

# AI-TRUST

Mensch-KI-Systeme Interaktionsgestaltung für  
Interieurs autonomer Fahrsysteme im ÖV



## Teil 1

ProjektpartnerInnen und Schwerpunkte

## ➤ ProjektpartnerInnen und Schwerpunkte

Das Forschungsprojekt "AI Lab TRUST" untersucht die Entwicklung von Mensch-KI-Systemen im Kontext autonomer Transportsysteme.

Die Projektpartner, das **Center for Cognition and Computation** an der Goethe Universität Frankfurt und das **Designinstitut für Mobilität und Logistik** der Hochschule für Gestaltung Offenbach, arbeiten hierfür interdisziplinär zusammen. Ziel ist es, ein **theoretisches Modell** zu erstellen, das maschinelles Lernen und Human-Centered Design vereint, um die **Interaktion zwischen Menschen und KI-Systemen** transparent und vertrauenswürdig zu gestalten. Sie verwenden VR-Simulationen, um den Einfluss visueller Designelemente auf die Vertrauensbildung zu untersuchen und Erkenntnisse für die Entwicklung von KI-Systemen zu gewinnen.

### **HfG Offenbach:**

Prof. Peter Eckart, Dipl.-Des. Ken Rodenwaldt, Dipl.-Des. Armin Arndt, Dipl.-Des. Daniel Rese, Leonard Neunzerling, Annika Storch

### **Goethe Universität:**

Prof. Dr. Visvanathan Ramesh, Martin Klingebiel, Dr. Iuliia Pliushch, Dr. Michael Rammensee

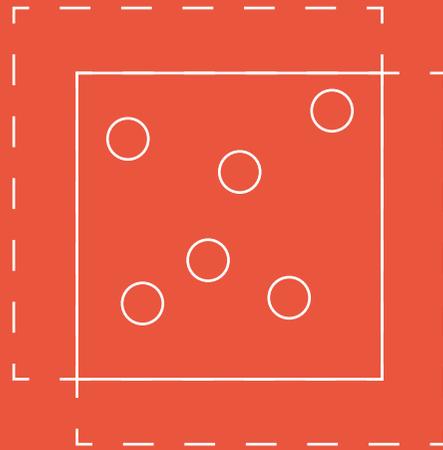


Human-Vision



Computer-Vision

AI•Lab



## Teil 2

Kontext und Forschungsziel der Studie

## ➤ Kontext und Forschungsziel der Studie

Wir verstehen die **Kommunikation zwischen autonomen Fahrzeugen und Passagieren** als entscheidend für die **Vertrauensentwicklung**. Es wurde ein Trust-Modell entwickelt, das die relevanten Aspekte dieser Kommunikation berücksichtigt. In einer VR-Simulation, die eine Bremsreaktion eines autonomen Fahrzeugs darstellt, untersuchen wir verschiedene Faktoren, darunter die **Art der vermittelten Informationen** (visuell), die **Menge der Informationen**, ihre **Position im Fahrzeug**, und den **Zeitpunkt**, zu dem diese Informationen den Passagieren mitgeteilt werden.

Das Ziel der VR-Studie ist es, herauszufinden, wie diese Faktoren das **Vertrauen der Passagiere beeinflussen** können. Insbesondere interessiert uns, ob eine zeitliche Vorankündigung von Fahrmanövern eine beruhigende Wirkung auf die Passagiere hat und ob es einen "sweet spot" gibt, in dem diese Ankündigungen am effektivsten sind. Die Ergebnisse sollen dazu beitragen, **effektive Designprinzipien** für die **Mensch-Maschine-Interaktion** in autonomen Fahrzeugen zu entwickeln und das **Vertrauen der Passagiere** in diese Technologie zu stärken.

