Tampereen yliopisto
Tampere University

Design Toolkit for Human-Centered AI (HCAI)

Created in the KITE project: human-centered AI solutions for the smart city (WP 4)

Lead author: Thomas Olsson, thomas.olsson@tuni.fi
Contributors: Kaisa Väänänen, Saara Ala-Luopa, Maria Hartikainen, Jouko Makkonen

What?

This toolkit offers a **collection of methods and tools for *human-centered AI design*** — that is, helping to be mindful of the possible users and surrounding society when *envisioning*, *specifying*, *prototyping*, and *testing* AI-infused applications.

For whom?

The toolkit is **intended for developer organizations** interested in human-centered AI.

For example, one might:

- Have experience in human-centered design but not in AI-infused applications
- Have experience in AI-development but not in human-centered design

How?

The slides offer summaries of recently published as well as classical methods that we consider especially useful.

For effective use of the methods and for continuous learning, please also see the links provided in the descriptions.

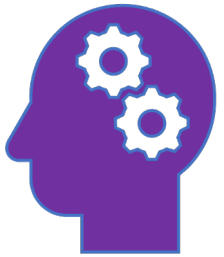
This document builds on the following earlier outputs of the KITE project:
1) Slide show: HCD in the age of AI
2) Public article: How Does AI Challenge Design Practice?

Please consult these for further background and terminological explanations.

Contents

How does AI change design thinking?

Summary of key points from earlier KITE documents



- **New agency:** The vast promises of AI introduce risk of thoughtlessly delegating complex decision making and reasoning to technology that only seems smart
- **New risks:** The complex and long-term sociotechnical implications of AI applications call for preventive risk mitigation
- **New mindsets:** AI applications challenge the conventional mental models of user-product relationships — e.g., from tools to collaborators

→ Need for methods that advocate forethought based on **diverse perspectives**, **holistic analysis** of the context, new **measures** and **criteria** of quality, and **reflexivity** within the design team.

→ From user-centric to **societal** and **ethical** design considerations

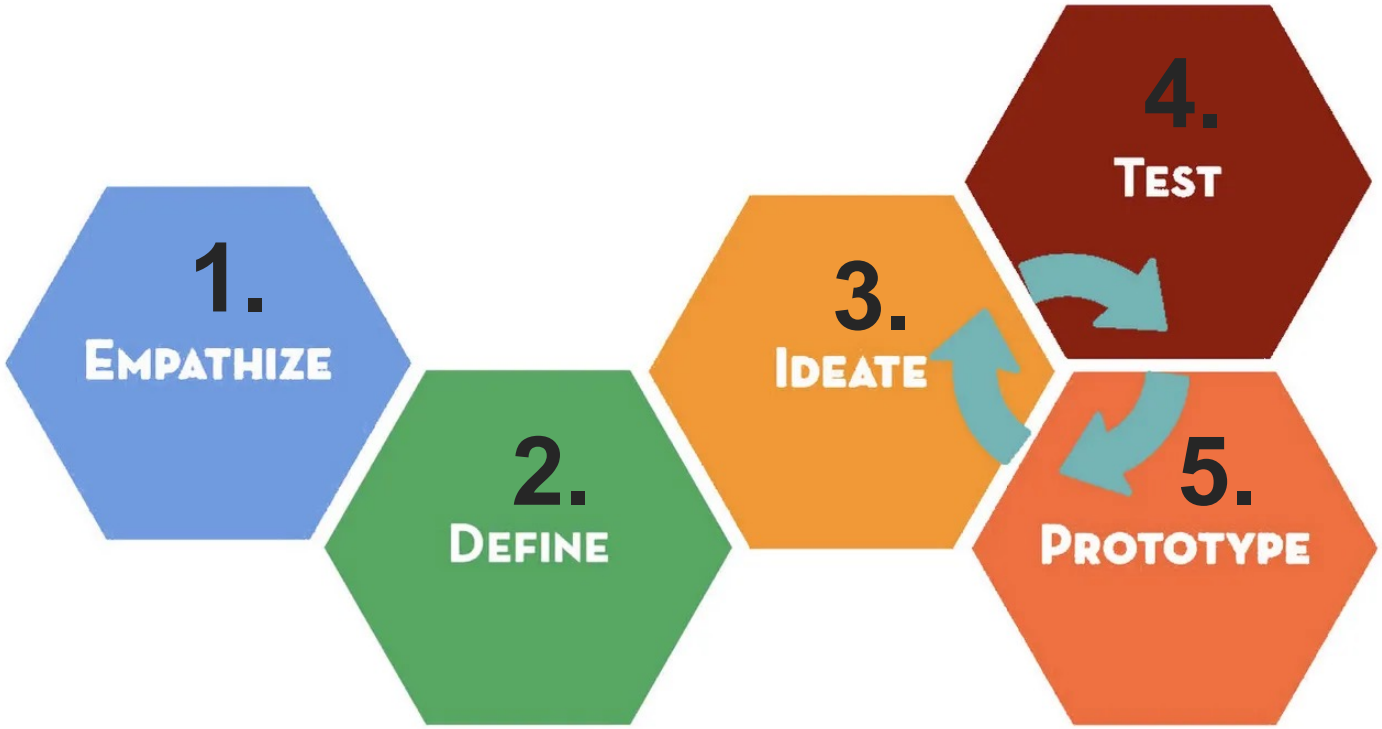
Tampereen yliopisto
Tampere University

Part 1.

Recommended methods and tools for HCAI design

Introduction to part 1

- The toolkit is organized according to the well-known *design thinking* process by Stanford's Hasso Plattner Institute of Design (d.school)
- For each of the five stages, we'll present alternative methods
  - We'll start from methods that appear easy to learn and adopt.
  - Many of the methods may be used throughout the design process!



Resource: An Introduction to Design Thinking PROCESS GUIDE

Tampereen yliopisto
Tampere University

All stages: **card deck** methods

Various card deck methods playfully offer **diverse viewpoints to consider** and **sources of inspiration** for all design projects

- The card decks are generally suitable for any project, organisation, and individual designer

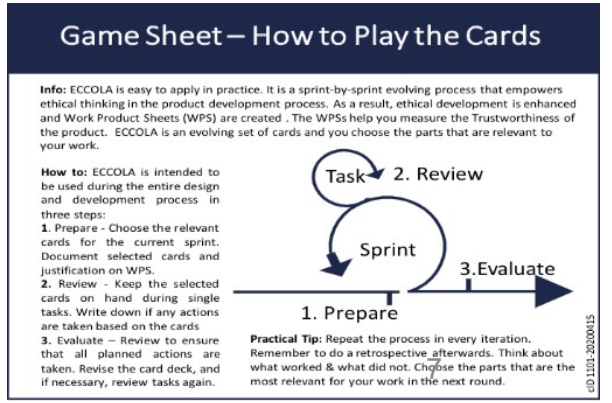
Each of the card decks (*see following slides*) may be used throughout the design process:

- To empathize with potential users, stakeholders, and various contexts of use
- To spur divergent thinking when envisioning new service concepts
- To assess possible effects in short- and long-term
- To refine early ideas

Tarot cards of technology

Envisioning cards

ECCOLA



Tarot cards of technology (link to webpage)

What: a set of provocations designed to help designers more fully consider the impact of technology on users, culture, and society, as well as think of various risks.

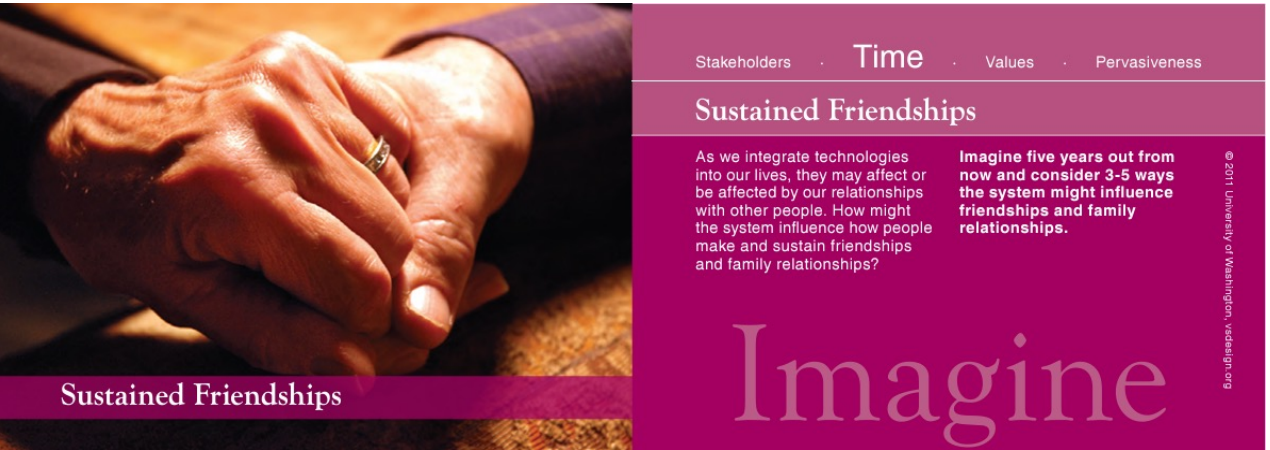
When: especially when defining, evaluating, and refining early service ideas.

