ADVANCED DESIGN OF INTERACTIVE SYSTEMS

12 FEBRUARY 2026

Janin Koch
Permanent Researcher
Inria Lille
janin.koch@inria.fr

LECTURES & TOPICS

1

Introduction

Introduction of Goals and Topic

Exercise: DOIS Quiz & Choose groups

Homework: Story Interviews

2

Understanding Humans

Understand underlying cognitive effects:

Grounding, Theory of Mind, Mental Models

Exercise: Concept Development

Homework: Initial Design

3

Agency in HAI

Explainability, Ironies of Automation

How to design for Error, Feedback and Control?

Exercise: Build Prototype

Homework: Storyboard

4

Iterating on Design

Socio-Technical Systems

Exercise: Prototype + Generative Walkthrough

Homework: Revise Concept

5

Evaluating interactive systems

Overview of evaluation methods

Exercise: Revise Video Prototype

Homework: Evaluation protocol

6

Reflection on HAI

Social, legal, sustainability impact of AI

Discuss the role of designers for HAI interaction

Exercise: Poster, Presentation

HOMEWORK

PREPARE REDESIGN ___

- Reassess your **users**
- Reassess your **ideas**
- Reassess your **design**
- Reassess your **system**

1) Gen. Walkthrough results:

- Summarize in an overview 1) interaction, 2) identified problem, 3) feedback/suggestion
- Make a list and provide improvements for each

2) Prepare Redesign: Reflect on your designs and consider context!

Decide on **at least** 3 new interaction snippets informed by the walkthrough to integrate!

Adapt your future scenario accordingly:

- Is it an interactive tool?
- How does your concept empower the user?
- What are the advantages and disadvantages of using your tool?
- Maybe consider potential other users - how would it change the interaction?

HOMEWORK
REDESIGN ___

3) Technological Feasibility:

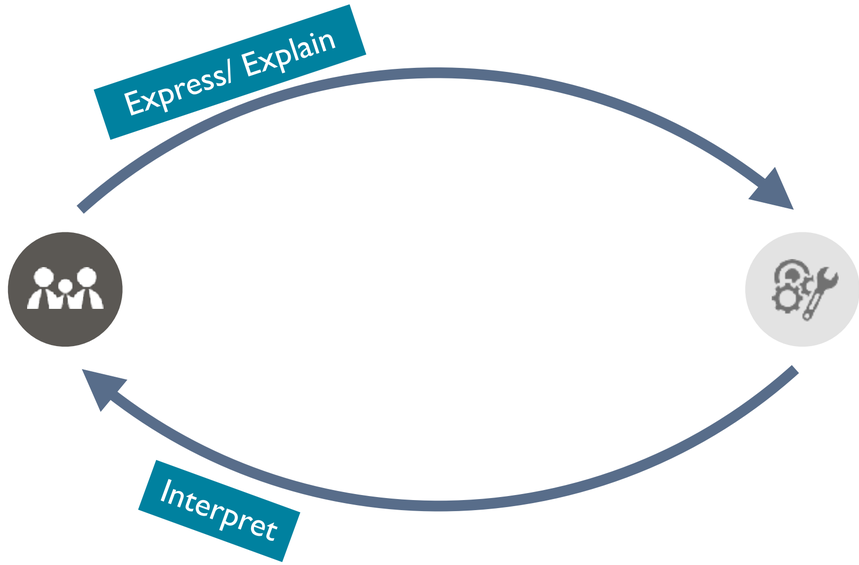
- Look at your baseline system/AI—would it support your use case?
- Go through your user interaction:
  - What do you need to translate to the system/user?
  - What other information would you need to collect or present?
- Make a list and suggest technical solutions.

4) Update your prototype based on these insights:

- Paper prototype
- Interactive prototype (optional)

EXERCISE

REVISED USER STORYBOARD ___



Revise your Storyboard:

- Revise the future scenario to incorporate the revised concept
  - Did you make the interaction simpler? more powerful?
- Add at least three new interaction snippets
  - Do they show off the new concept?
- How does it generalize to other users?

EXERCISE

REVISED VIDEO PROTOTYPE ___

- Shoot a new 3-5 minute video
- Reveal how your system works by showing how your target user interacts with it
- If you make a mistake, delete and reshoot

EVALUATION METHODS

EVALUATION

WHAT IS AN EVALUATION?

“Evaluation is a process that critically examines a program. It involves collecting and analyzing information about a program’s activities, characteristics, and outcomes. Its purpose is to make judgments about a program, to improve its effectiveness, and/or to inform programming decisions”

