



AI & Diversity - Digital Gender Equality

California State Visit 2022

Professor Sabiha Ghellal (she her hers) Ph.D.

June 21st 2022

About Me

- I am a HCI researcher. In my research, I focus on creating and investigating new interactive designs to **support future cultural and creative experiences**, emphasizing digital **games, mixed reality applications**, and transmedia storytelling.
- Based on the concept of **research through design**, I investigate "experience and game design" to derive design patterns, frameworks, and evaluation methods based on different **artifacts** (e.g., prototypes and workshops).



About the Study Path Mobile Media

- **Mobile media** is an interdisciplinary study path that fosters **development and design competence**.
- Alumni typically work as experience designers or developers, inclusive design experts, or in the games industry.
- It is also a study path in which a lot of research is done:
 - Accessibility (Inclusive Design)
 - Autonomous Driving
 - Serious Games (Museum and in schools)
 - Exergames
- **Minor Program for Internationals:** App Design & Development

Team Research Through Design & Ph.D. Research



Michael Möller
EDUTAIN ME

A blended Transmedia learning
experience in the classroom



Axel Braun

Mixed Reality Research
Exergaming & Autonomous
driving



Celina Retz

Exergames - Movement Science

**Ph.D. Student University
Stuttgart (1. Year)**

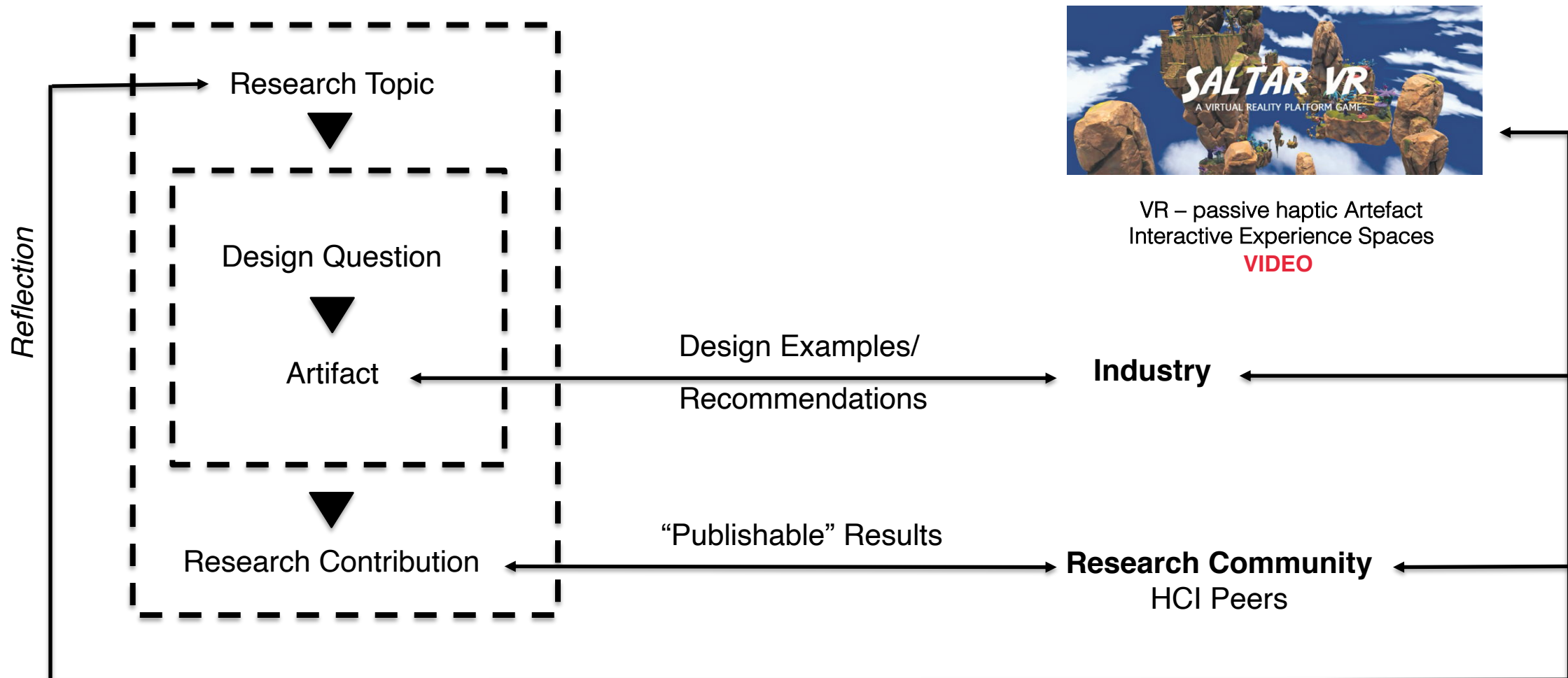


Tobias Schneider

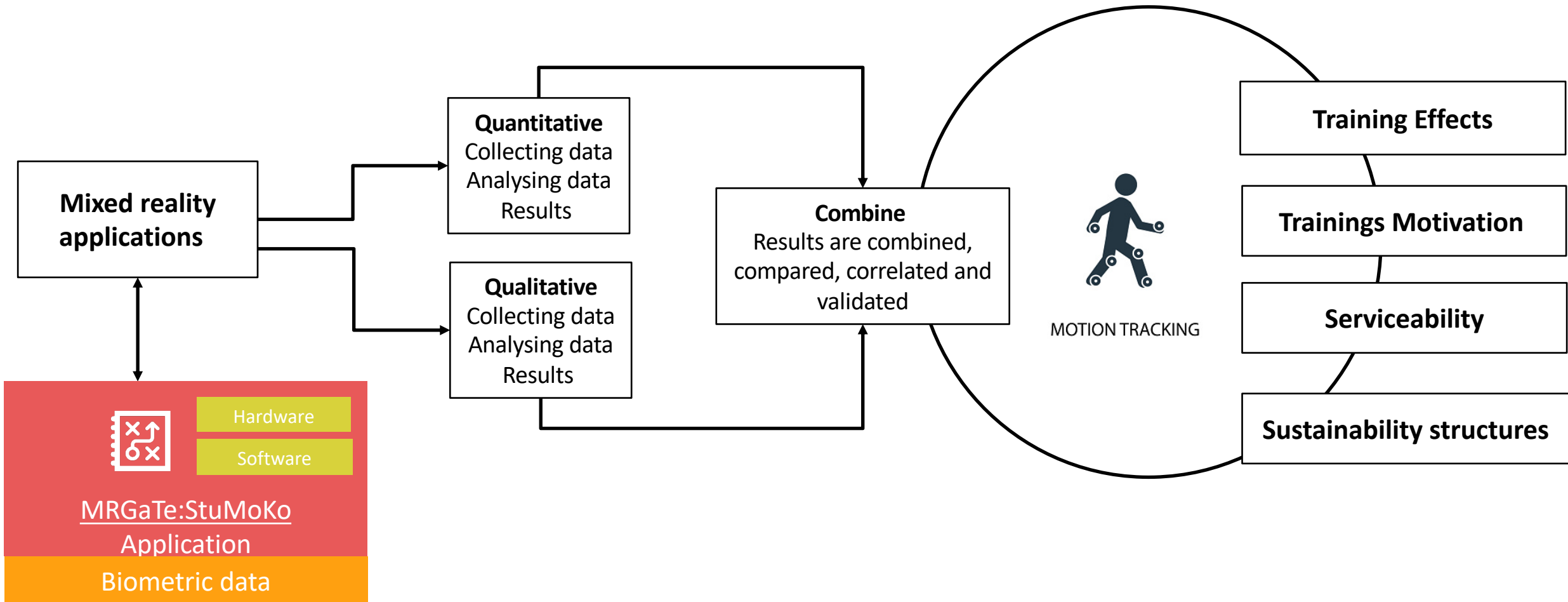
Autonomous Driving XD

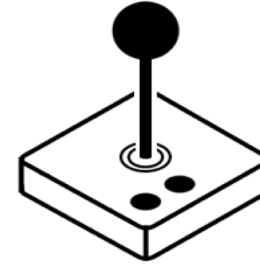
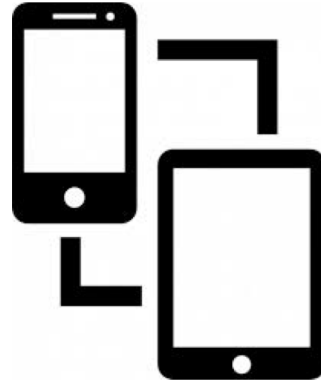
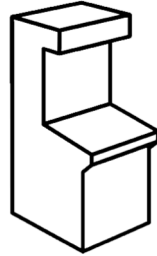
**Ph.D. Student Glasgow School
of Art (3. Year)**

Research Through Design (RtD)



PhD Research: Fall Prevention & MR Interventions





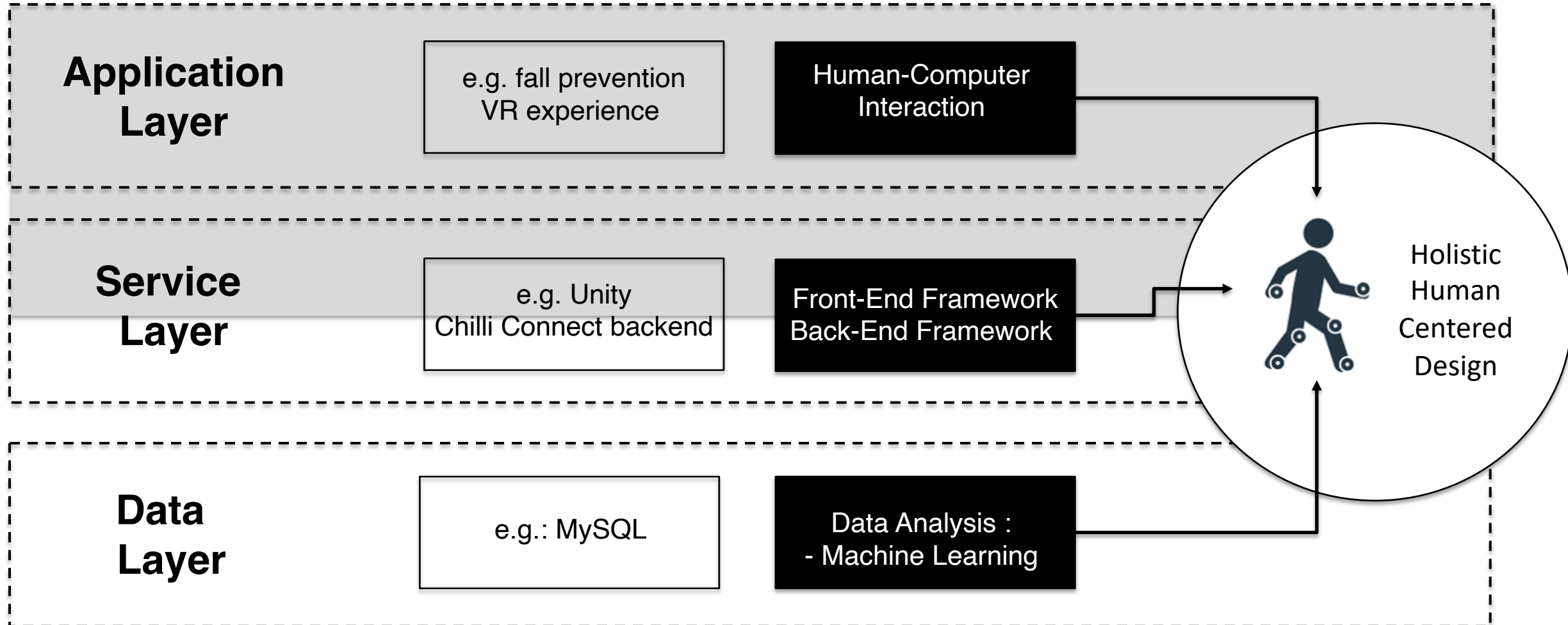
AI & Diversity - Digital Gender Equality