***Examples**:*
All cards are available online!

Envisioning criteria and cards (Friedman & Hendry, 2012)

What: 28 themes and related activities to provoke thought and remind of important viewpoints

- "A versatile toolkit for attending to human values during design processes - and discuss their early use"
- The method relates to a broader design philosophy *Value-Sensitive Design*
- *Envisioning criteria:* stakeholders, time, values, pervasiveness
- 5 example cards available here *(please consult the authors for more cards)*



ECCOLA: Method for implementing ethically aligned AI systems

(Vakkuri, Kemell & Abrahamsson, 2020)

What: 21 cards to be utilized in software development projects, sprint-by-sprint

- Focus on ethical considerations of AI systems: e.g., transparency, data quality, agency & oversight, fairness, and value trade-offs

When: may be used step-by-step (linear process) or more creatively throughout the design & development process

Examples:
All cards are available online!


Trans parency

#4 Documenting Trade-offs

Motivation: One important part of transparent system development is the documentation of trade-offs. Whenever you make a decision, you choose one option over other alternatives. However, documenting why and what the alternatives were is important.

What to Do: Ask yourself:

- Are relevant interests and values implicated by the system and potential trade-offs between them identified and documented?
- Who decides on such trade-offs (e.g. between two competing solutions) and how? Did you ensure that the trade-off decision and the reasons behind it were documented?

Practical Example: Documenting trade-offs can improve your customer relationship, allowing you to better explain why certain decisions were made over others. Moreover, it can reduce the responsibility placed on the individual developer(s) from an ethical point of view.

ECCOLA

cID 3401-20200415


Agency & Oversight

#11 Human Oversight

Motivation: AI systems should support human decision-making. They should not undermine human autonomy by making decisions for us, meaning they should be subject to human oversight.

What to Do: Ask yourself:

- Who can control the system and how? In what situations?
- What would be the appropriate level of human control for this particular system and its use cases?
- Related to the Safety and Security cards: how do you detect and respond if something goes wrong? Does the system then stop entirely, partially, or would control be delegated to a human? Why?

Practical Example: Assuming control is especially related to cyber-physical systems such as drones or other vehicles. For purely digital systems, the focus should be on supporting human decision-making instead of directing it.

ECCOLA

cID 9301-20200415


Wellbeing

#16 Environmental Impact

Motivation: Past the general wellbeing implications, ecological consciousness is a current trend. Being ecological can be a selling point for your organization.

What to Do: Ask yourself:

- Did you assess the environmental impact of the system's development, deployment, and use? E.g., the type of energy used by the data centers.
- Did you consider the environmental impact when selecting specific technical solutions?
- Did you ensure measures to reduce the environmental impact of your system's life cycle?

Practical Example: If you are hosting on a third party cloud, try to ascertain the sustainability of the service provider's services. If you are using hardware, are you processing the data in each physical device of your own or are you processing it in the cloud?

ECCOLA

cID 8101-20200415


Account ability

#20 Minimizing Negative Impacts

Motivation: Minimizing negative impacts of the system is financially important for any developer organization. Incidents are often costly.

What to Do:

- First, consider...
  - Is your stakeholder analysis is up-to-date (Card #0)
  - Have you discussed risks? (Card #13)
  - Have you discussed auditability? (Card #18)
  - Have you discussed redress issues? (Card #19)
- Are the people involved with the development of the system also involved with it during its operational life? If not, they may not feel as accountable.
- Are you aware of laws related to the system?
- Can users of the system somehow report vulnerabilities, risks, and other issues in the system?
- With whom have you discussed accountability and other ethical

ECCOLA

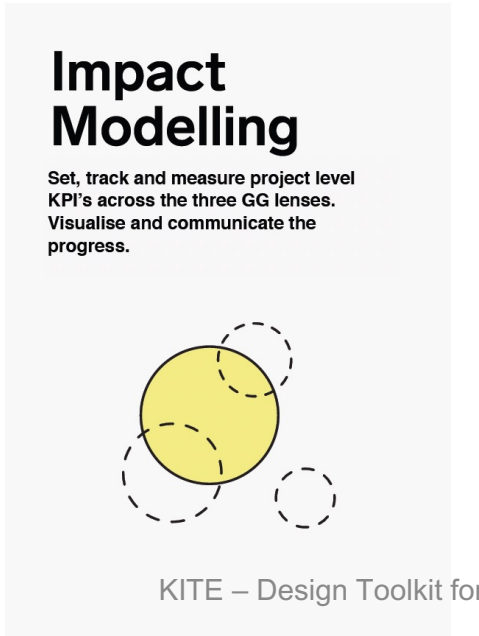
cID 9301-20200415

Good growth (Anton Schubert, Good growth company)

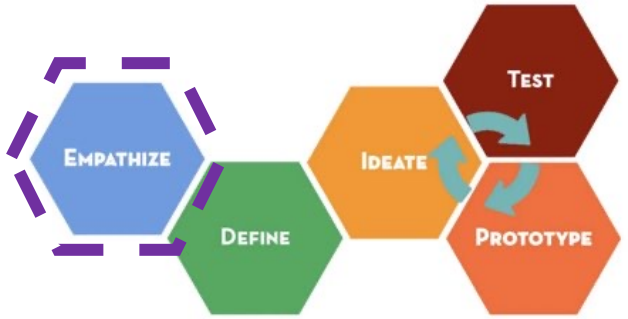
What: A set of reminders, goals, and considerations to spur sustainable business and digitalization

- Advocate for, e.g., impact modeling in respect to UN SDGs, digital resilience, rethinking the drivers of design work, and importance of validating solutions with a forum of respected external experts

Examples: *Consult the author for more!*



The **Empathize** stage



”...the work you do to **understand people**, within the context of your design challenge.” (Design Thinking Process Guide, d.school)

- Aims at understanding *possible user groups* and *stakeholders impacted* by the possible solutions. What is important to them? What are their key problems to be solved by technology?

Conventional HCD methods range from observation and ethnography to interviewing, surveys, and diary/journaling methods

- Ideo’s Design Kit: an extensive collection of user research methods
- GSA’s HCD Discovery Stage Field Guide: an easy-to-digest overview of HCD principles, processes, and methods
- **Challenge**: while many of these methods work also in HCAI, AI-powered solutions may introduce new systemic benefits and risks.
- **Challenge**: it is hard to foresee how AI-features might influence in long-term.

Overview of HCAI design methods for **Empathizing**

(order: from easier to more challenging)

- **1.1 Analogous projects:** learning from past design cases
- **1.2 Targets of benefit vs. harm:** who might be affected and how?
- **1.3 Pre-mortems:** preventive risk sweeping
- **1.4 Value scenarios:** clarifying the starting points
- **1.5 AI design probes:** provocations to understand user needs

See the following slides for more details and examples!

1.1 Analogous projects:
learning from the past

What: Identification of relevant *parallels with earlier projects* – both your own and public projects

- Learn from the past choices, risks analyses, development processes, evaluations of solution proposals, and strategies to mitigate issues

How: Analyze what makes the current and a past case similar / dissimilar. Build on their successes and avoid their failures.

- *Overall conditions*: technology maturity level, cultural & societal context, legislation
- *Aims*: key values, underlying premises, metrics of success, KPIs
- *Area of impact*: target users, key stakeholders, business domain
  - Consider economical, environmental, and social sustainability
- *The solution*: key features, mechanisms, agency & role of technology

1.2 Targets of benefit vs. harm

What: identifying the groups of people that the system might *indirectly* influence—either positively or negatively.

- Who might be discriminated by the application, whose interests might be compromized, will the produced benefits be fairly distributed?
- **To consider**: do the benefits significantly outweigh the harms?

How:

- List the secondary users and people whose jobs or activities might be affected indirectly or in long-term → Identify and list both the possible benefits and harms
- It is important to involve a diverse group of people (including known stakeholders) in the analysis to avoid groupthink and to employ diverse viewpoints.

1.3 **Pre-mortems**: preventive risk sweeping

What: In a pre-mortem, a project team imagines a scenario where the project/product has failed, and then works backward to determine what potentially could lead to the failure.

- The point is to **speculate** about possible futures to recognize risks and develop suitable preventive actions.

How: get inspiration from *design noir* (example canvas) and *speculative design* methods

(relevant books: Discursive design, Speculative design)

1.4 Value Scenarios (Nathan et al., 2007)

What: a narrative of *desirable sociotechnical future* where certain choices have made a positive impact on how a technology meaningfully fits the intended purposes and context of use.