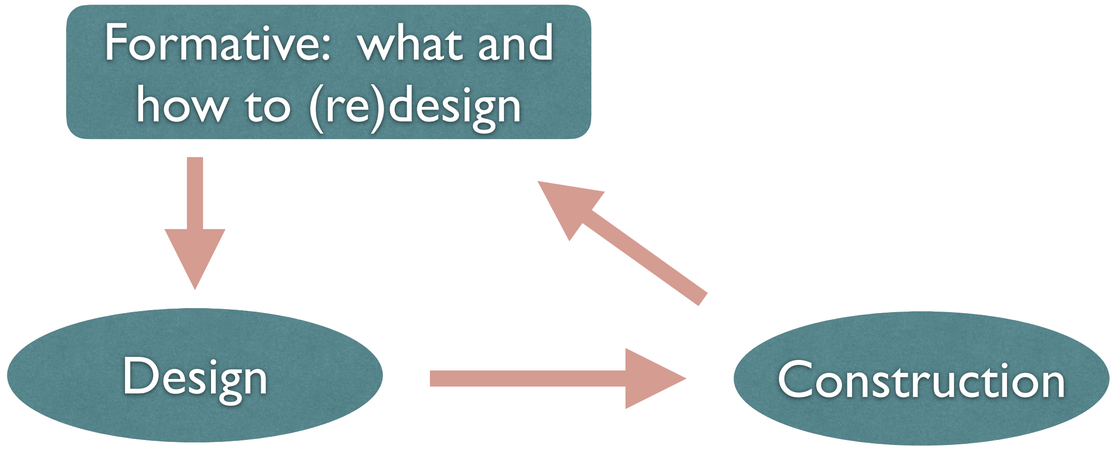
- Improves system design and implementation
- Demonstrate the system‘s impact
- Often reveal something new about people
- But how can we evaluate?

Patton, M.Q. (1987). Qualitative Research Evaluation Methods. Thousand Oaks

EVALUATION

WHAT EVALUATION?

FORMATIVE VS SUMMATIVE EVALUATION



Scriven, M. (2012). Conceptual revolutions in evaluation. Evaluation roots, 167-179.

WHAT EVALUATION?

FORMATIVE EVALUATION

Formative („How you design it?“)

- (Paper) prototyping, sketching
- Interviews, surveys, diaries, case studies
- Focus groups, user role playing, personas
- Generative Walkthroughs
- Participative Workshop

STUDIES

EXPLORATORY STUDIES

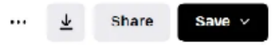
Purpose: To generate insights and hypotheses about user behavior or system design.

Methods:

- Focus Groups: Small groups of users discuss their experiences and opinions.
- Interviews: One-on-one conversations to gather in-depth qualitative data.
- Card Sorting: Users organize topics into categories to inform information architecture.

SUPPORTING

DESIGNERS TO USE GENAI IN PRACTICE



"a sunlit indoor lounge area with a pool with clear water and another pool with translucent pastel pink water, next to a big window, digital art"

Haris × DALL·E
Human & AI

Table 1. An example of the input data, the corresponding dataset image, and the image generated by our best ChatPainter model.

Input	Dataset image	Generated image
Caption: adult woman with yellow surfboard standing in water.		
Q: is the woman standing on the board? A: no she is beside it.		
Q: how much of her is in the water? A: up to her midsection.		
Q: what color is the board? A: yellow.		
Q: is she wearing sunglasses? A: no.		
Q: what about a wetsuit? A: no she has on a bikini top.		
Q: what color is the top? A: orange and white.		
Q: can you see any other surfers? A: no.		
Q: is it sunny? A: the sky isn't visible but it appears to be a nice day.		
Q: can you see any palm trees? A: no.		
Q: what about mountains? A: no.		

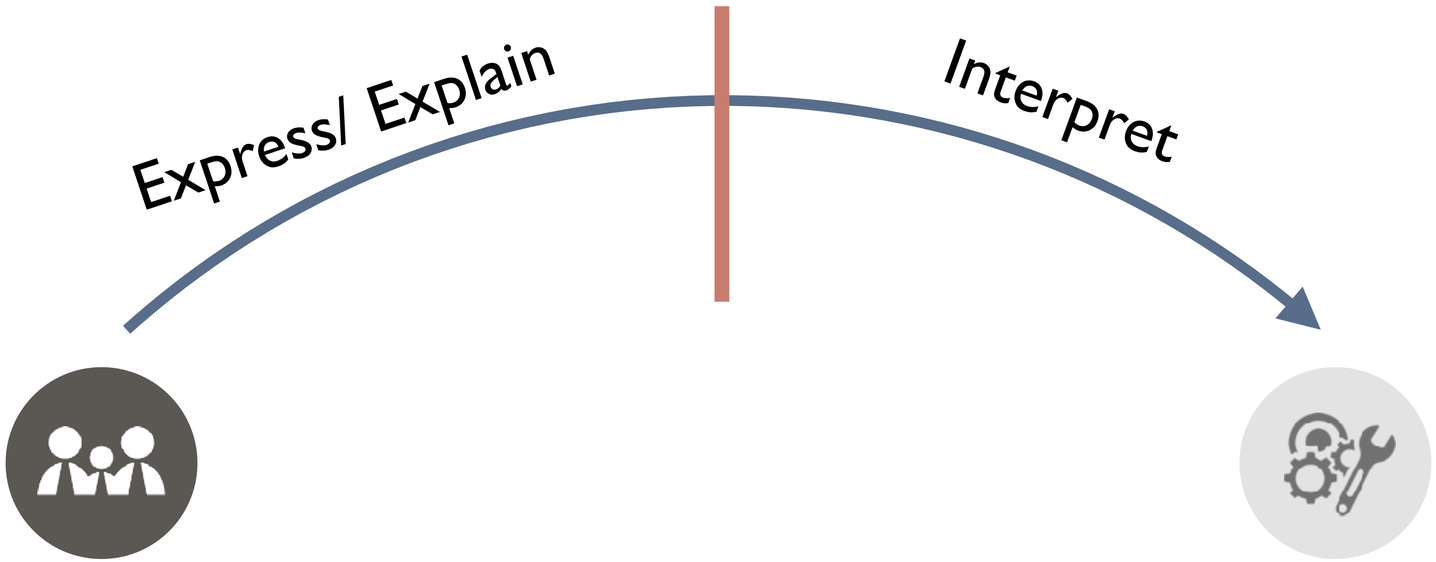
Is text prompting the best interaction in creative practice?