Sabiha Ghellal Ph.D.

Prof. Experience & Game Design

Hochschule der Medien Stuttgart 21.06.22

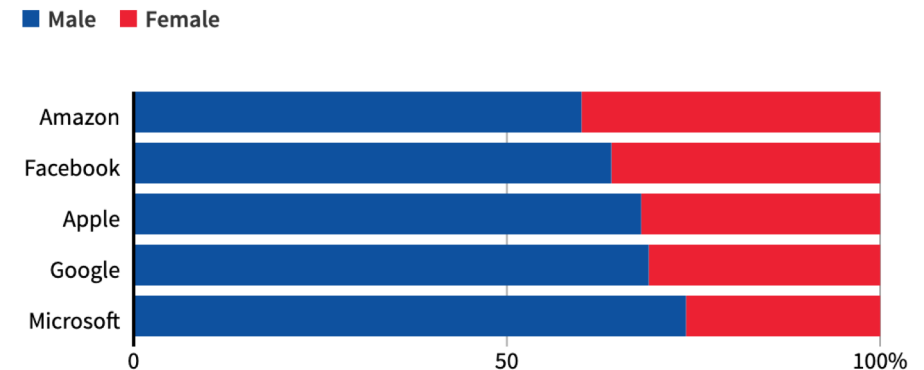
Human Centered AI – My New Perspective



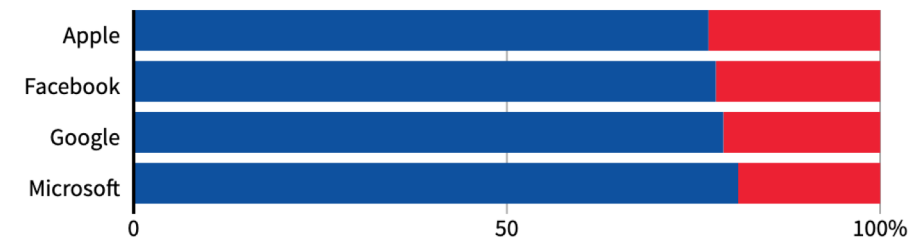
Famous Example 1: Diversity & AI

- **SAN FRANCISCO 2018 (Reuters) - Amazon.com Inc's AMZN.O** machine-learning specialists uncovered a big problem:
 - **their new recruiting engine excluded women.**
- Amazon's computer models were trained to vet applicants by observing patterns in resumes submitted to the company over 10 years.
- **Most came from men, a reflection of male dominance across the tech industry.**
- Amazon's experimental recruiting engine followed the same pattern, learning to penalize resumes including the word "women's" until the company discovered the problem.

GLOBAL HEADCOUNT



EMPLOYEES IN TECHNICAL ROLES



Ref: By Han Huang | REUTERS GRAPHICS 2017

Famous Example 2 : Diversity & AI

- **WIRED MAGAZINE 2019:** The Apple Credit card, launched in August 2019, ran into major **problems**. Users noticed:
 - It seemed to **offer smaller lines of credit to women than to men**.
- The response from Apple just added confusion and suspicion. No one from the company seemed **able to describe how the algorithm even worked**, let alone justify its output.
- A gender-blind algorithm could end up biased against women as long as it's **drawing on any input or inputs that happen to correlate with gender**.
- This is referred to as **“Proxy Discrimination”**.



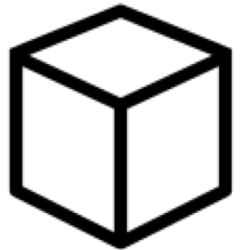
Discrimination by Proxy

An algorithm can have an adverse effect on vulnerable populations even without explicitly including protected characteristics. This often occurs when a model includes features that are correlated with these characteristics.

