

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

Find the problem

Find the solution

Divergent
Convergent
Divergent
Convergent

Discover
Research

Define
Synthesis

Develop
Ideation

Deliver
Implement.

General Problem

Specific Problem

Specific Solution
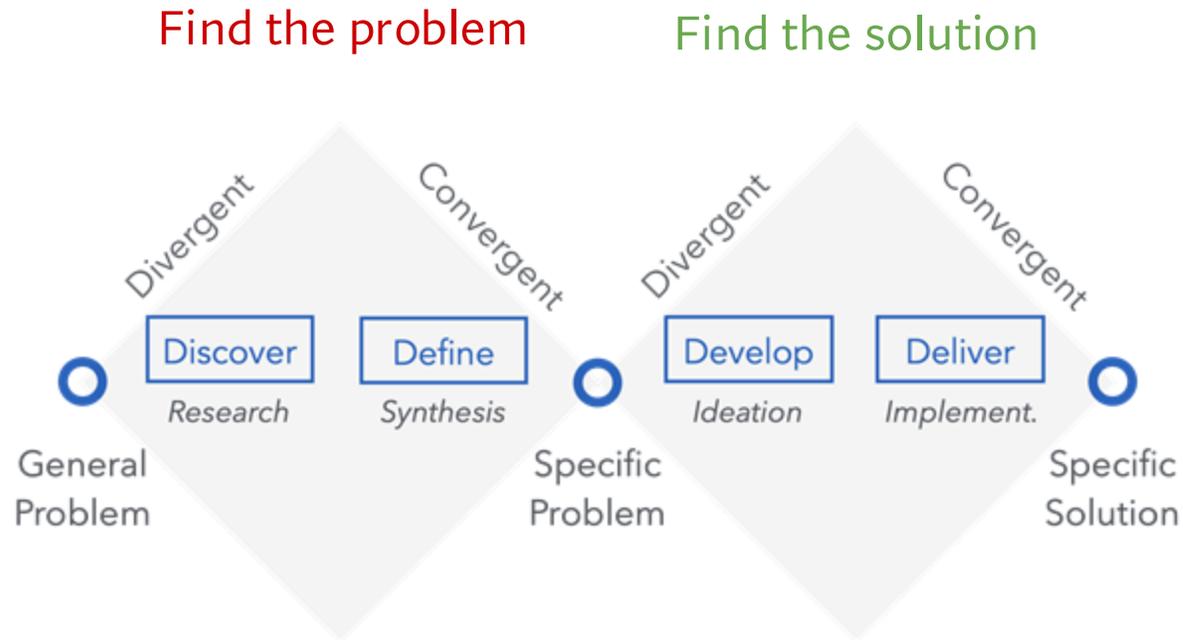
Four Steps:

- 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.

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

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

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

Discover: Interview

- 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

Codes

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.

- 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

Nielsen, J., & Landauer, T. K. (1993). A mathematical model of the finding of usability problems. INTERACT'93 and CHI'93.

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.

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

Develop Solution

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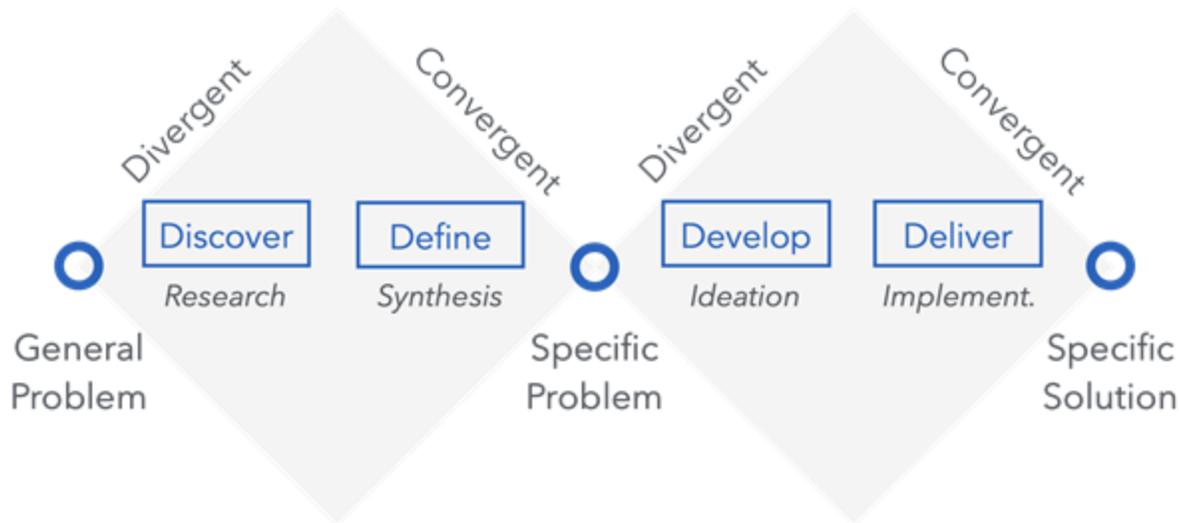
Storytelling