How can we use GenAI more meaningfully in design process?

Shikhar Sharma, Dendi Suhubdy, Vincent Michalski, Samira Ebrahimi Kahou, and Yoshua Bengio. Chatpainter: Improving text to image generation using dialogue, 2018.
https://openai.com/index/dall-e-2/

SUPPORTING

HUMAN EXPRESSION IN HAI INTERACTION



HAI INTERACTION IN DESIGN

UNDERSTANDING HUMAN NEEDS

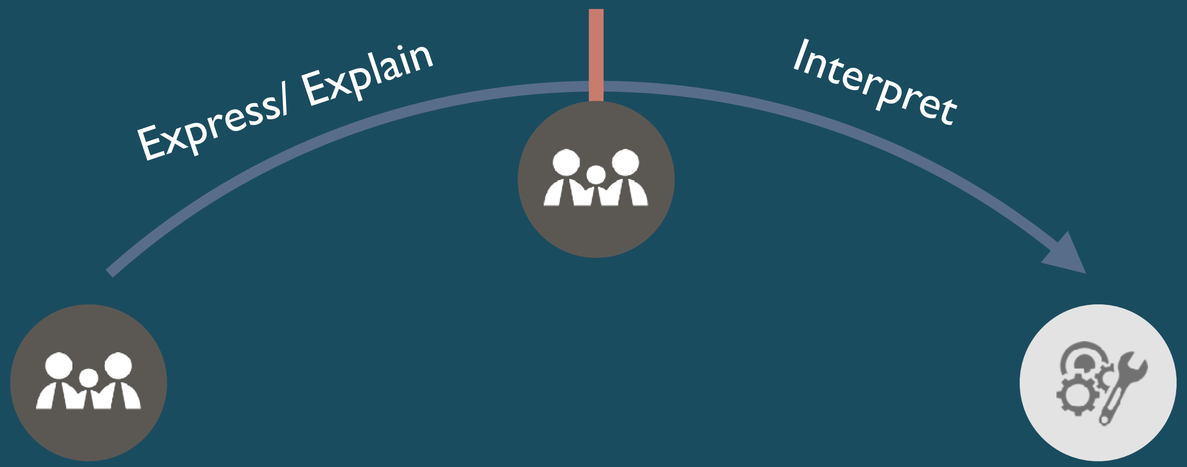
Study with 8 designers

Moodboard task (baseline condition):

- Participants create a moodboard that expresses a visual concept for a café interior in the style of their choice. They could use any image search engine or moodboard tools they prefer.

Generative AI task (experimental condition):

- Participants generated images of Café exterior with their preferred GenAI tool. They suggested potential GenAI prompt to the researcher, who entered it into the chosen tool. They then either used the image for their moodboard, or asked the researcher to adjust the prompt.



DESIGN IMPLICATIONS TO SUPPORT HAI INTERACTION

DI 1: Systems should support search or prompt input using different levels of image abstraction and semantics.

DI 2: Systems should help users to translate their abstract intentions to richer prompts.

DI 3: Systems should help users to identify the impact of prompts.

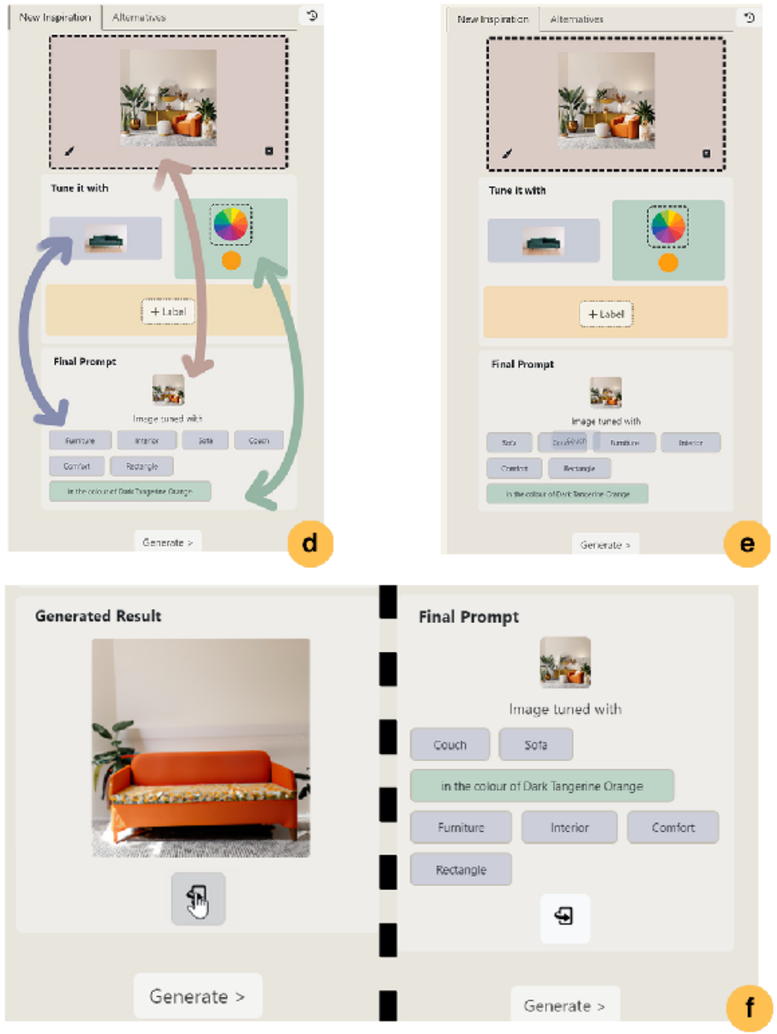
DI 4: Systems should allow users to control and manipulate images engagingly.