PLANNING
BUILDING
DEPLOYING
MONITORING

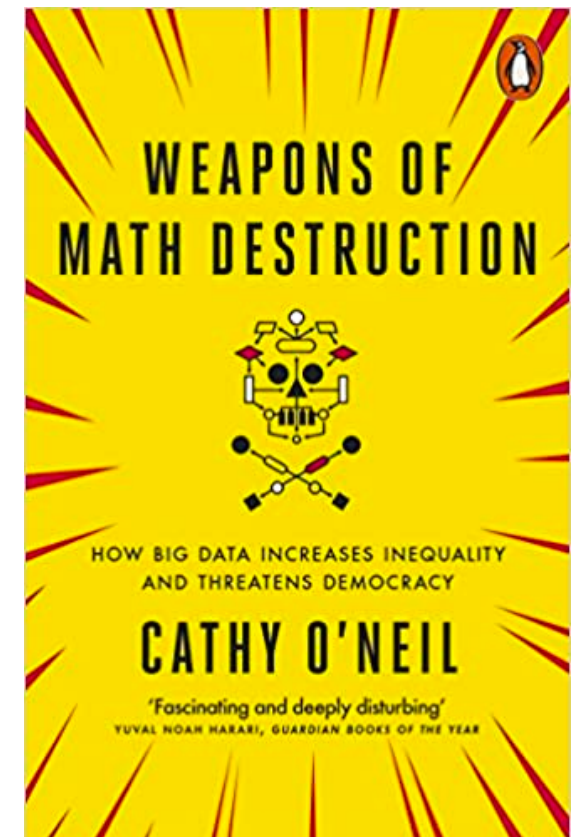
Transparency & Explainable AI (XAI)

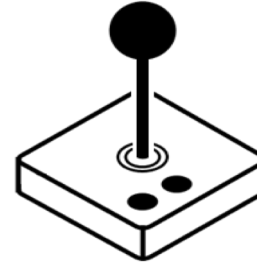
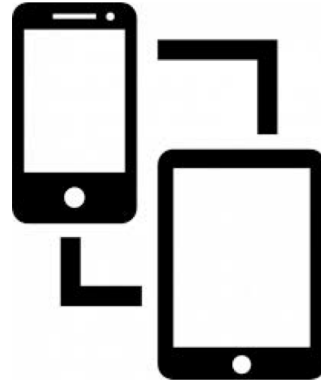
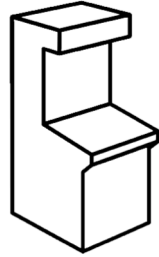
- The algorithms used in AI can be differentiated into white-box and black-box machine learning (ML) algorithms /solutions.
 - **White-box** models are ML models that provide results that are understandable for experts in the domain.
 - **Black-box** models, on the other hand, are extremely hard to explain and can hardly be understood even by domain experts.
- **Explainable AI (XAI)** describes ML results that can be understood by humans (not experts).
- XAI aims to explain what has been done, what is done right now, what will be done next, and unveil the information the actions are based on.



Discrimination by Proxy

- Proxy discrimination occurs when an apparently neutral characteristic is used as a proxy for a prohibited characteristic.
- In the past, companies have engaged in proxy discrimination to circumvent anti-discrimination laws.
- However, proxy discrimination may be unintentional e.g. „**web scraping**“ (collecting data from various webpages) may lead to unintentional proxy discrimination.
- **Claiming that „web scraping“ is unbaited because the data “does not lie” and does not “actively” discriminate is untrue.**