## ➤ Kontext und Forschungsziel der Studie

Nutzung eines autonomen Fahrzeuges im ÖPNV



# ➤ Kontext und Forschungsziel der Studie

## Feldstudie in autonomem Shuttle



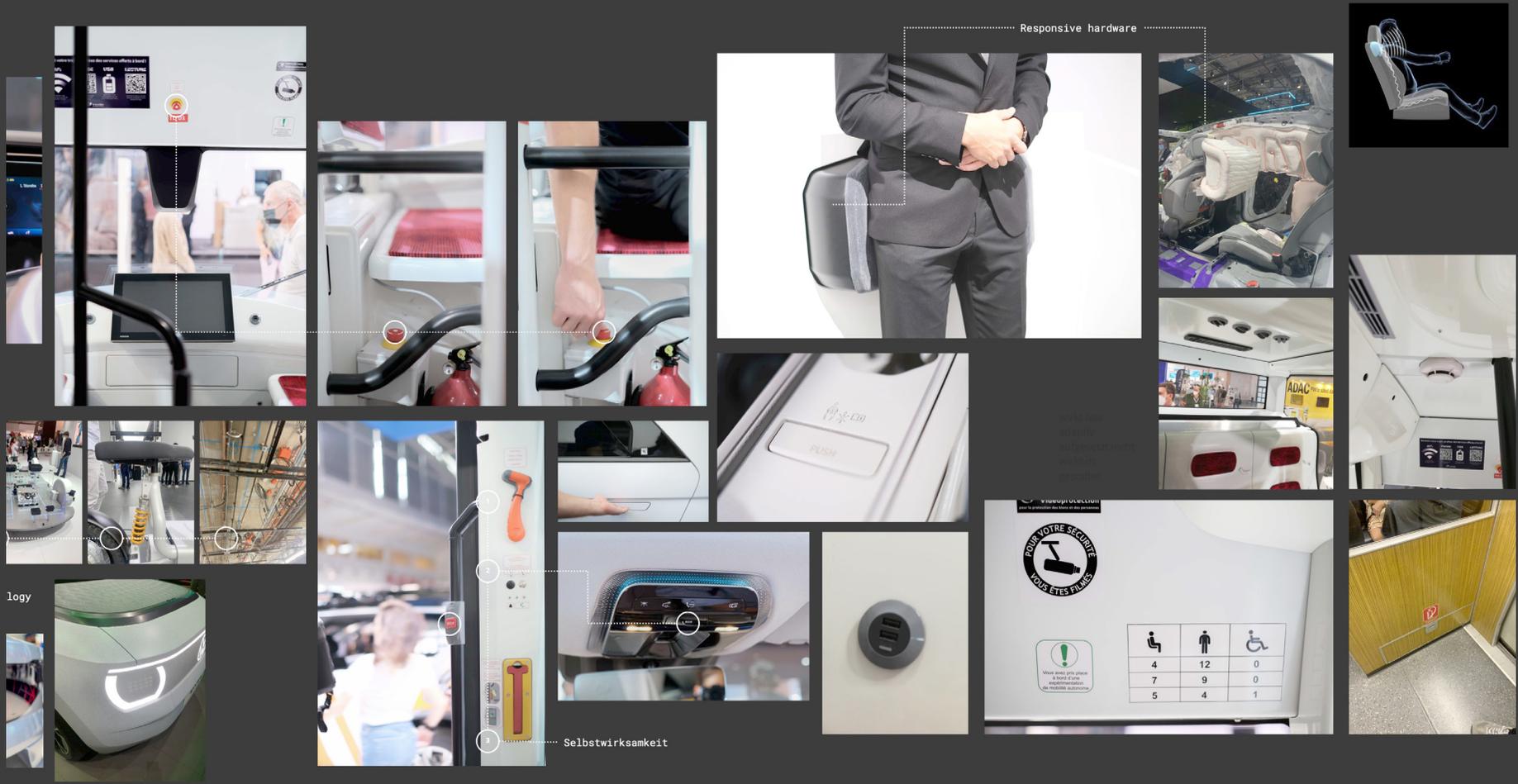
## ➤ Kontext und Forschungsziel der Studie

Besuch der IAA Mobility 2021 in München

» What will  
move us  
next

# ➤ Kontext und Forschungsziel der Studie

## Recherche zu vertrauensrelevanten Artefakten



## ➤ Kontext und Forschungsziel der Studie

Welche Faktoren machen Systeme vertrauenswürdig?



## ➤ Kontext und Forschungsziel der Studie

Was kann KI als Tool leisten, um Vertrauen im ÖPNV zu erhöhen?

KI kann gleichzeitig **fahren** und **kommunizieren**.

KI kann gleichzeitig **Gefahren** antizipieren und **kommunizieren**.

Zeitlicher Vorteil = Vertrauensgewinn?



