Design Challenges for Privacy-Preserving Explanation User Interfaces

Meetup GI SIG on User-Centered Artificial Intelligence | April, 7 2022

Univ.-Prof. Dr. Claudia Müller-Birn

Human-Centered Computing Research Group | Freie Universität Berlin

Human-Centered Computing (HCC) Research Group

Prof. Dr. Claudia Müller-Birn

HCC team (10 researchers with interdisciplinary backgrounds plus four students)

My vision is to advance the research and the design of socially responsible technologies of human-computer collaboration by investigating the interconnectedness of humans, data, algorithms, and environments based on a critical-reflexive practice.

HCC Profile

- **Research Areas:** Contributing to Collaborative Computing and Human-Computer Interaction
- **Methods:** Ranging from online experiments (MTurk) to participatory design
- **Scope:** Designing novel interaction concepts for augmenting human-decision making

HCC Projects

DocTalk

freemove
Privacy-Centered Urban Mobility Data

enkis

Berlin University Alliance

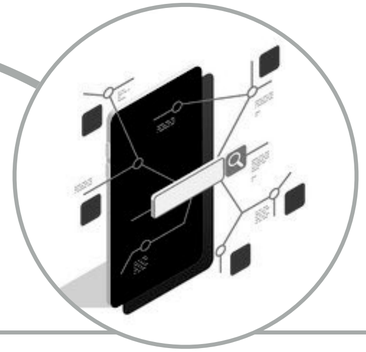
Matters of Activity
Image Space Material

Interactive Systems

complement humans

Human-Computer Collaboration

two or more actors work together to achieve shared goals

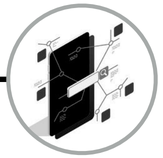


Autonomous Systems

emulate humans

- Agreeing on shared goals
- **Allocating responsibility**
- Coordinating activities
- Sharing context
- Learning mutually

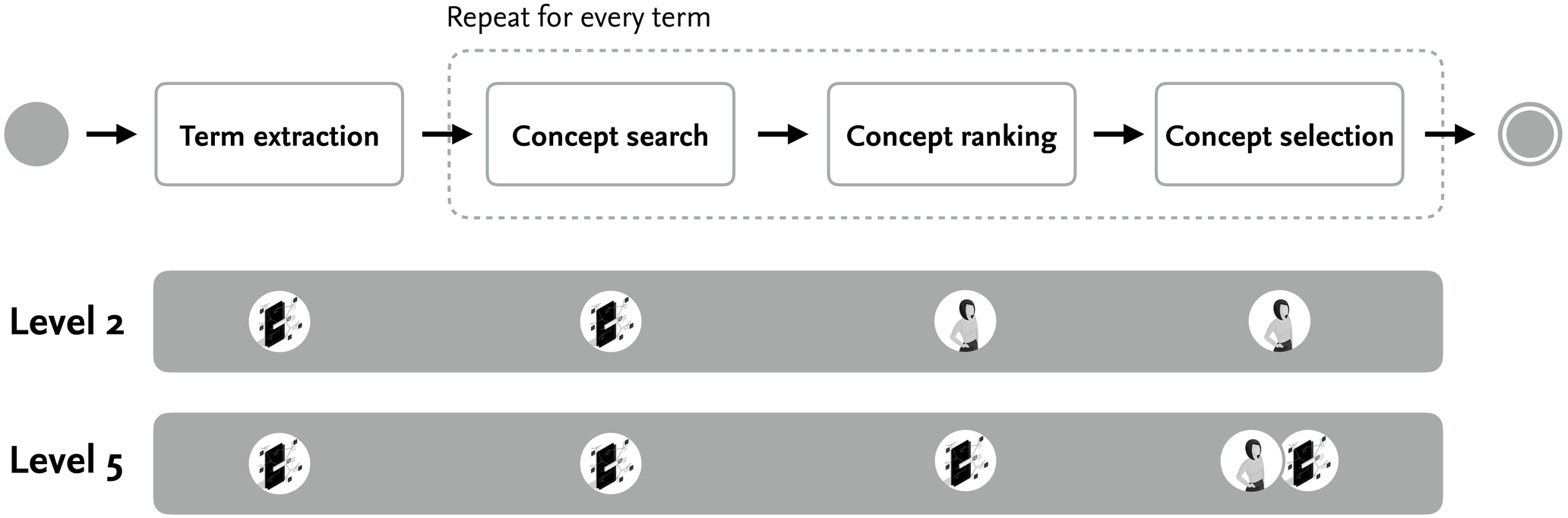
Levels of (Human Interaction with) Automation