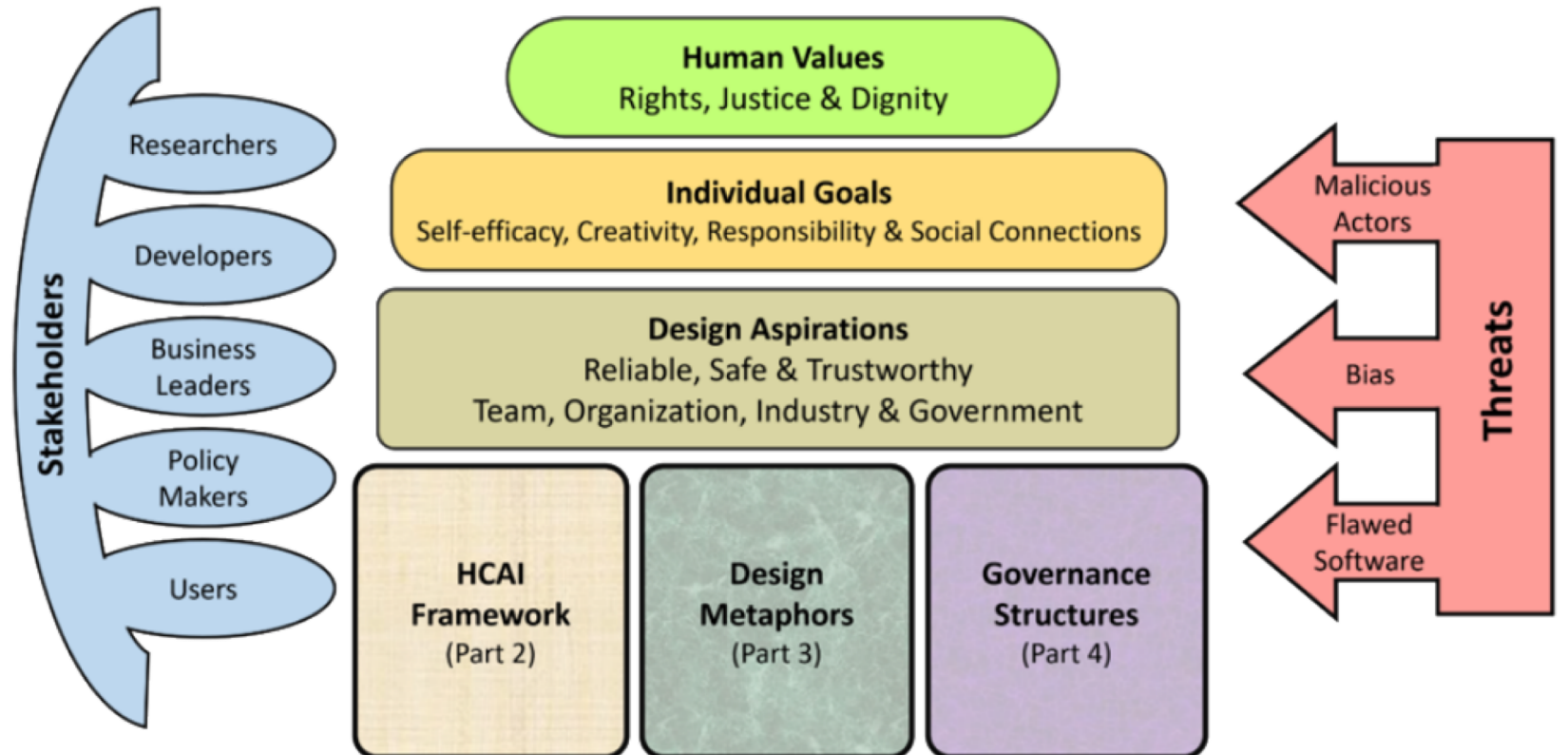


AI Diversity – Possible Solutions?

Sabiha Ghellal Ph.D

Hochschule der Medien Stuttgart

New Book from Ben Shneiderman



Research problems in HCI ^[1]

Empirical

Definition: Empirical research is aimed at creating or elaborating descriptions of real-world phenomena related to human use of computing.

Laudan ^[2] cites three characteristic subtypes:

- **Unknown phenomena**
- **Unknown factors**
- **Unknown Effects**

Conceptual

Definition: Work on a conceptual research problem is aimed at explaining previously unconnected phenomena occurring in interaction. Responses to this type of problem include theories, concepts, methods, principles, and models. Furthermore, Laudan ^[2] distinguishes among three characteristic subtypes:

- **Implausibility**
- **Inconsistency**
- **Incompatibility**

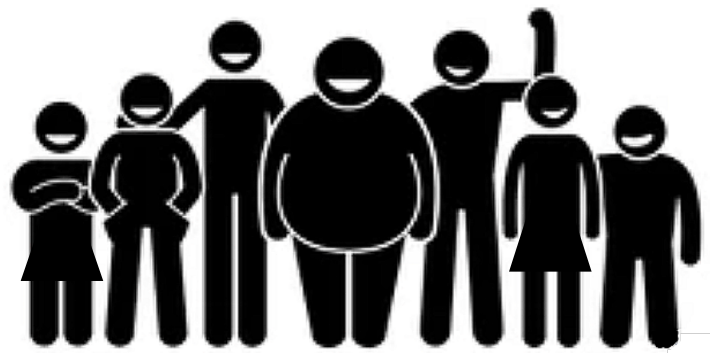
Constructive

Definition: Constructive research is aimed at producing understanding about the construction of an interactive artefact for some purpose in human use of computing. **Importantly, this problem type cuts across design and engineering.** Oulasvirta & Hornbæk ^[1] further distinguish three subtypes:

- **No known solution**
- **Partial solution**
- **Inability to deploy or implement**

Definition Ben Schneidermann

- A new synthesis that integrates AI technologies with HCI approaches to amplify, augment, empower, and enhance human abilities.
- The goal is to increase human performance, strengthen self-efficacy, support creativity, and promote social connections.
- This new synthesis recognizes human responsibility (legally and morally) for actions taken by the technologies they choose to use.
- HCAI researchers, designers, developers, and managers consider human values, rights, justice, and dignity. They seek to build reliable, safe, and trustworthy systems, then evaluate their usage so as to continuously improve them.



STAKEHOLDER	RESEARCHERS
	DEVELOPERS
	BUSINESS LEADERS
	POLICY MAKERS
	USERS

HUMAN VALUES
RIGHTS, JUSTICE & DIGNITY

INDIVIDUAL GOALS
SELF- EFFICACY, CREATIVITY, RESPONSIBILITY &
SOCIAL CONNECTION

DESIGN ASPIRATIONS
RELIABLE, SAFE & ACCURATE, TRUSTWORTHY
TEAM ORGANIZATION, INDUSTRY & GOVERNMENT

MALICIOUS ACTORS

BIAS

FLAWED SOFTWARE

THREAT

Ref: Ben Shneiderman (2022) Human-Centered AI ISBN: 9780192845290

HCAI
FRAMEWORK

DESIGN
METAPHORS

GOVERNANCE
STRUCTURES

Conversation about AI

- The conversation about **AI and ethics** needs to be open up wide open.
- We need to understand and address the subtleties and life cycle of AI systems and their impacts at each stage.
- Too often, conversations focus solely on the **development and deployment stages** of the life cycle, although many of the problems occur during **the earlier stages of conceptualization, research, and design**.
- Automated solutions will have to **consider humanity and ethics** next to mathematical possibilities.