DESIGNING HAI INTERACTION

- RQ1: Does using multimodal input to create GenAI prompts allow designers to explore and express their intents better?

- RQ2: Does revealing system interpretation of user prompts help users to produce results that are more aligned with their expectations?

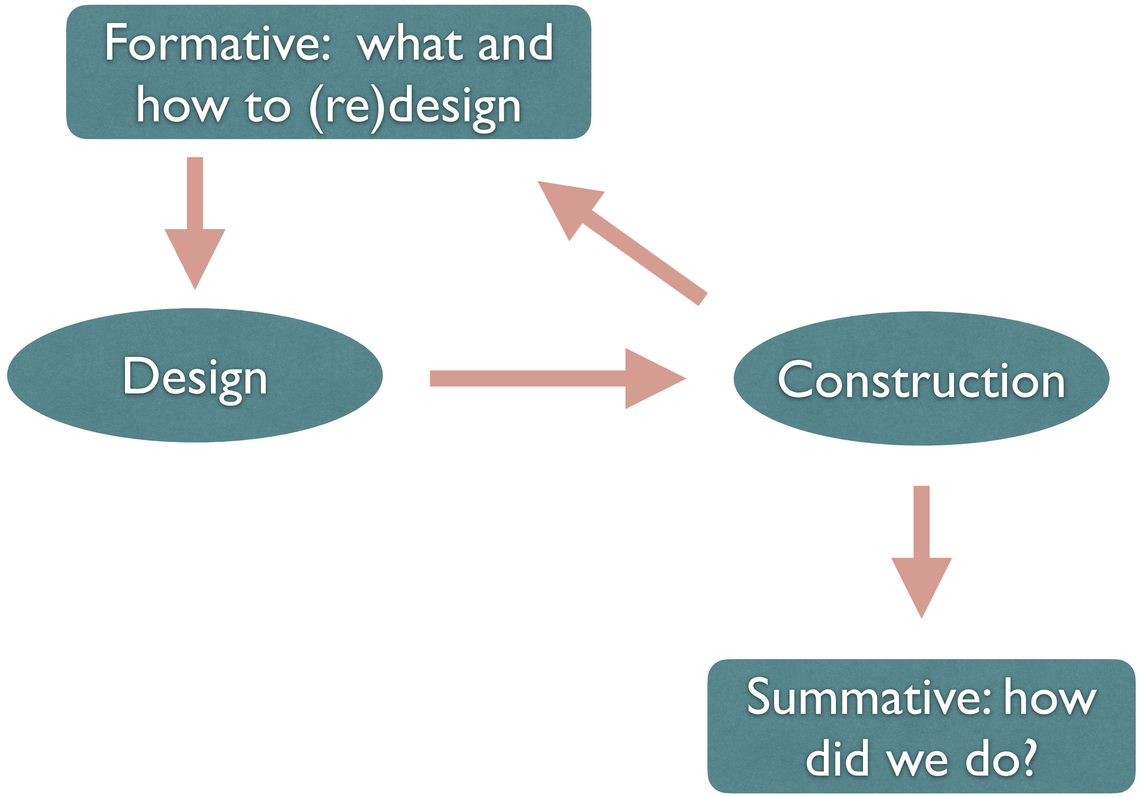
- RQ3: Does interactive and controllable GenAI input let users perceive the system as more transparent and useful for design practice?



EVALUATION

WHAT EVALUATION?

FORMATIVE VS SUMMATIVE EVALUATION



Scriven, M. (2012). Conceptual revolutions in evaluation. Evaluation roots, 167-179.

EVALUATION

WHAT EVALUATION?

SUMMATIVE EVALUATION

Summative („How good is it?“)

- Ratings, questionnaires, scales
- Efficiency measures
- Physiological measures
- Experiements
- Usability testing
- Field studies

STUDIES

USABILITY TESTING

Purpose: To evaluate how easy and efficient a system is to use.

Methods:

- Think-Aloud Protocol: Participants verbalize their thoughts while performing tasks.
- Task Analysis: Users complete specific tasks while researchers observe and measure success rates, errors, and time taken.
- Heuristic Evaluation: Experts evaluate the system against usability principles (e.g., Nielsen's heuristics).

STUDIES

CROWDSOURCED STUDIES

Purpose: To gather data from a large, diverse population quickly and cost-effectively.

Methods:

- Online Surveys: Distributed to a wide audience to collect quantitative data.
- Crowdsourcing Platforms: Platforms like Amazon Mechanical Turk are used to recruit participants for remote experiments.

STUDIES

COMPARATIVE STUDIES

Purpose: To compare different systems, designs, or interaction techniques.