How: Relevant elements in a value scenario include:

- *Stakeholders*: envision a range of effects of a technology, both direct and indirect stakeholders
- *Pervasiveness*: consider a vision in which a technology has become widespread, spanning various geographic regions, cultures, social classes, and other contexts
- *Time*: take into consideration what the world might look like five, ten, or twenty years after a technology has been deployed
- *Systemic Effects*: explore the multidimensional interactions among technology, psychology, society, culture, and the environment as use of the technology becomes pervasive years
- *Value Implications*: envision not only positive effects of technology but also its darker consequences

Example of a visually polished and thought-through scenario

An excerpt from a textual, early-stage scenario

On the first morning of the conference, Sam arrives in a hurry at the big lobby of the congress center. His mobile phone checks for available services and beeps to notify Sam. "Voice guidance available", reads on the phone display, and Sam puts his wireless earplug on. A pleasant voice tells him that there is a queue to the registration desk. The voice leads Sam to a public screen, where he can register as well as see a map and information about the happenings at the center.

Sam touches *Register for a conference* spot by the screen with his mobile phone and confirms the registration suggested on the phone display. The screen changes to show the program of Sam's conference and a list of people he might know present at the congress center. Before the conference Sam has set a meeting with Peter who he met in an interest group in the conference web portal. Peter's name is also in the list. The voice guidance remarks Sam that Peter has arrived and that their meeting is at 14 o'clock.

The voice on Sam's earplug however now tells his that the conference program is about to begin. Sam asks for the shortest route to the auditorium. On the way he sees several shared public screens that offer a connection to different sites around the world. Universities have screens on their auditoriums. In addition, registered students may virtually follow the conference presentations…

1.5 **AI design probes** *(cultural probes)*

What: A probe is a small artifact (e.g., a map, postcard, camera, or diary) along with evocative tasks, which are given to study participants to allow them to record specific events, feelings or interactions as part of their everyday life.

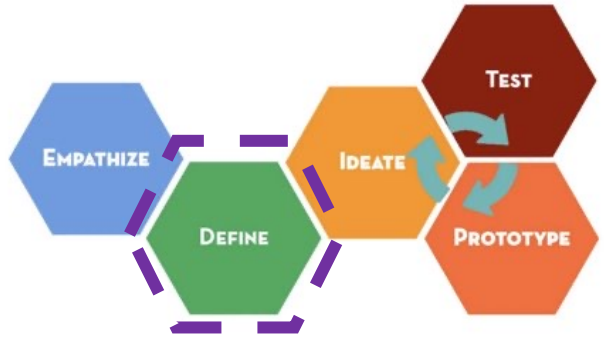
- The aim is to elicit interesting experiences, ideas, or needs from people, in order to understand their culture, thoughts and values, and thus provide design inspiration.

https://balslevdesignstudio.com/DESIGN-PROBES

How:

- Requires long-term participation and plenty of time but can provide foundational insights as basis for service design
- Concrete examples: ServiceDesignLab, Catherine Legros, and doctoral thesis by Tuuli Mattelmäki

The **Define** stage



Specifying ”a meaningful and actionable problem statement”, a ’point-of-view’ or ’design challenge’.

- *“Framing the right **problem** is the only way to create the right **solution**.”*
- What is the problem that provides **focus**, **inspires** your team, informs **criteria** for evaluating ideas, and **captures** the hearts and minds of people you meet?

Conventional HCD methods include, e.g., journey mapping, concept mapping, root cause analysis, analysis of pain points

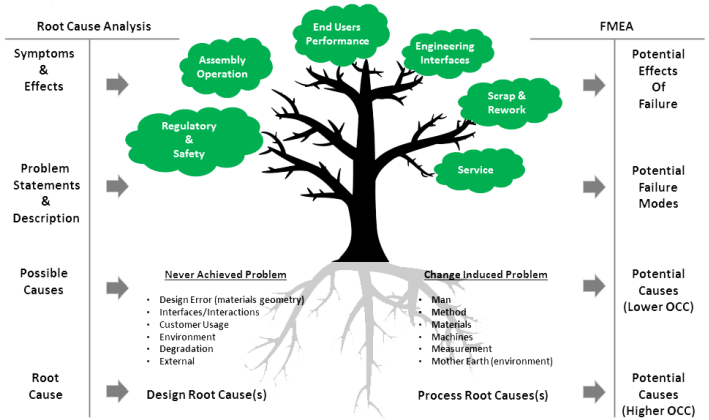
- See, e.g., Ideo’s Design Kit: an extensive collection of analysis methods
- **Challenge**: AI-powered solutions may support also problems that they were not originally designed for
- **Challenge**: AI-development can suffer from *solutionism* (a belief that all difficulties have benign technological solutions)

HCAI design methods for **Defining**

- **2.1 Root cause analysis:** finding a problem worth solving
- **2.2 Go or no-go:** maximizing the good or minimizing the bad
- **2.3 Canvas methods on design ethics** (by ethicsfordesigners.com)
- **2.4 User-product relationship**: redefining the roles of technology and humans
- **2.5 Value dams and flows:** recognizing issues and tradeoffs in values

See the following slides for more details and examples!

2.1 Root cause analysis



Example from http://theroutex.com/root-cause-analysis-rca/

- Search for other visual representations of RCA online!

What: Often the problems identified at early stages of a project are only *symptoms* of more profound problems

- To identify the "root cause" or a "problem worth solving", continue asking '*why*' (e.g., "*5 x Why*"). etc.

How? Example of an *identified problem*: "HR experts appear sceptical about the idea of a recruitment chatbot"

- *Why?* – they feel that chatbots shouldn't meddle with recruiters' work
- *Why?* – based on their experiences of other chatbots, they don't trust that it could answer applicants' tricky open-ended questions in Finnish
- *Why?* – they believe that applicants deserve human-based communication with the target company
- *Why is this important?* – it's necessary to have two-directional and unstructured communication with the candidate to find the best match → etc. etc.

2.2 Go or no-go:

How to maximize the positive and minimize the negative?

What: Reminding why are we doing this project, and for what *good ends*

- For example, what long-term sustainability aims do we contribute to (in addition to short-term business objectives)?
- See, e.g., UN sustainability goals for inspiration

How: Critically assess the premises of the project

- Will society, the world, and our customers really be better off with this solution than without it? Or are we trying to generate inauthentic needs or manufactured desires, simply to justify a new thing to sell?
- What are we willing to sacrifice to do this right?
- How probable are the expected benefits and identified risks of the solution?

2.3 Canvas methods on design ethics (www.ethicsfordesigners.com)

A collection of simple canvas-based methods to help specify and scope the solution appropriately.

Four canvases appear particularly relevant to HCAI *(see following slides)*:

- *De-scription*: analyzing the foundations of a design idea
- *Moral value map*: clarifying the values that drive your design work
- *Normative design scheme*: applying normative ethics theories to analyze the moral quality of a design idea
- *Ethical disclaimer*: dealing with moral responsibilities in a project

2.3 Canvas: **De-scription** (link)

May be used to both analyze existing systems or reconsider an ongoing project.

De-scription

This quick exercise trains your moral sensitivity as a designer. Learn to recognise and deconstruct the scripts of existing designs. By questioning why a design is the way it is, you'll uncover the underlying intentions and world-view of its designer.

SPECS

Suggested Time
30 minutes
Materials needed
This template or large sheet of paper, pens, an example product
Participants
Designer duo
Process phase
Framing, or whenever you want to train your moral sensitivity

PROCESS

1 Describe the 'WHAT' of the design. Use the questions on the template.

2 Describe the 'HOW' of the design. Use this to determine the script: Like the script of a movie or a theatre play, an artefact can 'prescribe' its users how to act when they use it. Write this down.