1	2	3	4	5	6	7	8	9	10
The computer offers no assistance; the human must take all decisions and actions.	The computer offers a complete set of decision/action alternatives.	The computer narrows the selection down to a few.	The computer suggests one alternative.	The computer executes that suggestion if the human approves.	The computer allows the human a restricted time to veto before automatic execution.	The computer executes automatic-ally, then necessarily informs the human.	The computer informs the human only if asked.	The computer informs the human only if it, the computer, decides to.	The computer decides everything and acts auto-nomously, ignoring the human.

Mackeprang, M.; Müller-Birn, C.; Stauss, M. (2019). Discovering the Sweet Spot of Human-Computer Configurations: A Case Study in Information Extraction. Proc. ACM Hum.-Comput. Interact. 3, CSCW, Article 195.

Applying Levels of Automation on Information Extraction



Mackeprang, M.; Müller-Birn, C.; Stauss, M. (2019). Discovering the Sweet Spot of Human-Computer Configurations: A Case Study in Information Extraction. Proc. ACM Hum.-Comput. Interact. 3, CSCW, Article 195.

Implications

» LoA provided a **better understanding of the design space** of human-computer collaboration.

» **As higher the level as more unlikely people are to challenge** the provided recommendations when unaccompanied by explanations.

» Choosing the ‘sweet spot’ of **human-computer collaboration is context-dependent,** thus, it is not a simple substitution problem.

» **LoA limits thinking about alternative designs** since it exhibits a certain understanding of human-computer collaboration (↑ automation ≠ ↓ control).

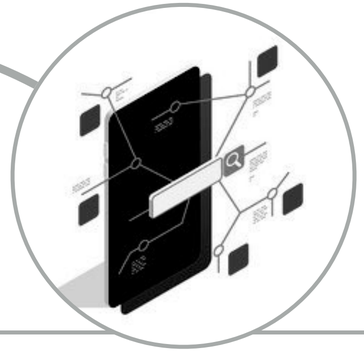
Mackeprang, M.; Müller-Birn, C.; Stauss, M. (2019). Discovering the Sweet Spot of Human-Computer Configurations: A Case Study in Information Extraction. Proc. ACM Hum.-Comput. Interact. 3, CSCW, Article 195.