Methods:

- Benchmarking: Comparing a system's performance against established standards or competitors.
- Feature Comparison: Evaluating specific features across systems to identify strengths and weaknesses.
- Comparative Structured Observation: Systematically comparing multiple conditions/systems in a semi-experimental setup using ecoloigical valid tasks.

STUDIES

SIMULATION STUDIES

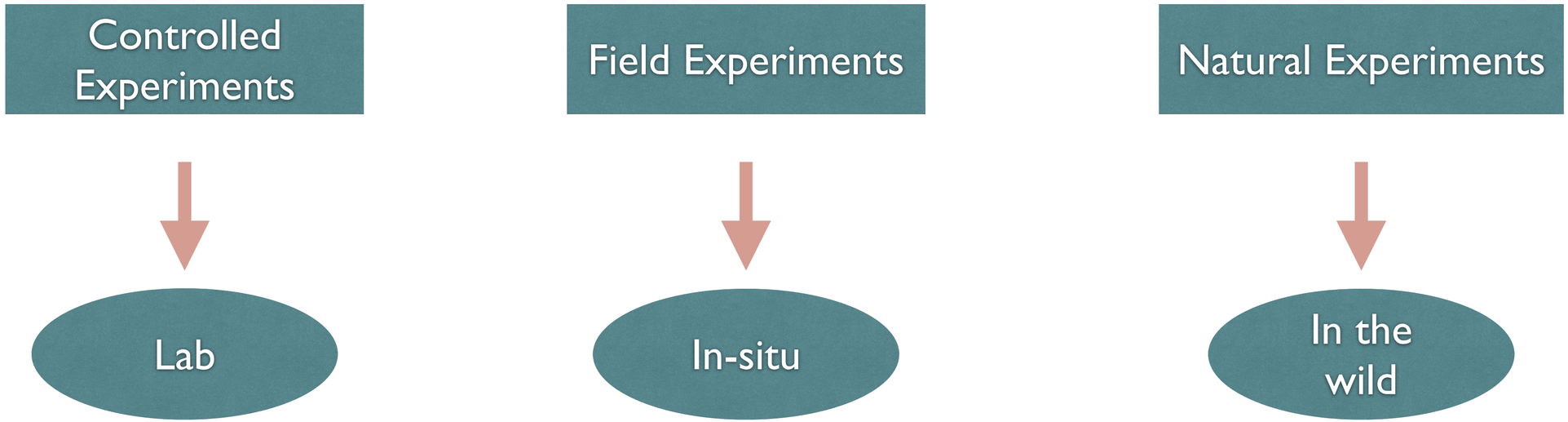
Purpose: To test systems or interactions in a controlled, simulated environment.

Methods:

- Wizard of Oz: A human simulates system responses to test user interactions with a prototype.
- Virtual Reality (VR) Simulations: Users interact with a system in a virtual environment to study behavior.

EVALUATION

EXPERIEMENT CONTEXT___



en.wikipedia.org/wiki/Experiment

STUDIES

CONTROLLED EXPERIMENTS

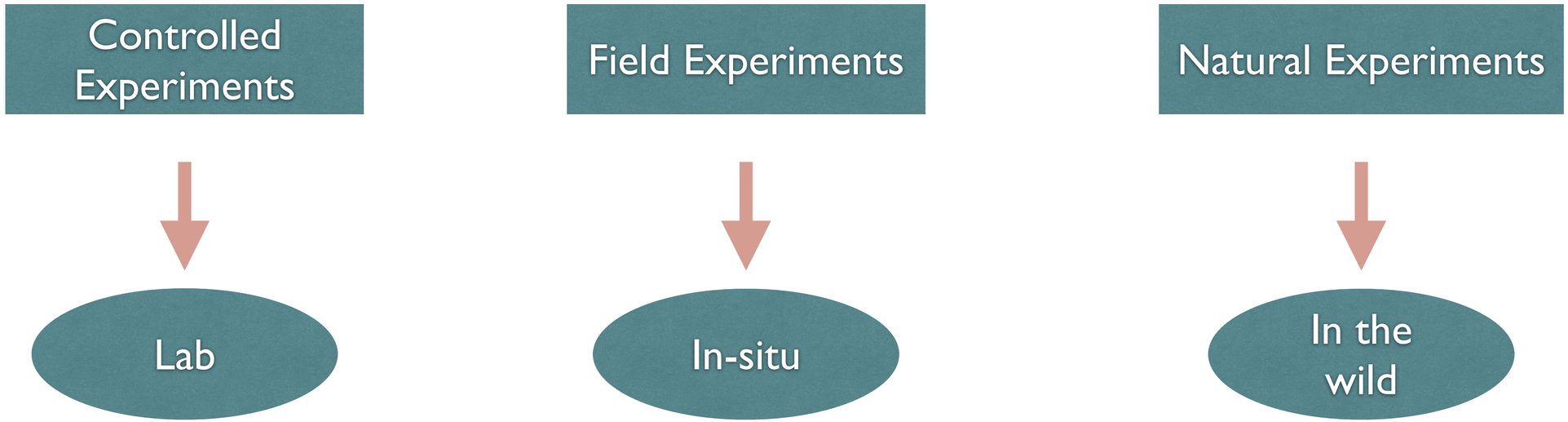
Purpose: To test specific hypotheses by manipulating one or more independent variables and measuring their effect on dependent variables. Determining the precise cause for our observation helps us to make any predictions about the world.