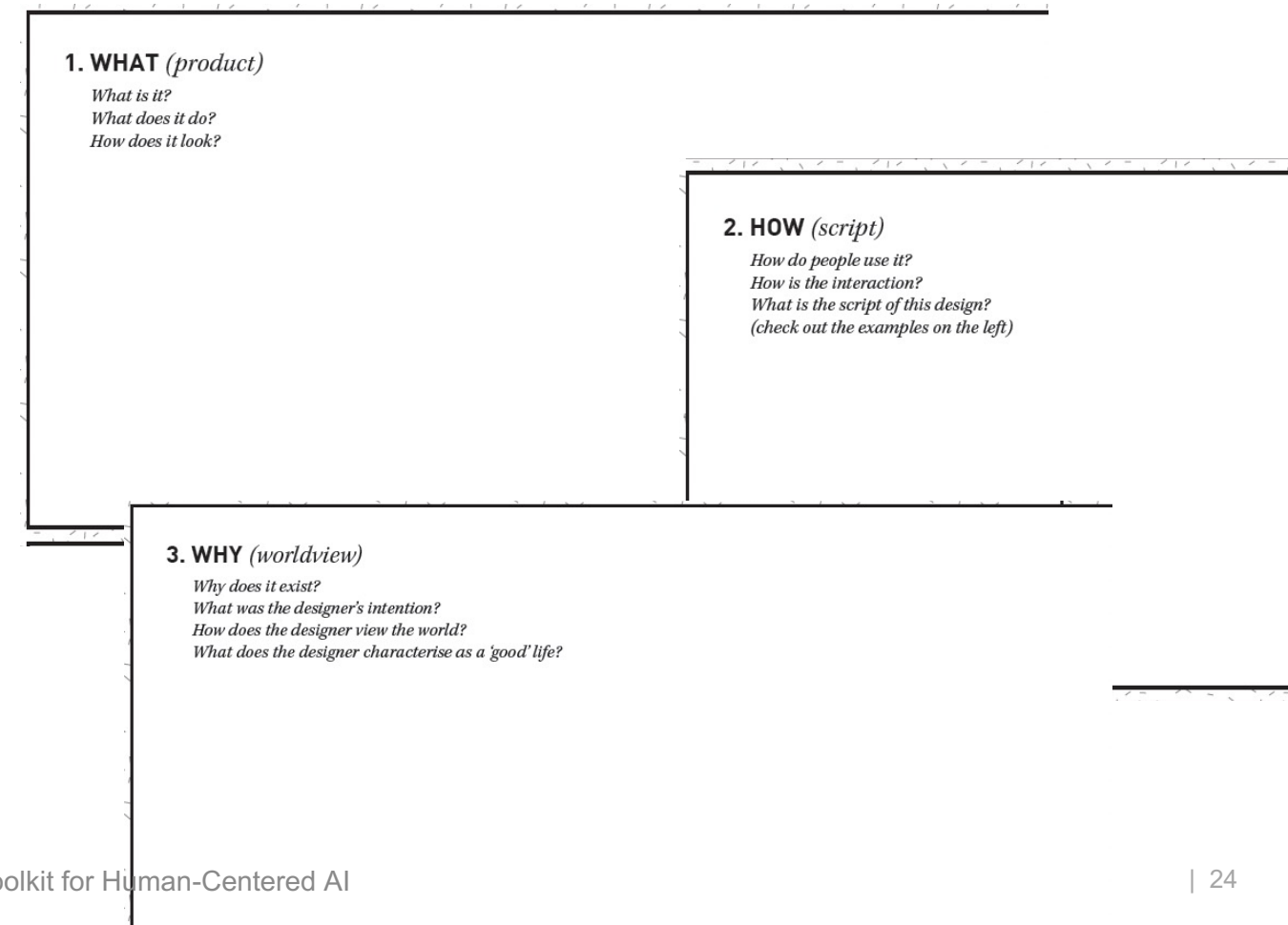
3 Describe the 'WHY' of the design. Use this to determine the underlying worldview. Write this down.

**For inspiration check out the readings on general philosophical worldviews, go to www.ethicsfordesigners.com/de-scription*

EXAMPLES

For speed bumps, the script is: 'Slow down!'

For paper coffee cups, the script is: 'Dispose me after use'



ethicsfordesigners.com

Tampereen yliopisto
Tampere University

2.3 Canvas: **Moral value map** (link)

To clarify the values that drive your design work.

Moral value map

Values are an important concept in ethics. With this tool you will look at which values are relevant to your design and how your design affects them. Doing this exercise with different stakeholders helps to understand everyone's value priorities.

ethicsfordesigners.com

SPECS

Suggested Time
30 - 60 minutes

Materials needed
Overview universal human values, this template, post-its, pens

Participants
Design team, stakeholders

Process phase
Framing, validating

2. SELECT VALUES RELEVANT TO YOUR DESIGN:

Entertainment: Experiencing excitement or heightened arousal

Physical well being: Feeling healthy, energetic or physically robust

Receiving: Obtaining approval, support or validation from others

Bodily sensations: Experiencing pleasure associated with the senses, physical movement or bodily contact

Exploration: Satisfying one's curiosity about personally meaningful events

Understanding: Gaining knowledge or making sense out of something

Positive self: Maintaining a sense of self-confidence, pride or self-worth

Transcendence: Experiencing optimal or extraordinary states of functioning

Harmony: Experiencing a profound sense of connectedness, harmony or oneness with people, nature or a greater power

Tranquility: Feeling relaxed and at ease

Belonging: Building or maintaining attachments, friendships, intimacy or a sense of community

Equity: Promoting fairness, justice, reciprocity or equality

Giving: Giving approval, support, assistance, advice or validation to others

Responsibility: Keeping interpersonal commitments, meeting social role obligations and conforming to social rules

Management: Maintaining order, organisation or productivity in daily life tasks

Creativity: Engaging in activities involving artistic expression or novel though

Safety: Being unharmed, physically secure and free from risk

Mastery: Meeting a challenging standard of achievement or improvement

Uniqueness: Feeling unique, special or different.

Autonomy: Experiencing a sense of freedom to act or make choices.

Superiority: Comparing favorably to others in terms of winning, status or success.

Intellectual stimulation: Engaging in activities involving original thinking or novel or interesting ideas

3. TRANSLATE VALUES TO CONCERNS:

HUMAN VALUE **X** CONTEXT OF USE **=** CONCERN

e.g. AUTONOMY X virtual reality gaming = "I want to be in control of my own actions in the game"

4. MAP THE CONCERNS:

Which concerns does each stakeholder prioritise?
Which concerns conflict?

5. DESCRIBE THE DESIGN'S EFFECT ON EACH CONCERN

*The design **inhibits/supports/limits/enhances/prevents/enables/reinforces/undermines/challenges** this concern.*

2.3 Canvas: **Normative Design Scheme** (link)

Helps to consider the moral nature of your design goal, using different normative ethics theories

ethicsfordesigners.com

2. OUR DESIGN GOAL:

WE WANT TO...	BY DOING...	IN ORDER TO...
THE INTENTION	THE DESIGN	THE EFFECT

3. ASSESS YOUR DESIGN GOAL:

VIRTUE ETHICS	DEONTOLOGY	CONSEQUENTIALISM
"An action is morally right just because it would be done by a virtuous person acting in character."	*"An action is morally right just because it abides moral rules and serves a person's moral duty."*	*"An action is morally right just because it produces the best actual or expected results for the most people."*
Is your design morally right? Why (not)? How could you change that?	*Is your design morally right? Why (not)? How could you change that?*	*Is your design morally right? Why (not)? How could you change that?*

4. BRAINSTORM WITH NORMATIVE THEORIES:

1. Which virtues could your design affect? Positively or negatively? 2. How could your design stimulate virtuous behaviour? 3. Consider the following virtues: wisdom, courage, moderation, justice and honesty. 4. What kind of person does your design appeal to?	1. What would the world be like if everyone used your design? 2. Would you use your design? 3. Would you let your children use your design? 4. Which moral or social norms apply to your design?	1. What are the consequences of your design for the different stakeholders? 2. How could your design have positive consequences? 3. How could your design cause the greatest happiness for the greatest number?

Tampereen yliopisto
Tampere University

2.3 Canvas: **Ethical disclaimer** (link)

To outline the moral responsibilities of a project & developer organization.

Ethical disclaimer

This tool helps you set the ethical terms at the start of your project. Use your imagination to think of unethical situations and discuss what you'll take responsibility for as designers.

SPECS

ethicsfordesigners.com

Suggested Time
45 - 60 minutes

Materials needed
This template, post-its, pens

Participants
Design team, stakeholders if possible

Process phase
Framing, envisioning, validating

PROCESS

1. Describe the current situation of your design context.
2. List all the people/companies/ institutions that have an interest in, or are affected by your design. Both direct stakeholders such as users and your client and indirect such as maintenance.
3. Write down your intentions. Be specific in terms of users, context and purpose. The more explicit, the easier you can discuss them with stakeholders throughout the project.
4. Imagine the context with your design in it. (If you don't know yet what you're designing, think of ways to change the current situation.) Think of situations with your design which might be unethical. Think from the perspective of the different stakeholders. How will they use/contribute to/gain from/be harmed by your design? Use post-its to speed things up.
5. Discuss for which of the unethical situations you will take responsibility by moving them down to the left or right section of the template. Substantiate your choices.

3. OUR DESIGN INTENTIONS ARE...

What do you want to change in the current situation? Why?

4. UNETHICAL SITUATIONS WITH OUR DESIGN...

What dark, wrong or weird situations can you come up with? What would be really wrong?
What happens when it is used over a longer period of time?
What if it is used by the 'wrong' people? Or shared by different people?
What if it is used in a different context? Or for a different purpose? Or combined with other technology?
Who can access it? What happens if people hack it?

5. WE TAKE RESPONSIBILITY FOR...

Which situations fall within the ethical scope of the project? Why?

5. WE DO NOT TAKE RESPONSIBILITY FOR...

Which situations are outside the ethical scope of the project? Why?
Which might be someone else's responsibility?