Interactive Systems

complement humans

Human-Computer Collaboration

two or more actors work together to achieve shared goals



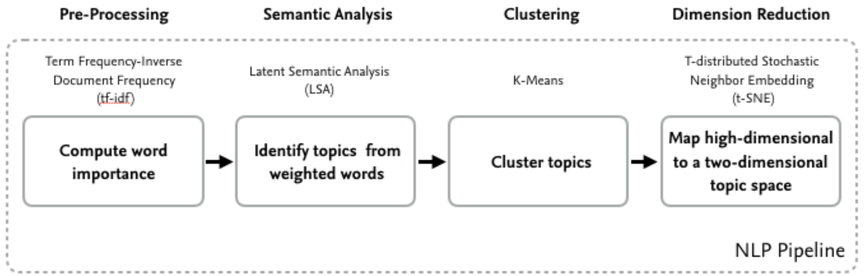
Autonomous Systems

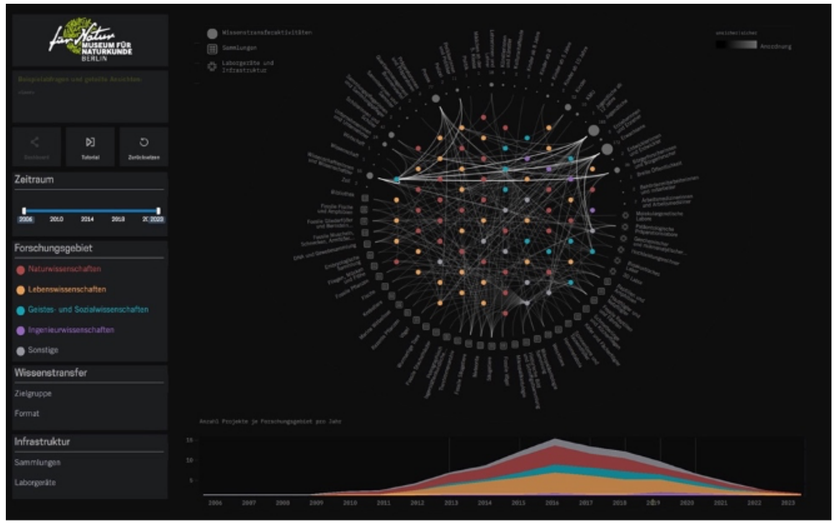
emulate humans

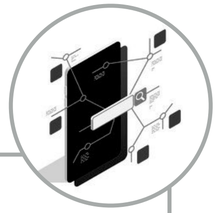
Agreeing on shared goals

- **Allocating responsibility**
- **Coordinating activities**
- **Sharing context**
- **Learning mutually**