### Teil 3

Versuchsaufbau

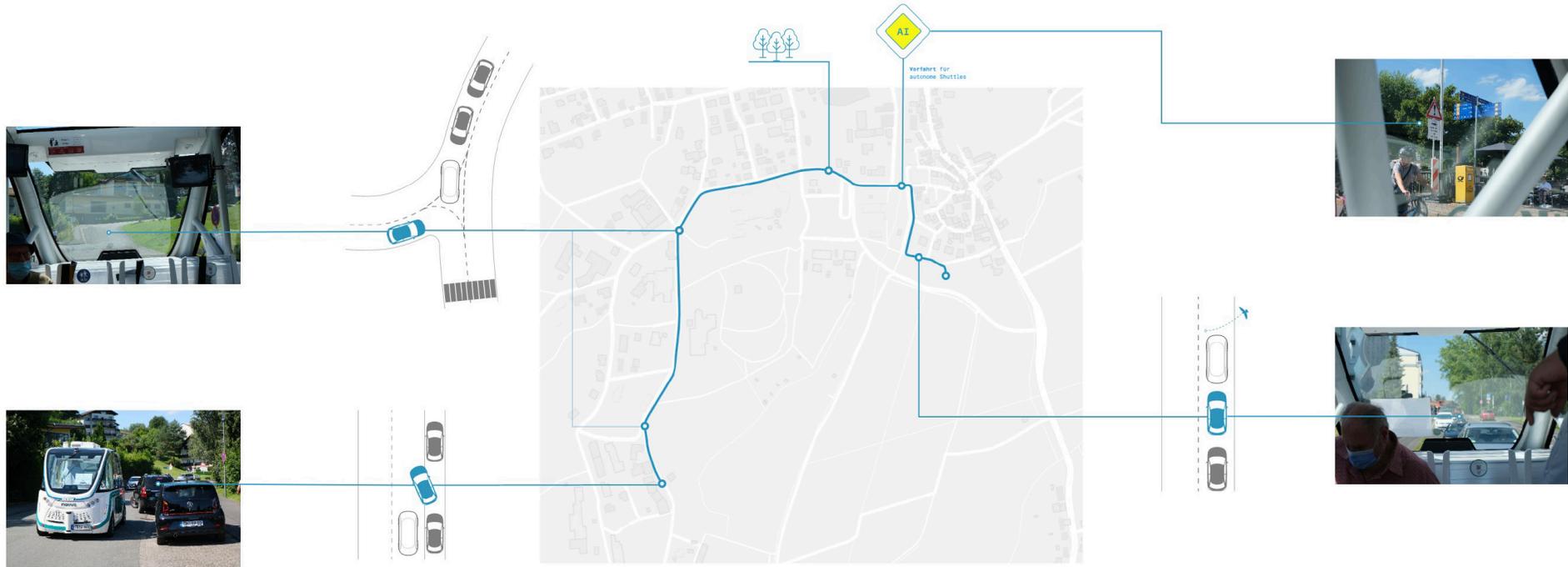
## ↗ Versuchsaufbau

Es wurde ein **Mixed-Reality-Testbed** aufgebaut, das das Mitfahren im **Interior eines autonomen Fahrzeugs** und die Fahrt durch einen **virtuellen Stadtverkehr** für Proband:innen erlebbar macht. An bestimmten Abzweigungen wurde eine vordefinierte Situation programmiert, in der dem Fahrzeug durch ein anderes Fahrzeug die Vorfahrt genommen wird. In der Simulation wurden anhand spezifischer Gestaltungsprinzipien vier **verschiedene Zeitpunkte einer Alarmierung (Pre-Notification)** getestet.

Für die Pre-Notification eines Bremsvorgangs wurden vier Zeitpunkte für ein visuelles Alarmsignal bis zum Stillstand des Fahrzeugs (**0 sek., 0.3 sek., 0.6 sek., 0.9 sek.**) miteinander verglichen. Anhand von Fragen, die in Bezug auf die **Gestaltungsprinzipien** der **ISO-Norm 9241** ausgearbeitet wurden, wurde die Einschätzung des Vertrauens vor und nach dem Fahrerlebnis verglichen. Um Informationen über die situative Empfindung der ProbandInnen während der verschiedenen Bremsvorgänge selbst zu sammeln, wurden zudem **biometrische Daten** wie **Atmung, Herzfrequenz und Hautwiderstand** erhoben.