Examples:

- A/B Testing: Two versions of a system are compared to determine which performs better on specific metrics (e.g., click-through rates).

EVALUATION

EXPERIEMENT CONTEXT___



en.wikipedia.org/wiki/Experiment

STUDIES

FIELD STUDIES___

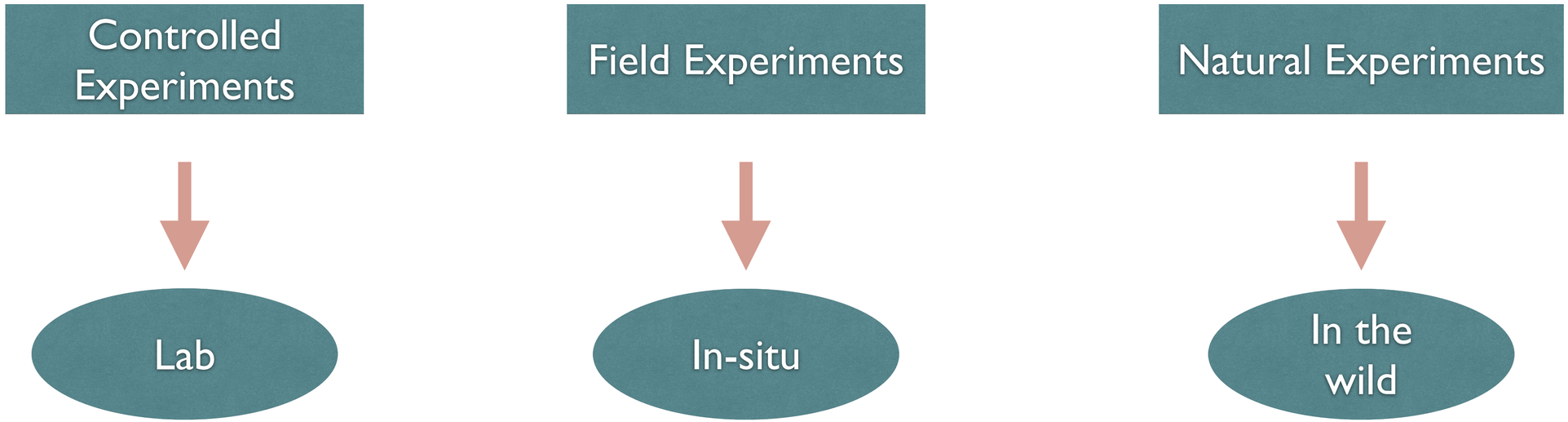
Purpose: To observe and analyze user behavior in real-world settings, providing ecological validity.

Methods:

- Ethnographic Studies: Researchers immerse themselves in the user's environment to observe natural interactions.
- Diary Studies: Users record their experiences and interactions with a system over time.
- Contextual Inquiry: Combines observation and interviews to understand user tasks and needs in their natural context.

EVALUATION

EXPERIEMENT CONTEXT___



en.wikipedia.org/wiki/Experiment

STUDIES

LONGITUDINAL STUDIES

Purpose: To study user behavior and system performance over an extended period.

Methods:

- Repeated Measures: Participants interact with the system multiple times, and their performance or satisfaction is tracked over time.
- Cohort Studies: A group of users is followed over time to observe changes in behavior or attitudes.
- Often in the-wild studies

EVALUATION

SETUP AND CONTEXT

What is the location and environment you conduct your study in:

- Context: lab, in-situ,..
- Technical setup: computer/ tablet, mouse/ phone….
- Environmental setup: personal computer, at their workplace,..
- Time setup: how long is the study, one time or long-term,..

EVALUATION

PARTICIPANTS

Who is the end-user of your system?

- What is their background? (Representation)
- What do they have to know? (Assumption)

How many condition and how are they tested?

- Within-Subjects Design: The same participants experience all conditions, reducing variability.
- Between-Subjects Design: Different participants are assigned to different conditions to avoid learning effects.

Rule of thumb:

~12 “small-n” statistics (student’s t)

~30 normally distributed tests (Z)

EVALUATION

PARTICIPANTS

Time Slot	Name	Age	Self-identified gender	Years of Experience Using Moodboard	Design Education	Current Job Title	Condition+Design Brief Combination	Familiarity with Generative AI
14/12 14:00	A	25	M	2 years	1 year of a research-creation program, one internship in HCI but I've been working very closely with designers on several projects	HCI Researcher	1A+2B	Used chatgpt and AI image generation
15/12 16:00	B	27	F	8 Yesrs	Bachelor in design 3 years; Master in design for art direction, 2 years	UI/UX Designer	2B+1A	Almost never
17/12 15:30	C	34	F	2 Years	Bachelor in Fashion Design; Master in UI UX Design	Product designer	1B+2A	Used Dalle before, know some basic prompt engineering
19/12 15:30	D	27	F	2 years	Bachelor in Architecture; Master in Human-computer interaction	HCI Researcher	2A+1B	Frequently used chatgpt, prob everyday, but not so much

EVALUATION PROCESS

In general:

- All subjects receive the same instructions
- All subjects perform same tasks under the same conditions
- All instructions are simple and clear
- Informal contact kept to a minimum

What is your baseline? What do you want to compare?