Study Context: Data Visualization for Collaborative Research



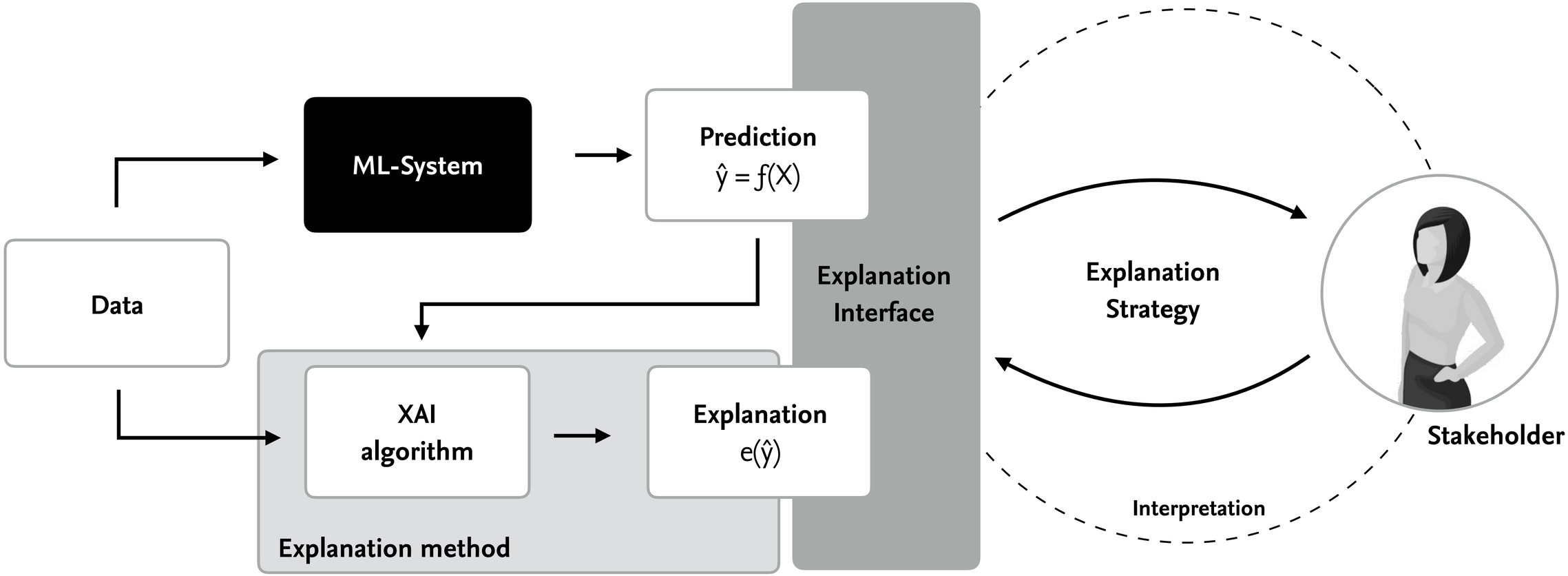




Research Information → **NLP Pipeline** → **Human Interpretation**

Jesse Josua Benjamin, Christoph Kinkeldey, Claudia Müller-Birn, Tim Korjakow, and Eva-Maria Herbst. 2022. Explanation Strategies as an Empirical-Analytical Lens for Socio-Technical Contextualization of Machine Learning Interpretability. Proc. ACM Hum.-Comput. Interact. 6, GROUP.

Ensuring Interpretability by Explanation User Interfaces



Jesse Josua Benjamin, Christoph Kinkeldey, Claudia Müller-Birn, Tim Korjakow, and Eva-Maria Herbst. 2022. Explanation Strategies as an Empirical-Analytical Lens for Socio-Technical Contextualization of Machine Learning Interpretability. Proc. ACM Hum.-Comput. Interact. 6, GROUP.

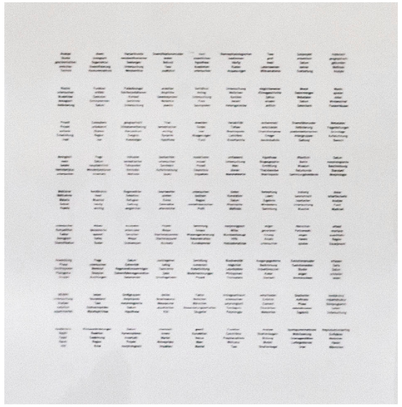
Selected Explanation Methods

Explanation Methods

Local Techniques

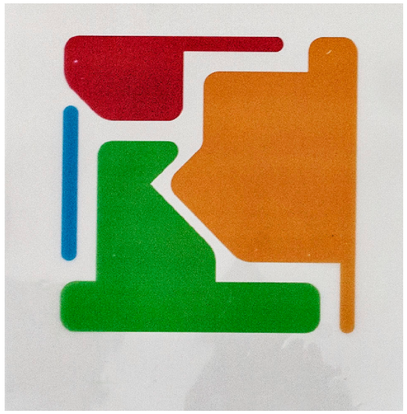
Feature Importance

Project Topwords



Similarity

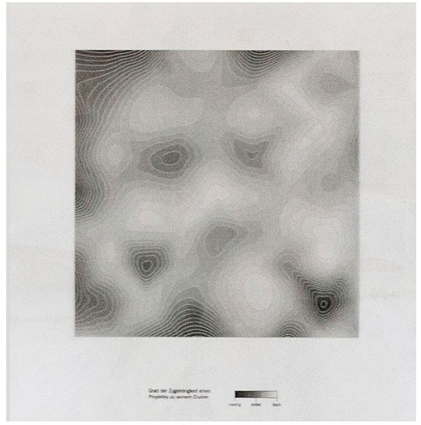
Cluster Regions



Global Techniques

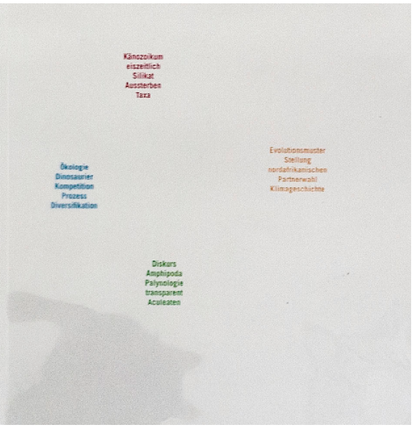
Regions of Error

Uncertainty Map



Feature Importance

Cluster Topwards



Jesse Josua Benjamin, Christoph Kinkeldey, Claudia Müller-Birn, Tim Korjakow, and Eva-Maria Herbst. 2022. Explanation Strategies as an Empirical-Analytical Lens for Socio-Technical Contextualization of Machine Learning Interpretability. Proc. ACM Hum.-Comput. Interact. 6, GROUP.