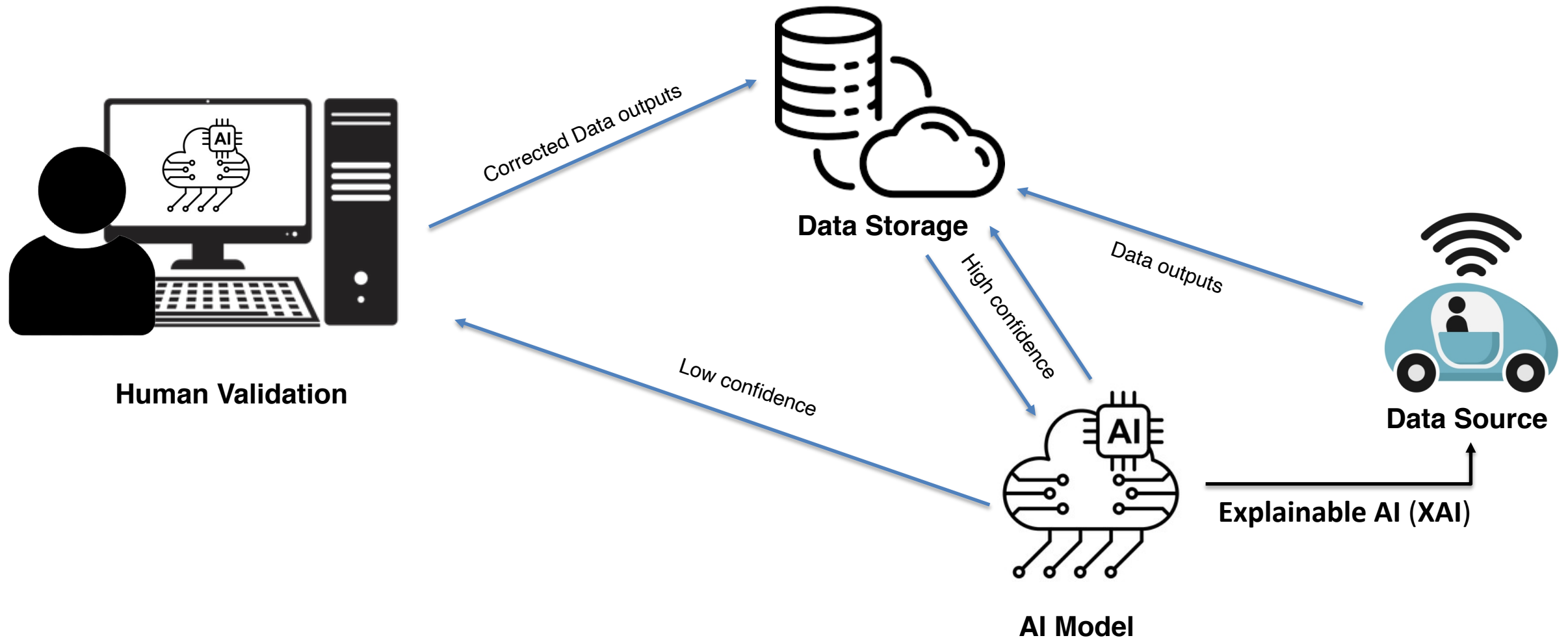
AI Diversity – Path Forward

1. **Self-Reflection** (80 % of my research is based on data produced by male researchers – In hardly any of my research did I personally consider that- Why?).
2. **Ensure Transparency** (Explainable AI).
3. **Humanity in the Loop!** Activate the Human in the Loop for Diversity!
4. Produce **diversity algorithms** that searches, highlights, and removes potentially biased data.
5. Keep the **discussion open** as research and development continues.



HUMAN IN THE LOOP

Definition: The integration of human workforce in the AI pipeline in order to train and validate models in a continuous way