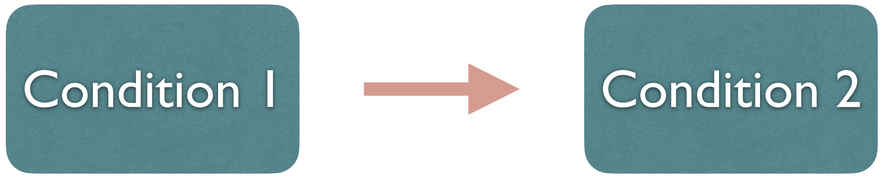
- What happens if there is no existing tool to compare?

EVALUATION PROCESS

How to avoid biases?

- Order effects/ Training effects
- Familiarity effects
- Carryover effects
- …

— > Balance primary and secondary factors!



Latin Square

- a Latin square is an n × n array filled with n different symbols, each occurring exactly once in each row and exactly once in each column
- a balanced Latin square additionally ensures that one symbol never follows another twice

EVALUATION PROCESS

	Condition 1	Condition 2
Task 1	A	B
Task 2	C	D

Set

P1: A +D

P2: C+ B

P3: D + A

P4: B + C

Latin Square

- a Latin square is an n × n array filled with n different symbols, each occurring exactly once in each row and exactly once in each column
- a balanced Latin square additionally ensures that one symbol never follows another twice

EVALUATION PROCESS ___

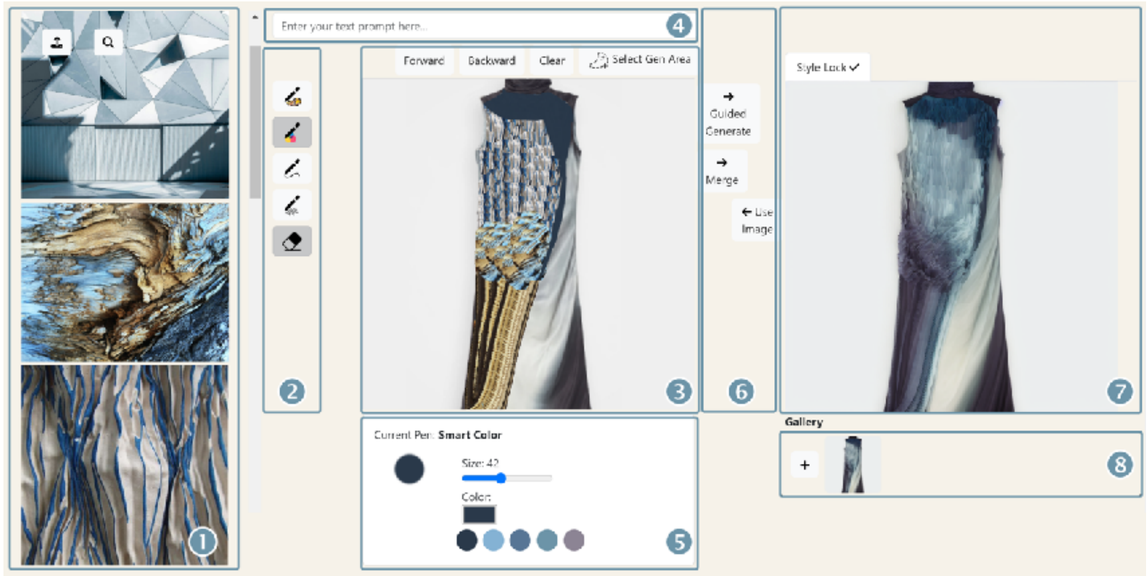
Example 1: Between system study

We use a [2x2] within-participant design with two factors:

- The **primary factor** related to the interaction strategies: Firefly condition (baseline) and Pen condition (variant).
- The **secondary factor** is the task, for which we created two equivalent design briefs.

The tasks are designed to require general design skills without needing highly specialized expertise in a specific niche.

Both, the design task and interaction strategies, were counterbalanced using a Latin-Square to account for order effects.



Xiaohan Peng, Janin Koch, and Wendy E. Mackay. 2025. FusAIn: Composing Generative AI Visual Prompts using Pen-based interaction. (CHI '25)

EVALUATION PROCESS

Example 1: Between system study

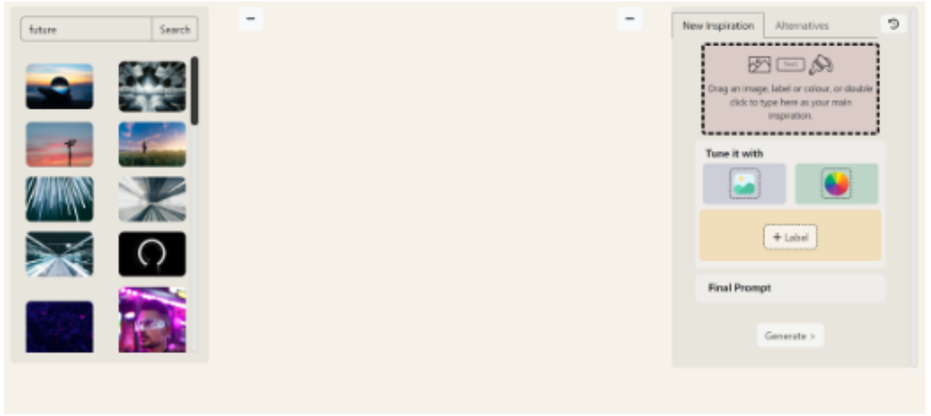
The study protocol consists of five steps:

1. **Introduction:** The experimenter first describes the study and obtains informed consent.
2. **Training**: Each task begins with a 10-15 minute training session. The experimenter walks users through an interactive tutorial describing all AI and basic image operation features. The training includes hands-on exercises,..
3. **Design Tasks:** Participants perform a 15-minute design task with the assigned condition using a think-aloud protocol. After each condition, participants fill out a **short questionnaire** with three Likert-scale questions regarding the ease of visual composition, intention alignment, and versatility. Then participants repeat the process with the second design task and other condition.
4. **Comparative Questionnaire**: The participant fills out a Likert-scale **questionnaire** with ten questions **to compare the two conditions.**
5. **Semi-Structured Interview:** The experimenter **conducts a semi-structured interview** that asks participants to reflect upon the different workflows, interaction techniques, and expectation alignment…

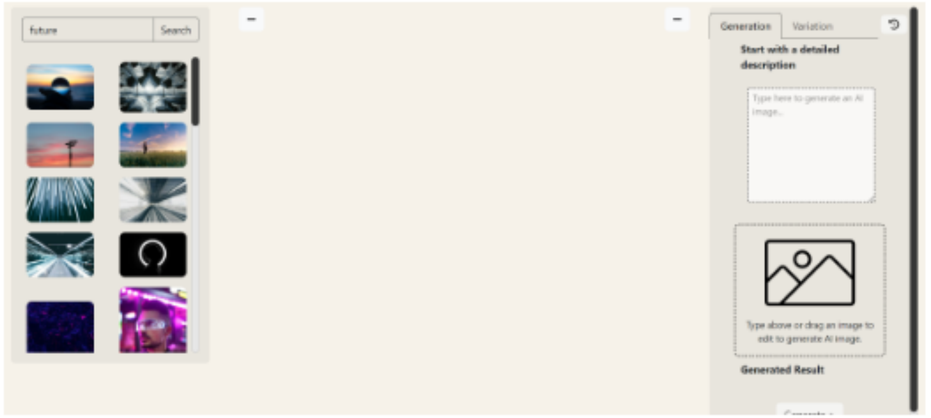
EVALUATION PROCESS

Example 2: Within system study

We use a [2x2] within- participant design with two factors.
- The **primary factor** related to prompt strategy: Text condition (baseline, text-only prompts) and Multi-modal condition (variant with multimodal prompts).
- We created **two equivalent design briefs** to generate the design tasks, with a Latin Square to counter-balance for order across conditions and participants.



(a) ***Multi-modal condition***



(b) ***Text condition***

Xiaohan Peng, Janin Koch, and Wendy E. Mackay. 2024. DesignPrompt: Using Multimodal Interaction for Design Exploration with Generative AI. (DIS '24)

EVALUATION PROCESS

Example 2: Within system study

The study protocol consists of four steps:

1. **Introduction and Tutorial:** The experimenter first describes the study and obtains informed consent. Participants then view a tutorial that describes how DesignPrompt works and have several minutes to familiarize themselves with different prompt strategies.
2. **Design Tasks:** Participants perform two moodboard design tasks with a think-aloud protocol . Each task begins with a short video tutorial with details about the current prompt strategy. Participants then read the assigned design brief and then perform a 12-minute moodboard task with the assigned prompt strategy. After answering a short questionnaire, they repeat the process with the second design task and the other prompt strategy.
3. **Comparative Questionnaire:** The participant fills out a Likert-scale questionnaire to compare and contrast the two prompt strategies.
4. **Comparative Interview:** The experimenter conducts a semi-structured interview (10-20 minutes) that asks participants to reflect upon the different prompt strategies…

EVALUATION

DATA COLLECTION ___

Data	Quantitative	Qualitative
Subjective	• Questionnaires • Visual assessment scales	• Verbal statements • Textual feedback
Objective	• Task completion time • Physiological measures	• Grounded Theory

https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=bc202e6b1fbde29b65e9470beacbea5daf4ed26a
(Albrecht Schmidt)

EVALUATION

DATA ANALYSIS

Use a mixed-method approach for Human-AI interaction!

Qualitative:

- Likert Scale (frequency analysis)
- Semi-structured Interviews (Thematic Analysis)

Quantitative:

- Logs (statistical analysis — see ANOVA etc.)
- Questionnaires

EVALUATION

DATA ANALYSIS

Use a mixed-method approach for Human-AI interaction!

Qualitative:

- Likert Scale (frequency analysis)
- Semi-structured Interviews (Thematic Analysis)

Quantitative:

- Logs (statistical analysis — see ANOVA etc.)
- Questionnaires

HOMEWORK

EVALUATION STUDY___

- Make a study plan for Comparitive or Wizard of Oz study
- Consider:
  - Setup: What is the context of the study?
  - Participants: Who? How many?
  - Process: What are the tasks the user has to do?
  - Data Collection: What would you measure?
- Conduct a mini study (2-3) participants
  - Summarize 2 insights
  - Update prototype if necessary

ADVANCED DESIGN OF INTERACTIVE SYSTEMS

12 FEBRUARY 2026

Janin Koch
Permanent Researcher
Inria Lille
janin.koch@inria.fr