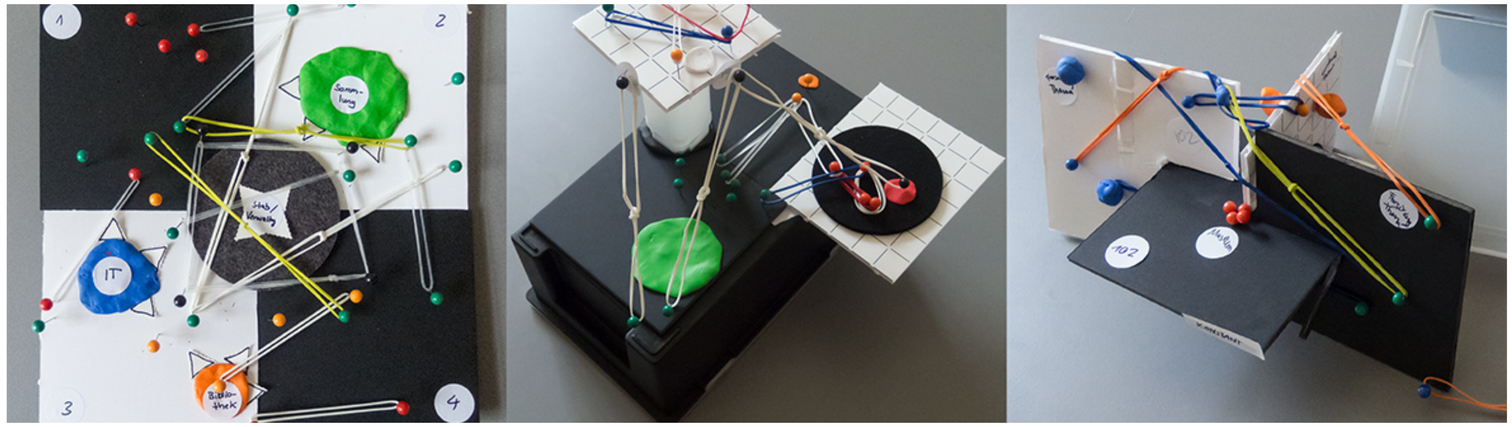
Participatory Design for Revealing Explanation Strategies

Reactive Part

- **Phase I: Application**
- **Phase II: Enactment**

Reflexive Part

- **Phase III: Reification**
- **Phase VI: Reflection**



Jesse Josua Benjamin, Christoph Kinkeldey, Claudia Müller-Birn, Tim Korjakow, and Eva-Maria Herbst. 2022. Explanation Strategies as an Empirical-Analytical Lens for Socio-Technical Contextualization of Machine Learning Interpretability. Proc. ACM Hum.-Comput. Interact. 6, GROUP.

Implications

Methodological Implications of Explanation Strategies

» The co-creation workshop is a proof of concept for studying human interpretation in context. Explanation strategies function as the ‘’what’’ that links context, explanation method and explanations.

» Explanation strategies learned from a co-design workshop could be used within explainability scenarios, which could then be probed.

Design Implications for Explanations for Non-ML Experts

» Enable combinations of explanations to support explanation strategies.

» Supplement explanations with contextual cues.

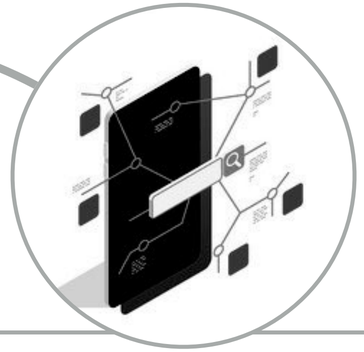
Jesse Josua Benjamin, Christoph Kinkeldey, Claudia Müller-Birn, Tim Korjakow, and Eva-Maria Herbst. 2022. Explanation Strategies as an Empirical-Analytical Lens for Socio-Technical Contextualization of Machine Learning Interpretability. Proc. ACM Hum.-Comput. Interact. 6, GROUP.