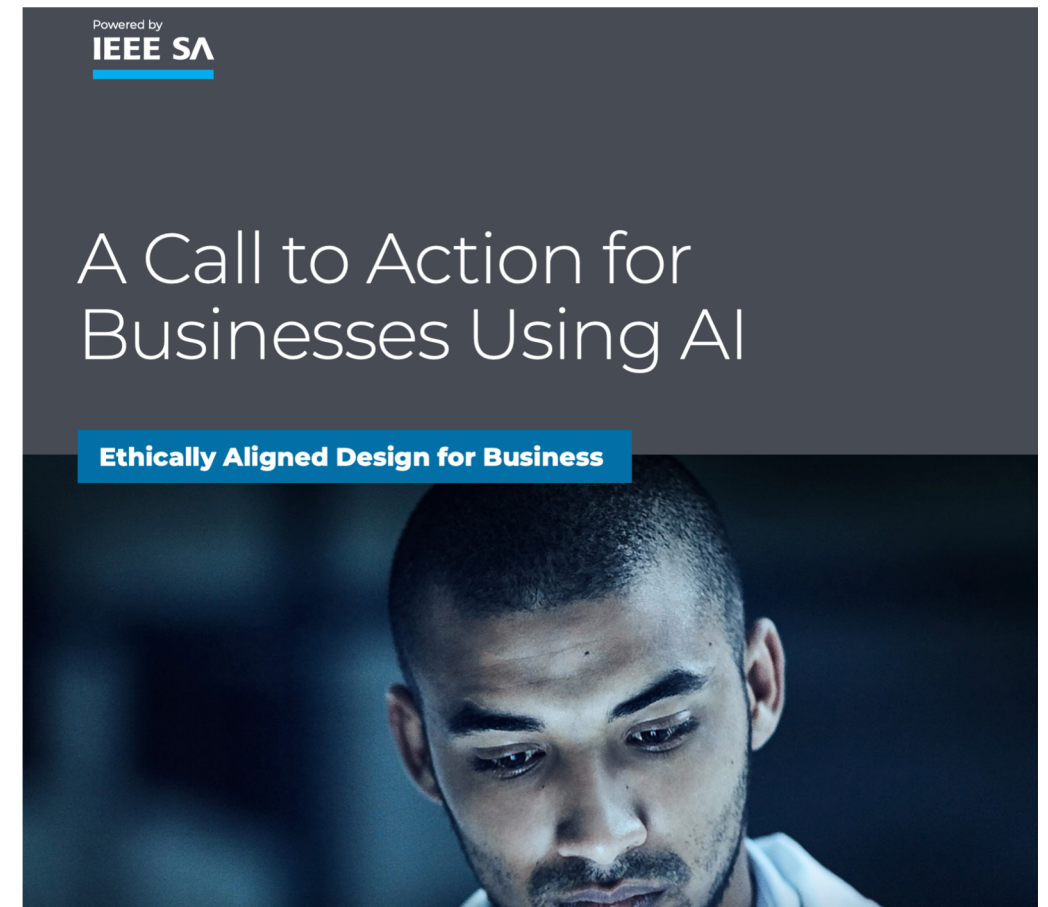


IEEE Ethically Aligned Design Report

The IEEE Ethically aligned Design Report is based on three-year research involving more than 200 people. It offers clear statements grouped into the following eight principles:

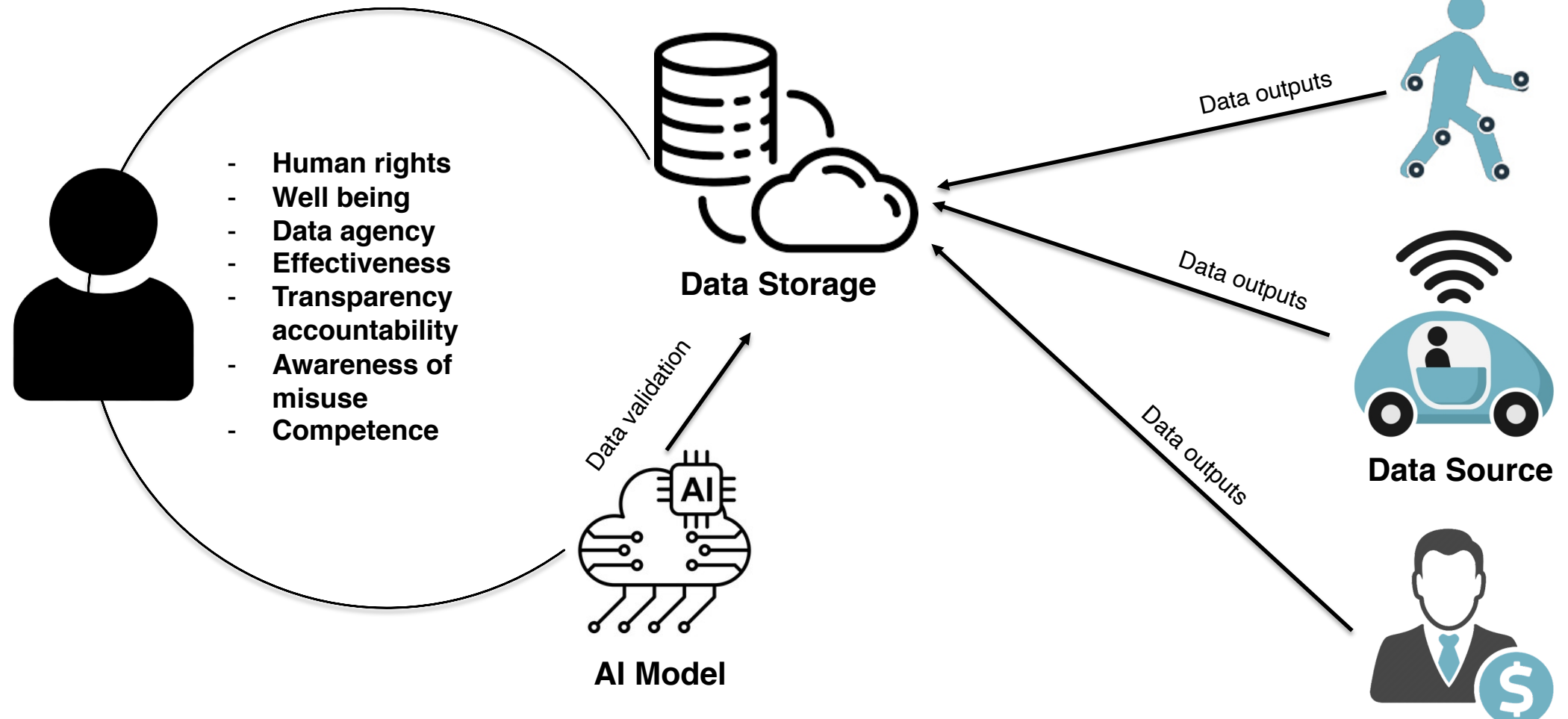
- **Human rights**
- **Well being**
- **Data agency**
- **Effectiveness**
- **Transparency**
- **accountability**
- **Awareness of misuse**
- **Competence**