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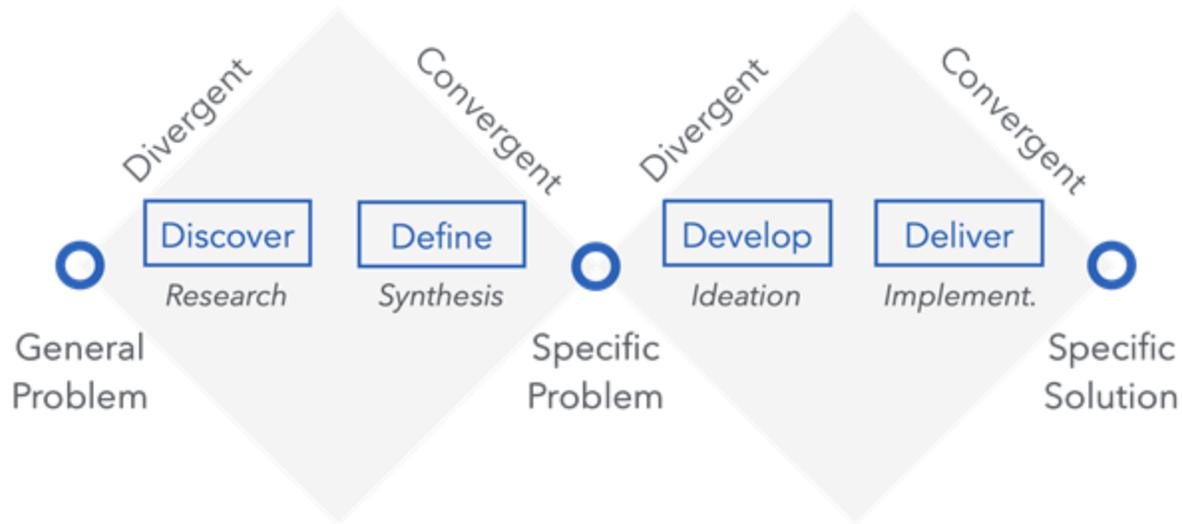
Rapid Prototyping

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

40

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."

Holtzblatt, Karen, and Hugh Beyer Contextual Design : Defining Customer-Centered Systems, Elsevier Science & Technology, 2016.

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*

Holtzblatt, Karen, and Hugh Beyer Contextual Design : Defining Customer-Centered Systems, Elsevier Science & Technology, 2016.