2.4 User-product relationship

What: (Re)defining the roles of both technology and its users.

- AI may turn human-computer interaction into *human-machine integration, human-machine collaboration,* or *human-machine teaming*.
  - AI can either automate certain tasks or augment human intelligence

How:

Viewpoints to consider:

- Understanding the user needs in certain context: evaluating and assessing both the benefits, risks, and consequences of introducing AI-based solutions
  - AI enthusiasm involves the risk of thoughtlessly delegating complex decision making and reasoning to opaque algorithms
  - What are the challenges or decision-making stages where AI could assist the user and how?
- *What might be the "scripts of collaboration" or "conflict-management procedures"?*
  - *Step-by-step implementation, and user control and autonomy on focus*
- How does the user *coach* their AI companions? How to motivate the user to continue teaching the system after the initial setup?

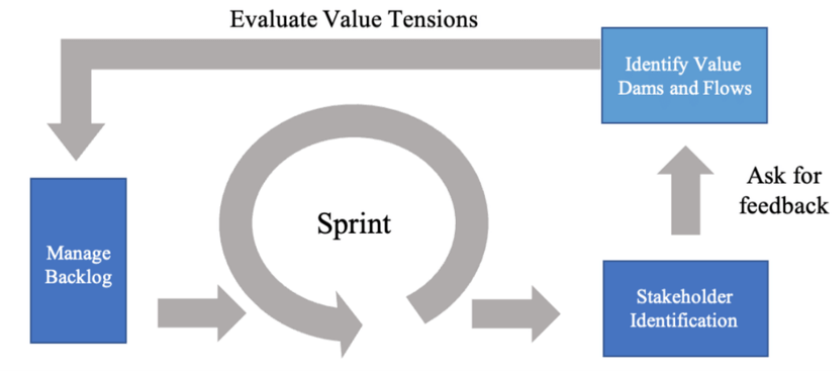
2.5 Value Dams and Flows

Part of Value-Sensitive Design methods (Miller et al., 2007)

What: Analytic method to reduce the solution space and resolve value tensions among design choices

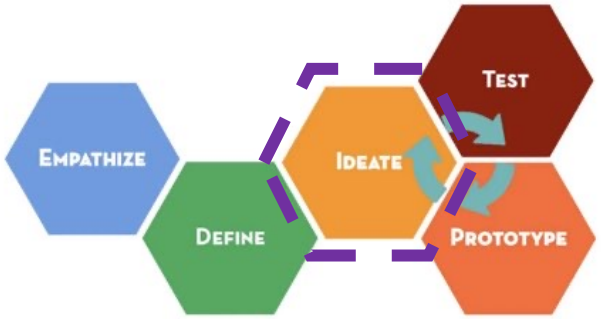
- *Value dams* are defined as technical features or organizational policies that some stakeholders strongly oppose, causing a value tension
- *Value flows* call attention to features and policies that a large number of stakeholders are in favour of incorporating.

How: the point is to either address the tension or emphasize the flow, or to mark these for attention at a later stage in the design process.



Value dams and flows in Agile development: Umbrello & Gambelin (2021)

The **Ideate** stage



Stage of innovation and generating ideas to solve the previously defined problem

- Aims at providing large numbers of alternative solutions through lateral thinking

Conventional HCD methods work perfectly fine also when ideating AI-powered applications!

- Examples include: Facilitated brainstorming, Ideation canvases, How Might We, bodystorming, Mindmapping, etc.
- General rules and tips for brainstorming
- The following slides highlight some viewpoints that are important in this context

HCAI design methods for Ideating

- 3.1 Perspective taking
- 3.2 Distant concepts
- 3.3 Extreme and contrafactual imaginaries

Also remember the card deck methods mentioned earlier!

3.1 Perspective taking: "How might we...?"

E.g.: Nielsen & Norman group

What: In idea creation, "How might we" can be used to take different perspectives and thereby fuel creativity.

How: Take a starting point (existing product, problem, target of improvement) and think how to address from different viewpoints.

Example of possible perspectives in a taxi redesign case:

Emphasize the good > "How might we ...make taxi rides feel like racing?"
Remove the bad > "… remove the perception of waiting in traffic?"
Question an assumption > "… rethink the number of drivers per car?"
Use analogy > "… make taxi rides as enjoyable as visiting hairdresser?"
Identify unexpected resources > "… make the passenger more engaged?"
Change the temporal perspective > "… make the rides feel like in the 1800's or 2100's"
Change context > "… make underground taxis? ... create a feeling of a rickshaw ride?"
Add superpowers > "… make taxi rides feel superhuman?"

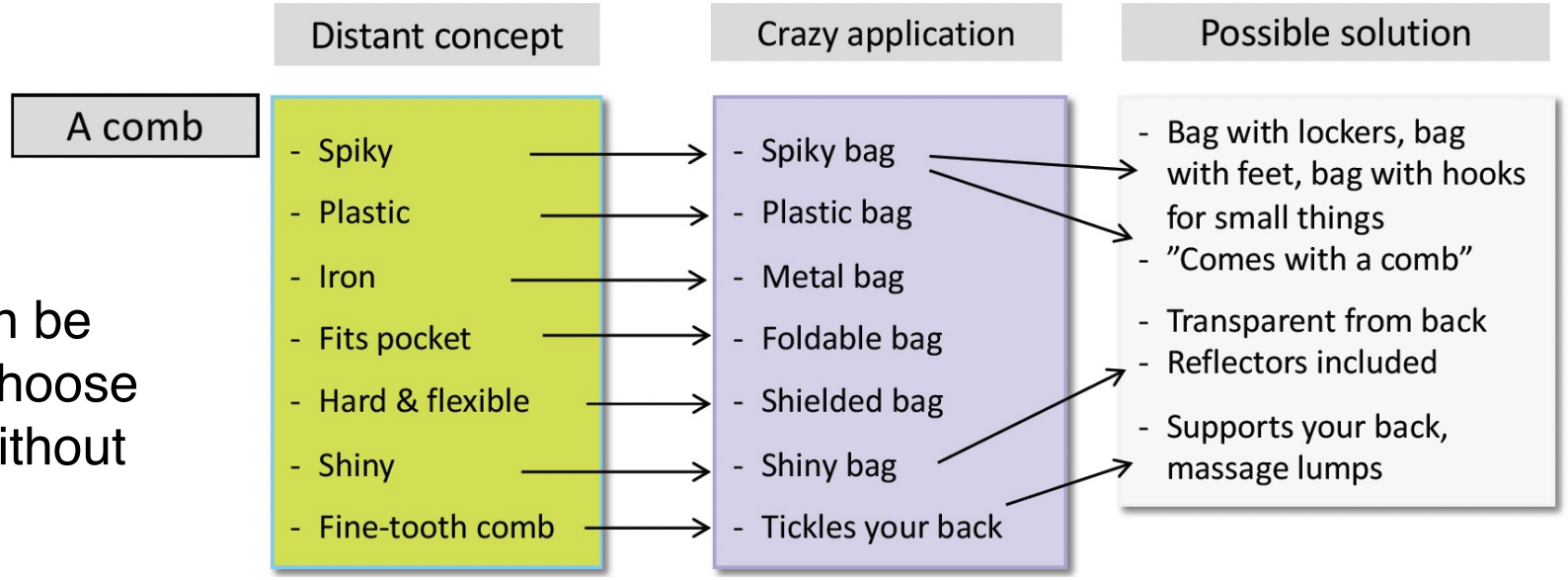
3.2 Distant concepts

What: Using metaphors to create unexpected ideas and trigger *lateral thinking*.

How: Take an object that is *conceptually far* from what you are designing. Consider its characteristics and what they could mean in your case.

Example: designing a new schoolbag by thinking of a *comb*.

The distant concept can be almost anything. Just choose something randomly, without hesitation!



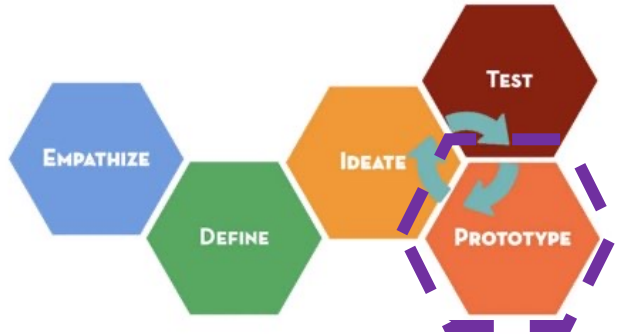
3.3 Extreme and counterfactual imaginaries

What: imagining versions of the future that are **extreme** in some regard, in order to identify desirable compromises and risks to avoid.