Ref: <https://ethicsinaction.ieee.org/wp-content/uploads/ead-for-business.pdf>



Suggestion: Humanity in the Loop

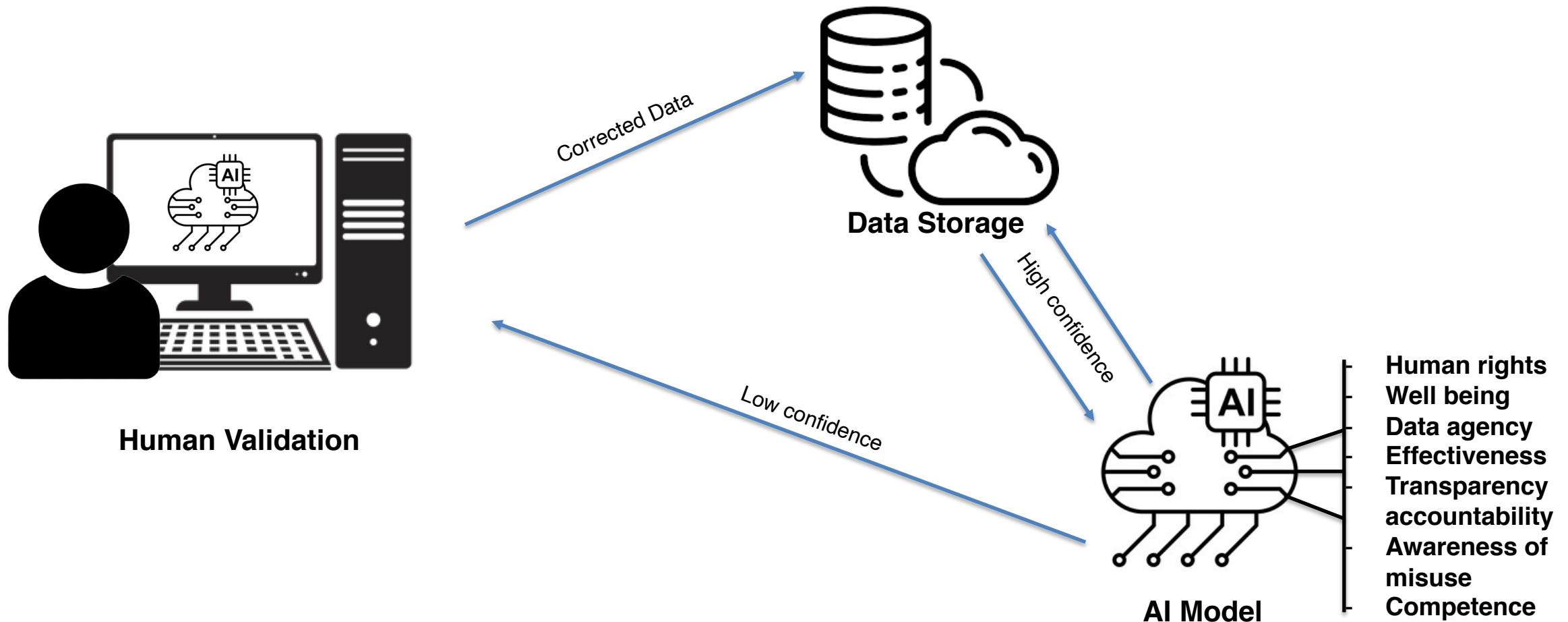
Definition: The integration of humanity (principles) in the AI pipeline in order to train and validate AI models in a continuous way.



Work on a Humanity Algorithm?

Question: Is all available Data corrupt? How can we ensure we are using unbiased data?

Suggestion: Let's work on Humanity Algorithms that will validate existing AI models.



ADDITIONAL LINKS



DIVERSITY.AI
Inclusion | Balance | Neutrality

<https://diversity.ai/>

 Artificial Intelligence
Index Report 2021

**CHAPTER 6:
DIVERSITY IN AI**

https://aiindex.stanford.edu/wp-content/uploads/2021/03/2021-AI-Index-Report-_Chapter-6.pdf

**MIT
Technology
Review**

[Featured](#) [Topics](#) [Newsletters](#) [Events](#) [Podcasts](#)

[Sign in](#) 

THE BIG STORY: OPINION

How our data encodes systematic racism

Technologists must take responsibility for the toxic ideologies that our data sets and algorithms reflect.

<https://www.technologyreview.com/2020/12/10/1013617/racism-data-science-artificial-intelligence-ai-opinion/>

Thank You!

Sabiha Ghellal Ph.D.

She.Her.Hers

Prof. Experience & Game Design

Hochschule der Medien Stuttgart

www.hdm-stuttgart.de

www.hdm-stuttgart.de/IFG

www.game.de

Contact: ghellal@hdm-stuttgart.de