Interactive Systems

complement humans

Human-Computer Collaboration

two or more actors work together to achieve shared goals



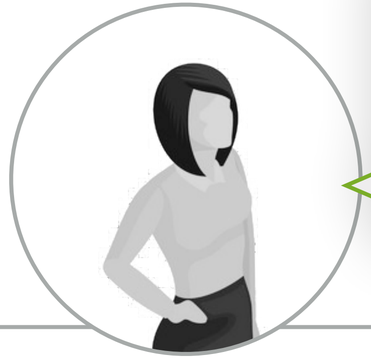
Autonomous Systems

emulate humans

Agreeing on shared goals

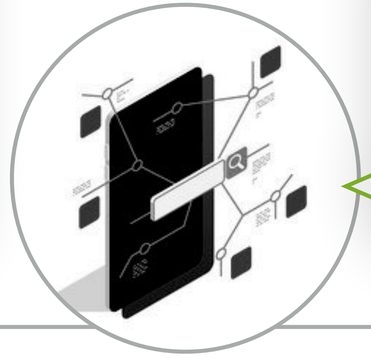
- **Allocating responsibility**
- **Coordinating activities**
- **Sharing context**
- **Learning mutually**

Study Context: Data Sharing for Personalized Medicine



Patients have privacy concerns

Sharing Medical Data



Medical researchers need ‘more’ data

ML-Based System in the Medical Domain

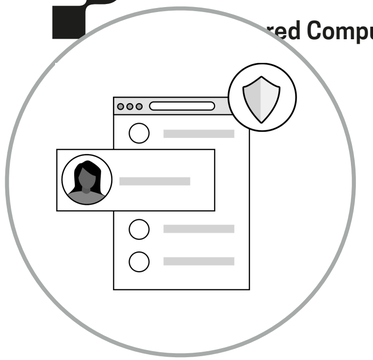
Privacy-Preserving Computation

Security researchers explain that more privacy might impact the utility of the ML model

Sörries, Peter, Müller-Birn, Claudia, et al. (2021). Privacy Needs Reflection: Conceptional Design Rationales for Privacy-Preserving Explanation User Interfaces. Mensch und Computer 2021-Workshopband.

Privacy Preserving Explanation UIs

Sörries, Peter, Müller-Birn, Claudia, et al. (2021). Privacy Needs Reflection: Conceptional Design Rationales for Privacy-Preserving Explanation User Interfaces. Mensch und Computer 2021-Workshopband.



Design Rationales

We aim to enable people to make an informed decision by having an appropriate understanding

» of the privacy-preserving technology used (privacy literacy)