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
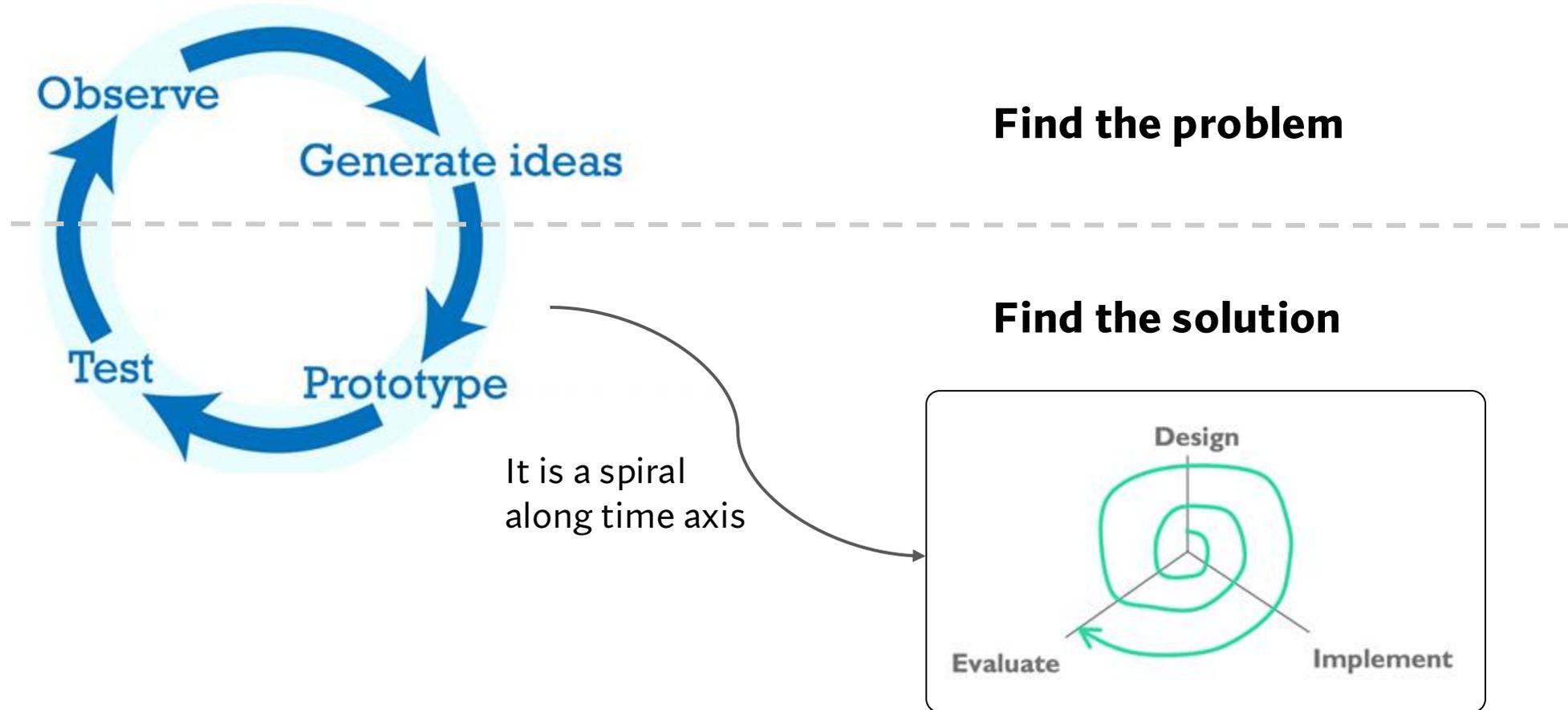
The "Double Diamond" Method
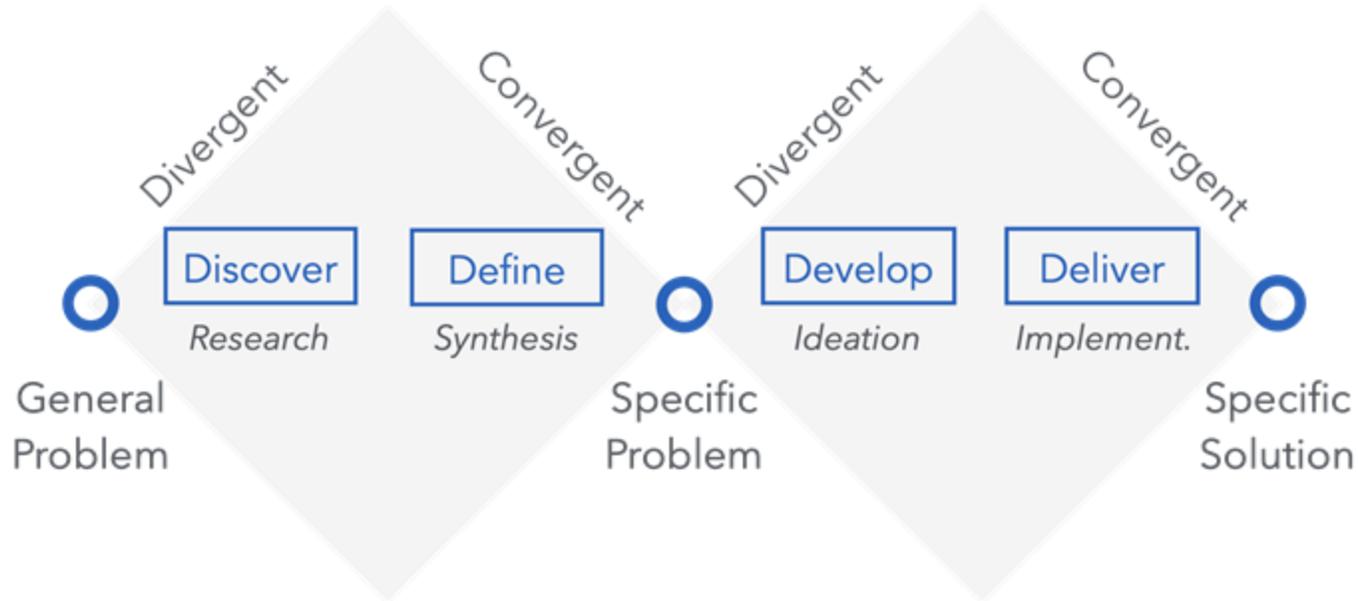
Four Steps:

- 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

Observe → Generate ideas → Prototype → Test (cycle)

Find the problem

Find the solution

It is a spiral along time axis

Design, Implement, Evaluate (spiral)

Norman, Don. The design of everyday things: Revised and expanded edition. Basic books, 2013. DreamEndState.com
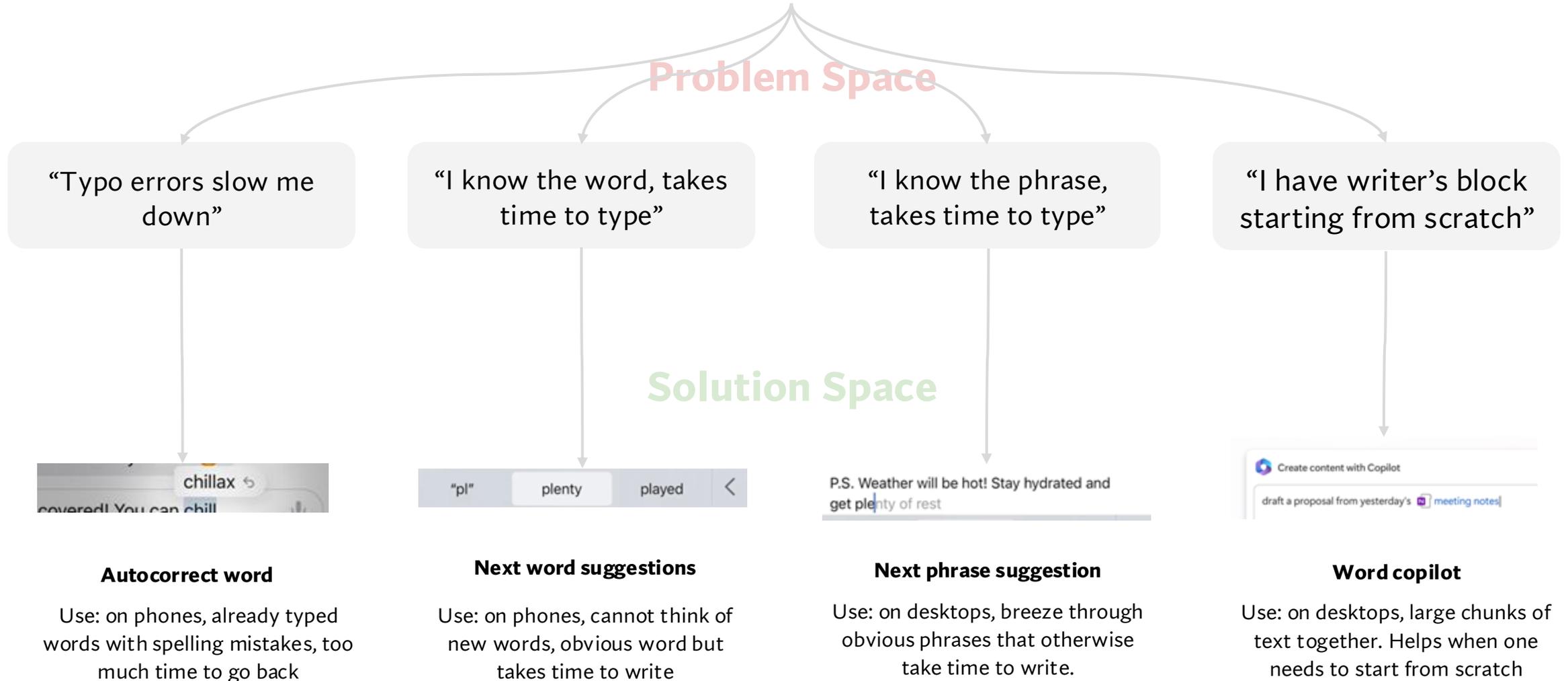
Example Task: Make writing faster for humans

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

Example Task: Make writing faster for humans

Problem Space

Solution Space

"Typo errors slow me down"

"I know the word, takes time to type"

"I know the phrase, takes time to type"

"I have writer's block starting from scratch"

Autocorrect word

Use: on phones, already typed words with spelling mistakes, too much time to go back

Next word suggestions

Use: on phones, cannot think of new words, obvious word but takes time to write

Next phrase suggestion

Use: on desktops, breeze through obvious phrases that otherwise take time to write.

Word copilot

Use: on desktops, large chunks of text together. Helps when one needs to start from scratch

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

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.