How:

- Expose yourself to fictional stories or scenarios about possible futures, for example those in sci-fi literature/games/movies that haven't become mainstream.
- Think what could be an extremely *utopian* vs. *dystopian* versions of a future where a brainstormed idea has been deployed at large. What could go extremely wrong or well – considering society, culture, and the planet?
- Consider *counterfactual* versions of the *present* to identify desirable *futures:* e.g., where would we be if internet hadn't been free in the beginning?
  - E.g. the Future Ripples canvas

The **Prototype** stage



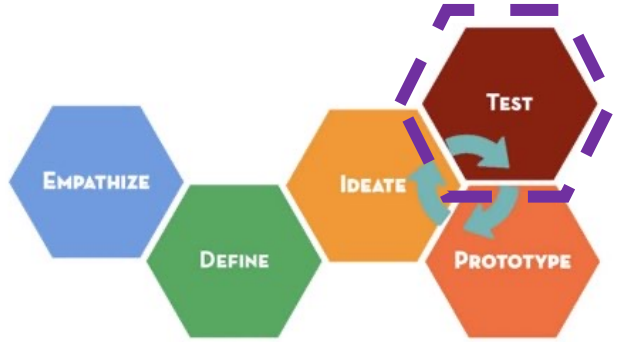
”Iterative generation of artifacts intended to answer questions that get you closer to your final solution.” (Design Thinking Process Guide, d.school)

- The aims are to solve the problems and to understand it better by outlining solutions, to communicate ideas, to learn (”fail fast”) or to show next directions

Conventional HCD methods: prototype is something a user can interact with

- A wall of post-it notes, a gadget, a role-playing activity, a storyboard, Wizard of Oz simulation
- https://www.designkit.org/methods/determine-what-to-prototype
- https://www.designkit.org/methods/rapid-prototyping

The **Test** stage (a.k.a. Evaluation)



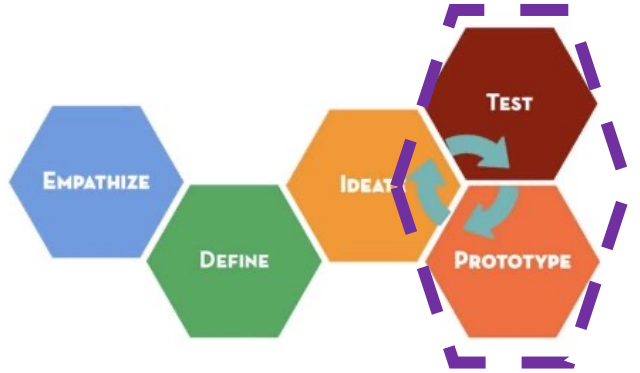
Gathering feedback and experiences about the prototypes and identifying issues to be remedied.

- Iteratively provides feedback to earlier design stages
- Gain more empathy and understand the people impacted by the technology
- Helps to improve the prototypes and scenarios
- Possibly also refine the Point of View (the design problem)

Conventional HCD methods include: Guerrilla testing to validate the concept, experimental comparisons between several prototypes, Wizard of Oz simulation, field testing

- https://www.interaction-design.org/literature/article/stage-5-in-the-design-thinking-process-test

HCAI methods for **Prototyping** & **Testing**



Note that many traditional prototyping & testing methods can be used also when developing AI-based applications! Therefore, the following only highlights a few key methods and examples from recent literature.

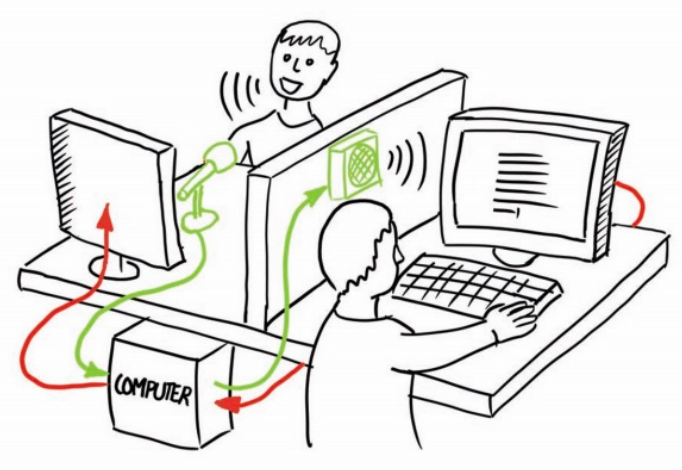
- 4.1 Wizard of Oz: lo-fi simulation of the customer's experience
- 4.2 Life-cycle assessment of scaled up designs
- 4.3 UI tool to evaluate AI fairness

4.1 Wizard of Oz – simulated user experiences

What: users may interact with a system that they *believe* to be autonomous, but in reality it is controlled by a human operator

- With high-agency AI applications, it's crucial to also consider the acceptability of having a technological actor instead of a human
- Typically used to test costly concepts inexpensively and to focus on the user/customer experience of a new service concept.
  - E.g. chatbots, recommender systems, decision-support systems

- **How**: plenty of good-old HCI resources are still useful, e.g., by Nielsen & Norman group, UC San Diego, a practical service design book, UX4sight



A wizard of Oz prototype for a listening typewriter by Adam Russ-Wood & Bill Buxton

4.2 Life-cycle assessment of scaled up designs (see, e.g., Brown et al., 2019 and life-cycle assessment in general)

What: Focus on *scaling up* to explore the long-term impact of concept ideas or early prototypes. Create scenarios that describe imaginary futures of the given design prototype.

How:

- *Upscale* your ideas in terms of: amount of training data; number of users; variety of user groups; different activities or use cases it could support; the broader societal & political conditions in which it would operate; and different technological *ecologies* the solution might become part of
- Think of the possible impacts of your AI application in terms of society, culture, specific user groups, and environmental burden
- Consider different *time horizons*: 1... 5... 10... 20 years, etc.

Resources: Tarot Cards of Technology (slide 8), end-of-life analysis, various sustainability impact assessment frameworks (e.g., this master's thesis)

4.3 UI tools to evaluate AI fairness

- **What**: Nakao et al. 2022 offer a user interface that supports not only data scientists but also domain experts to investigate AI fairness (in this case study, evaluated in the context of loan applications)
- **For whom:** Data scientists, designers and domain experts