Explaining the technology used for data protection

» of the potential consequences (risks) caused by sharing their data

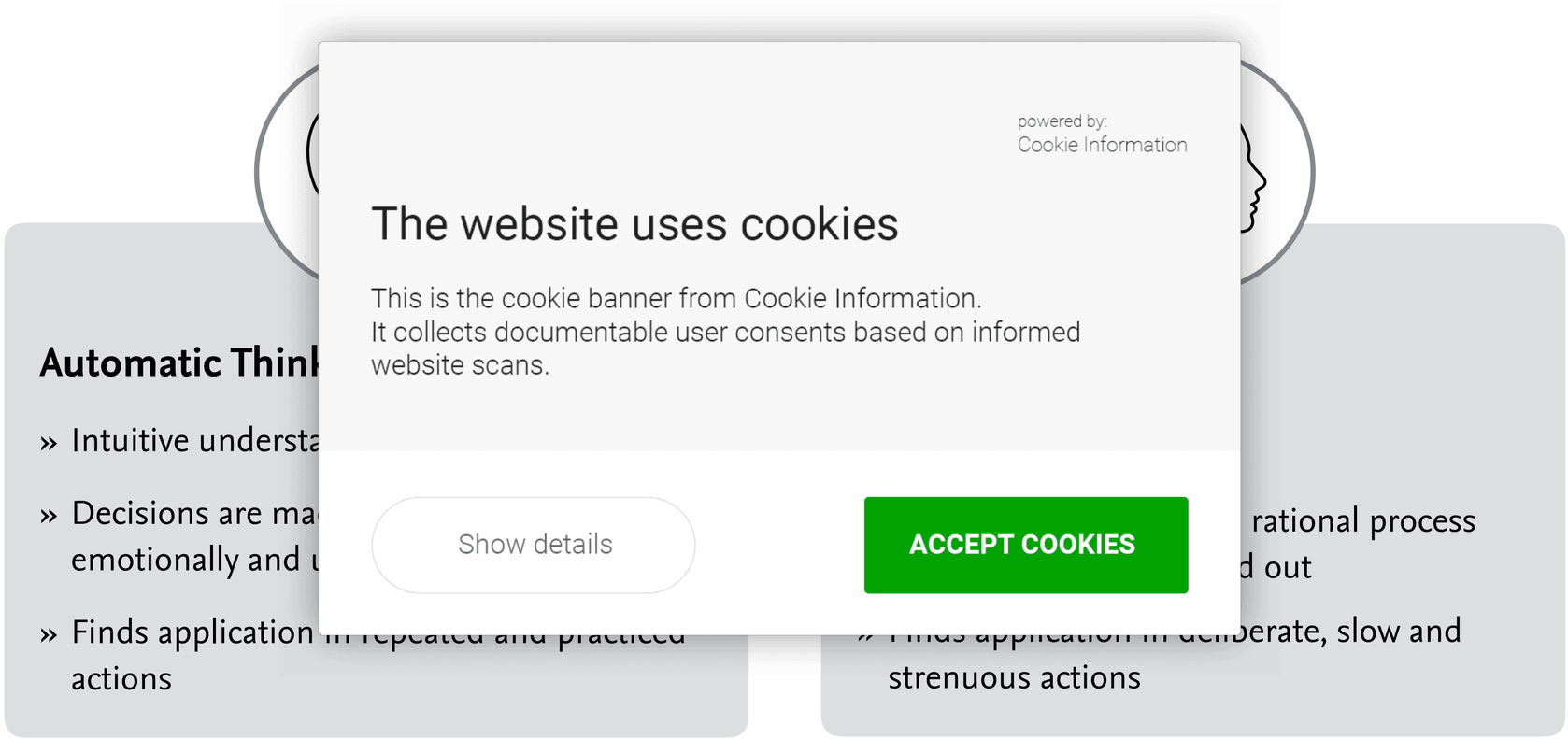
Communicating the risks of data sharing

» by supporting them to reflect critically (reflection) on their sharing decision.

Encouraging reflective thinking through meaningful decision support

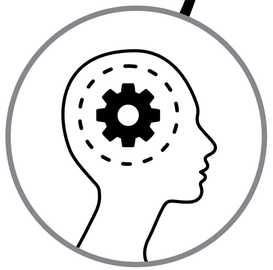
Sörries, Peter, Müller-Birn, Claudia, et al. (2021). Privacy Needs Reflection: Conceptional Design Rationales for Privacy-Preserving Explanation User Interfaces. Mensch und Computer 2021-Workshopband.

Dual Process Theory



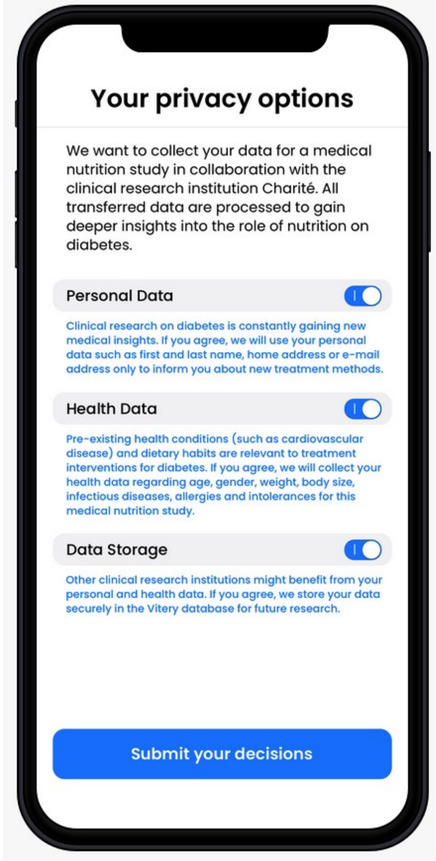
Kahneman, Daniel. Thinking, Fast and Slow. London: Penguin Books, 2012.

Study Scenario

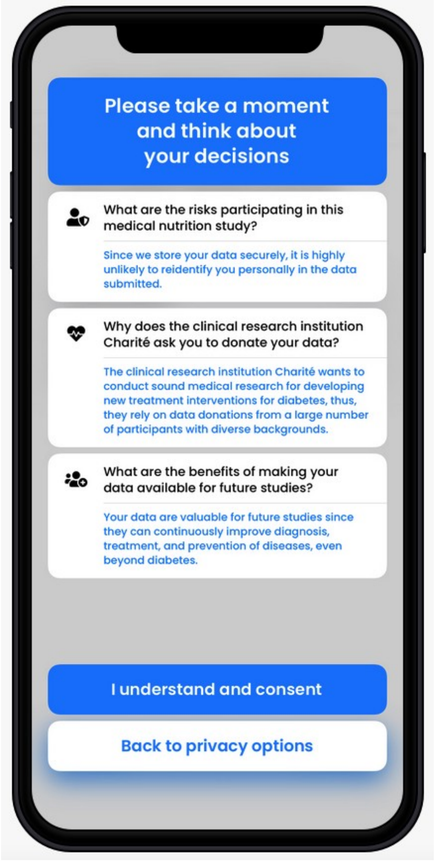


Simple Nudge

Privacy Options



Reflection Prompt



Breakdown

Inquiry

Transformation

Peter Sörries, Diane Linke, Claudia Müller-Birn. 2022. Designing for Reflection: Exploring Responsible Nudge Design for Informed Decision-Making for Data Donation in the Medical Domain. (in preparation)

Implications

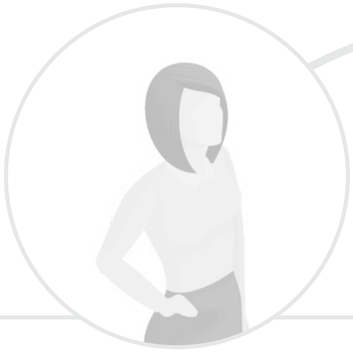
> “people greatly appreciate a sensible default”
>
> (Thaler & Sunstein 2008, S. 89)

» Designs should support reflection and transparency, which also impacts the perceived trustworthiness of the UI provider (see also Paunov et al., 2019)

» Design should consider individual characteristics, for example, people with a high need for cognition prefer complex user interfaces (**“no one fits all UI”**)

» Designs should allow people to learn, thus, we should understand our designs as enablers.

Peter Sörries, Diane Linke, Claudia Müller-Birn. 2022. Designing for Reflection: Exploring Responsible Nudge Design for Informed Decision-Making for Data Donation in the Medical Domain. (in preparation)



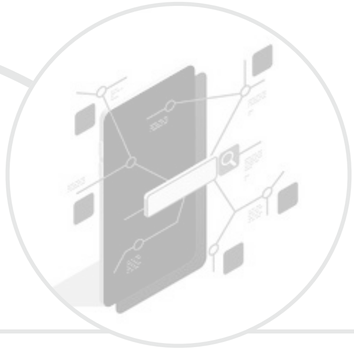
Interactive Systems

complement humans

Human-Computer Collaboration

Questions and Comments

- Agreeing on shared goals
- Allocating responsibility
- Coordinating activities
- Sharing context
- Learning mutually



Autonomous Systems

emulate humans

CONNECT clmbirn