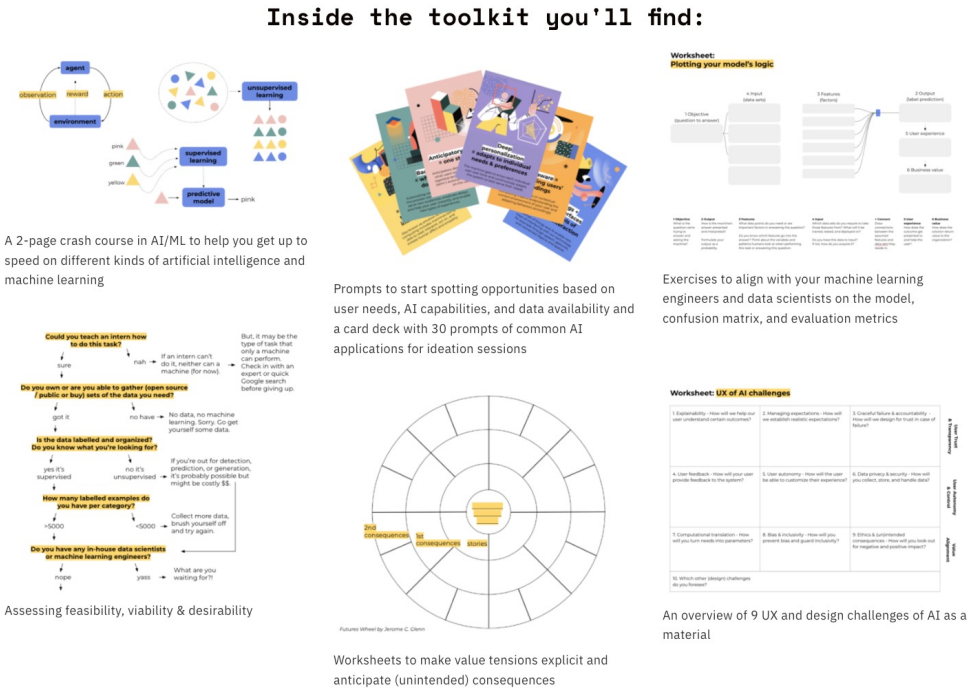
Tampereen yliopisto
Tampere University

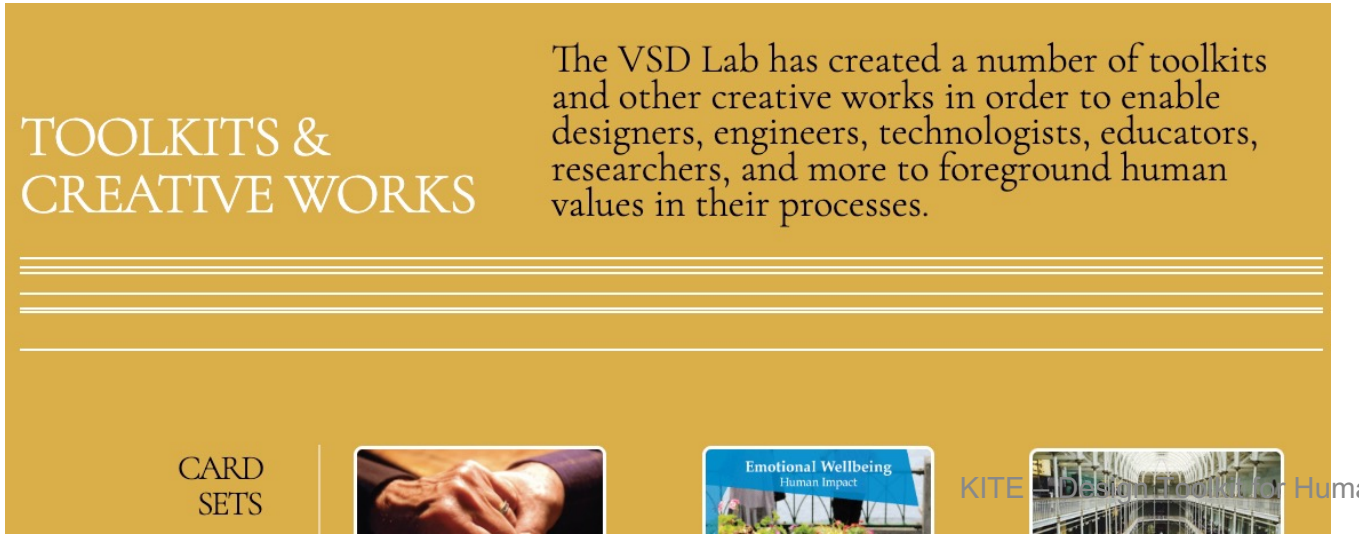
Part 2.

General HCAI guidelines & principles + additional resources

Other general HCAI toolkits and relevant resources

- https://www.aixdesign.co/toolkit
- https://vsdesign.org/toolkits/
- https://www.ethicsfordesigners.com/





Recommended guidelines & checklists:

Large corporations' HCAI guides

- IBM Design for AI
- Microsoft guidelines for human-AI interaction
- Microsoft Responsible AI Standard, & impact assessment
- Google People & AI guidebook
  - Google Flights example: quite general-level guidance
- Delloitte: Why artificial intelligence needs human-centered design
- Designing responsibly with AI
- Toward human-centered ai: a perspective from human-computer interaction
- Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI

Relationship between design, data, and AI models throughout the process

- There are several approaches on when and how designers are involved with **data** and AI models (Windl et al. 2022)
  - Design team is not required to deeply engage in the data aspects of the project.
    - A-priori approach assumes the model to be finished before the start of the UI development. Designers need to fulfil the requirements of the model.
    - Post-hoc approach: the UI design is completed before the model, implying that the model is built to fit the requirements of the design.
  - The design team is engaged with the data
    - Model-centric approach, the model is at the core of the project and the designers are strongly involved in nearly all technical aspects.
    - Competence-centric approach leverages a clear focus on the heterogeneous expertise within the team. Designers engage with the data, however not in primary role.
- Pair design
  - Intensive **cooperation of designers and data scientists** throughout the design process
  - https://bootcamp.uxdesign.cc/human-centered-ai-design-process-part-2-empathize-hypothesis-6065db967716
  - https://pair.withgoogle.com/guidebook

Tampereen yliopisto
Tampere University

Part 3.

Holistic design approaches and methodologies relevant to HCAI

Introduction to part 3

While parts 1 and 2 focus on individual methods, tools, and general HCAI guidelines, there are various **design approaches and methodologies** that appear useful for HCAI design.

- The following enriches the toolkit by offering more in-depth material for those who are interested in developing advanced design competences.

After a brief overview, we cover two well-established methodologies that we consider to offer particularly relevant perspectives and practices:

- Value-Sensitive Design
- Data-enabled Design
- Critical Design

Overview of approaches that underline viewpoints relevant to HCAI – six categories with relevant links

Responsibility-oriented approaches

- Socially responsible design
- Social design

Sustainability-oriented approaches

- Sustainable design (environmental)
- Design for sustainability
- Planet-centric design

Data-driven *(or data-enabled)* **design**

Diversity- and equality-sensitive approaches

- Inclusive design
- Design4all
- Universal design
- Feminist design
- Post-colonial design

Provocative and futuristic approaches

- *Critical design*
- Discursive design
- Speculative design

Value-sensitive approaches

- *Value-sensitive design*
- Worth-centered design
- Reflective design

Tampereen yliopisto
Tampere University

Value-Sensitive Design

(Batya Friedman et al., since 90's)

A pioneering design approach that comprehensibly and proactively accounts for human values in design

- Meant to encourage designers to consider ethics and values in the design process
- Strongly connected to the early discussions on *bias in computer systems*: envisioned as an approach to minimize bias

VSD has seen some success out on the field as well, with e.g. Intel and Microsoft utilizing it in some projects (Manders-Huits, 2011).

But which values to consider?

Friedman et al. (2008)

Value: *"what a person or group of people consider important in life"*

The authors' original list includes:

- Human Welfare,
- Ownership and Property,
- Privacy,
- Freedom From Bias,
- Universal Usability,
- Trust, Autonomy,
- Informed Consent,
- Accountability,
- Identity,
- Calmness,
- and Environmental Sustainability.

But many others may be considered as well:
peacefulness, compassion, love, warmth, creativity, humor, originality, vision, friendship, cooperation, collaboration, community, purposefulness, devotion, diplomacy, kindness, aesthetics, harmony...

Values are embedded in almost all decisions

Importance of differentiating between **facts vs. values**

Example of a value conflict between individuals' desires and societal safety during the Covid-19 pandemic, Dailypress.net

Designers' values
vs.
Users' values
vs.
Customer's values
vs.
Values that *emerge* while specifying a product

Three orientations of VSD:
conceptual, empirical and technical investigations

Conceptual investigations: identification and definition of relevant values and tensions/trade-offs between competing values.

Empirical investigations refine and expand key values identified in the conceptual investigations

assessing stakeholders' experience of the value-oriented features of a system at various stages of the design, development, and deployment process.

Technical investigations focus on the design of the technology itself

- how the technical properties support or hinder human values, and the proactive design of the system to support values identified in the conceptual investigation.

VSD methods

Friedman et al. 2017 report a survey of 14 VSD methods:

(1) direct and indirect stakeholder analysis
(2) value source analysis
(3) co-evolution of technology and social structure
(4) value scenarios
(5) value sketch
(6) value-oriented semi- structured interview
(7) scalable information dimensions
(8) value-oriented coding manual
(9) value-oriented mock-up, prototype, or field deployment
(10) ethnographically informed inquiry regarding values and tech
(11) model of informed consent online
(12) value dams and flows
(13) value sensitive action-reflection model
(14) Envisioning Cards

Data-driven / data-enabled design (DED) - designing with and based on data

Data-driven design (King et al.) emphasises the power of data to evaluate and validate design decisions

- Data can be qualitative or quantitative

Data-enabled design (DED) (Kollenburg & Sanders; Noortman et al.) provides a more designerly perspective, where data is used as creative material to inspire and inform new design interventions

- DED is characterized as a "quick and iterative approach that sets out to design, deploy and remotely adapt prototypes while they are situated in the daily life of people" (Funk et al.)

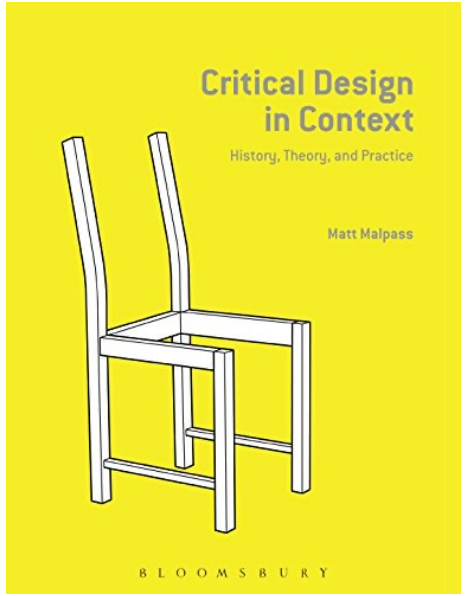
A resource to support this approach is the Data Foundry at TU Eindhoven: https://data.id.tue.nl/

References:

- Rochelle King, Elizabeth F Churchill, and Caitlin Tan. 2017. Designing with Data. O'Reilly Media.
- Janne van Kollenburg and Sander Bogers. 2019. Data-Enabled Design. Ph. D. Dissertation. Eindhoven University of Technology.
- Mathias Funk, Eden Chiang, and Eva van der Born. 2019. Data foundry: a data infrastructure for design research. In Data Science Summit 2019: The Annual Data Science Summit Eindhoven, organized by TU/e-DSCE
- Noortman, R. R., Lövei, P., Funk, M., Deckers, E. J. L., Wensveen, S. A. G. & Eggen, J. H. Breaking up data-enabled design: expanding and scaling up for the clinical context. 19 May 2022, In: Artificial Intelligence for Engineering Design, Analysis and Manufacturing. 36, 1, p. 1-13 13 p., e19.

Critical Design and Discursive Design

(e.g., Bardzells, 2013, Dunne & Raby, 2001 & Johannessen, 2013)



Aim to **provoke** new questions and reflection, rather than to solve problems

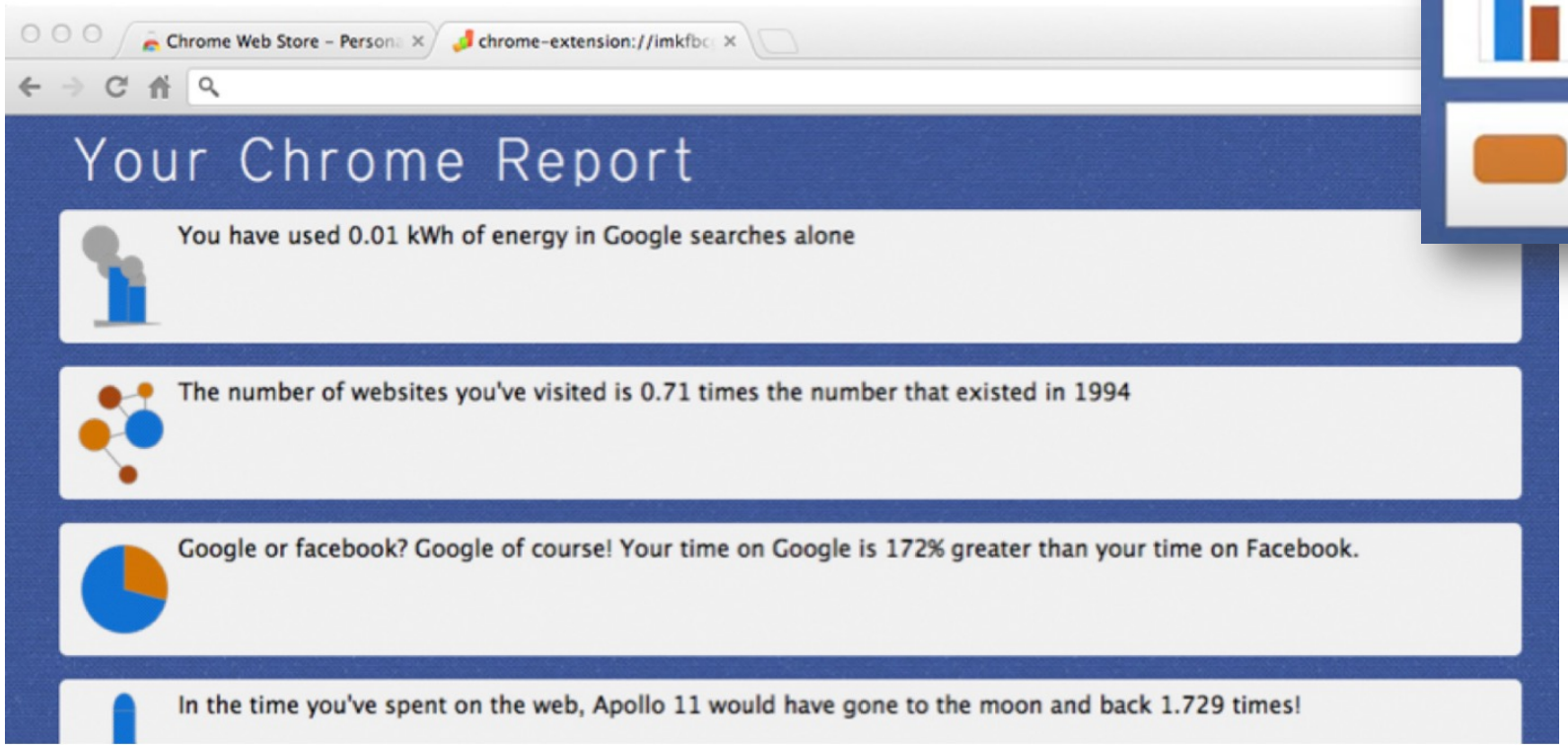
- Applies knowledge from social sciences and humanities for *reflective design* of artefacts

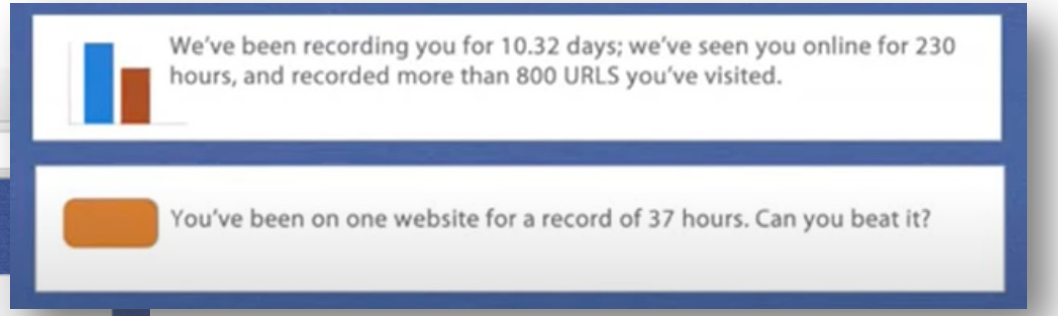
Foreground the *ethics of design*, aiming to reveal potentially *hidden agendas and values*, and explore *alternatives*

Underline the identification of design conventions: *twisting*

Example of a *digital* artefact:

"Everybody Knows What You're Doing": A Critical Design Approach to Personal Informatics (Khovanskaya et al. 2013) – abstract in Youtube





Instead of motivating users to examine their own behaviors, can *creepy*, *malfunctioning*, and *strange* personal informatics promote a\ data mining infrastructures themselves?

Examples of *non-digital* artefacts

Critical artefacts involve "propositional content" and ”counter-factuals”

Teddy Bear Bloodbag Radio
(Dunne & Raby)

Pets/animals as entertainment & source of power

Porcupine Dress
(Amisha Gadani)

Violence towards women

CCD-Me-Not
(Mark Shepard)

Hiding from surveillance systems

Contrasting “traditional” with critical

(Johannessen, 2017:, The Young Designer’s Guide to Speculative and Critical Design)

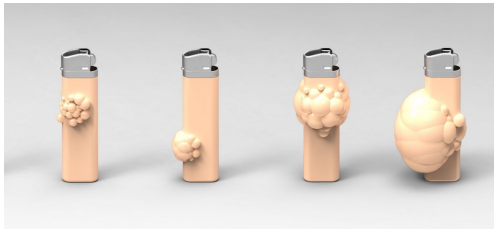
	Traditional design	SCD
Attitude	**Normative**	**Critical**
Foundation	**Information**	**Speculation**
Mindset	**Pragmatic** Productive	**Idealistic** Dreaming
Purpose	**Commercial** Satisfy industry’s need to make money	**Discursive** Spur debate on the development of society
Goal	**Develop solutions** Provide answers by solving problems	**Explore ideas** Find problems by asking questions
Intent	**Serve a user** In seriousness provide clarity	**Provoke an audience** Use ambiguity to make satire

Four uses of Critical Design

(adapted from Johannessen, 2017)

Fully-fledged design

Questioning conventions



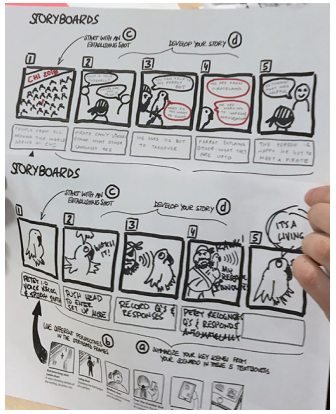
Influencing people

For other designers

For broad audience

Study of designers' values

Provoking thought on values in user research



Lo-fi prototype

Questioning the conventions

E.g., Pierce et al. (2015): Expanding and Refining Design and Criticality in HCI

A key approach to Critical Design is questioning and twisting existing conventions and trends in design

Some possible tactics for *twisting*:

- **Subvert**: destroying or damaging something essential in a design
- **Reinterpret**: seeing a design in new light
- **Exaggerate**: represent something as better, worse, larger, etc. than it really is
- **Juxtapose**: place close to something for a contrasting effect
- **Apply irony & satire**: expressions that signify the opposite

Thank you!

For further info, please contact us at:

thomas.olsson@tuni.fi
kaisa.vaananen@tuni.fi
saara.ala-luopa@tuni.fi
maria.hartikainen@tuni.fi
jouko.makkonen@tuni.fi