# ➤ Versuchsaufbau

Festlegung auf ein spezifisches Szenario



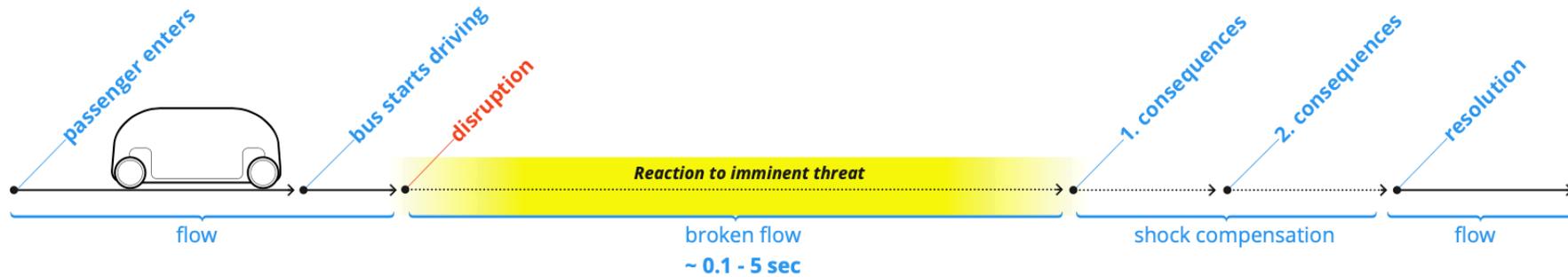
# ➤ Versuchsaufbau

Festlegung auf ein spezifisches Szenario

Context of system behaviour

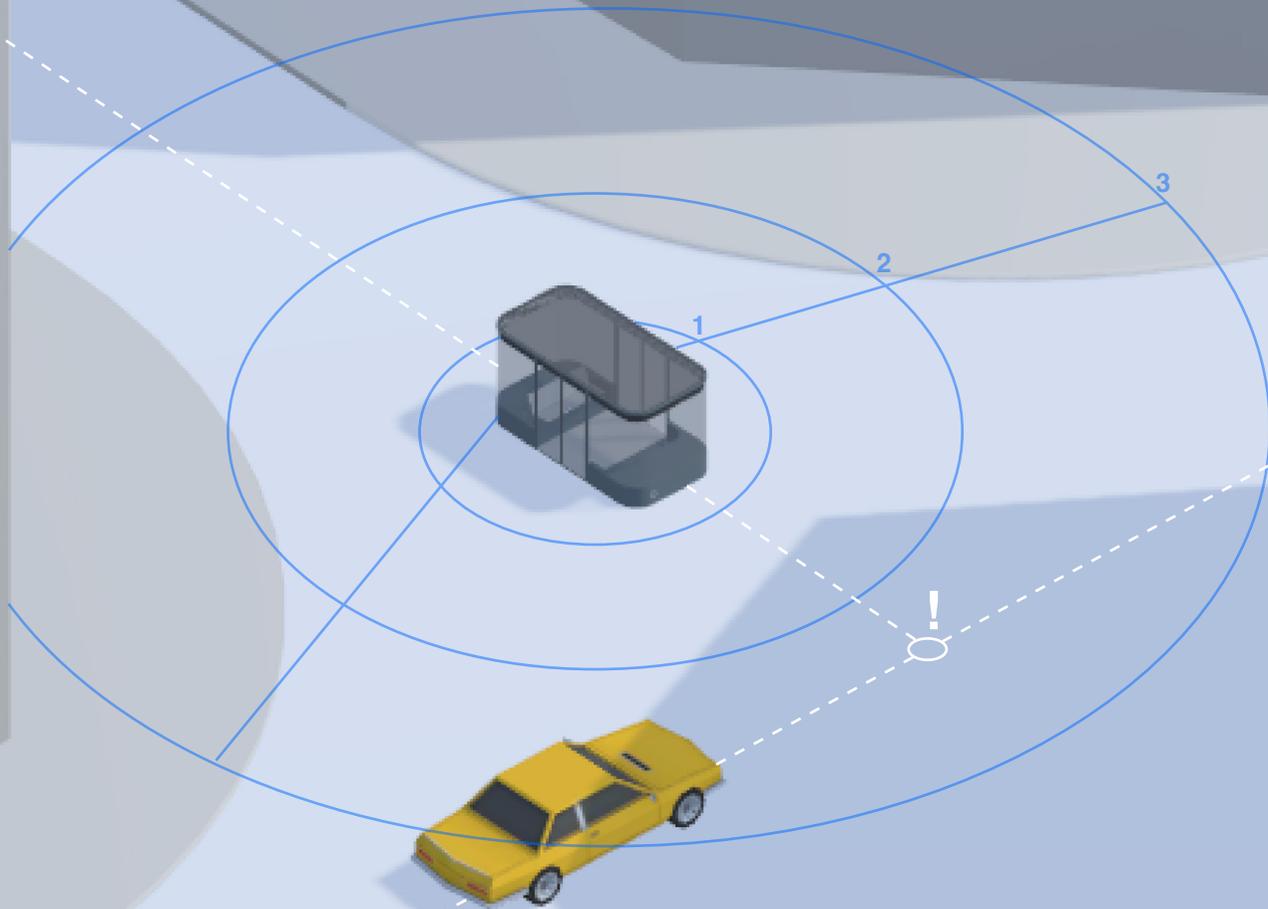


Timeframe in which context takes place



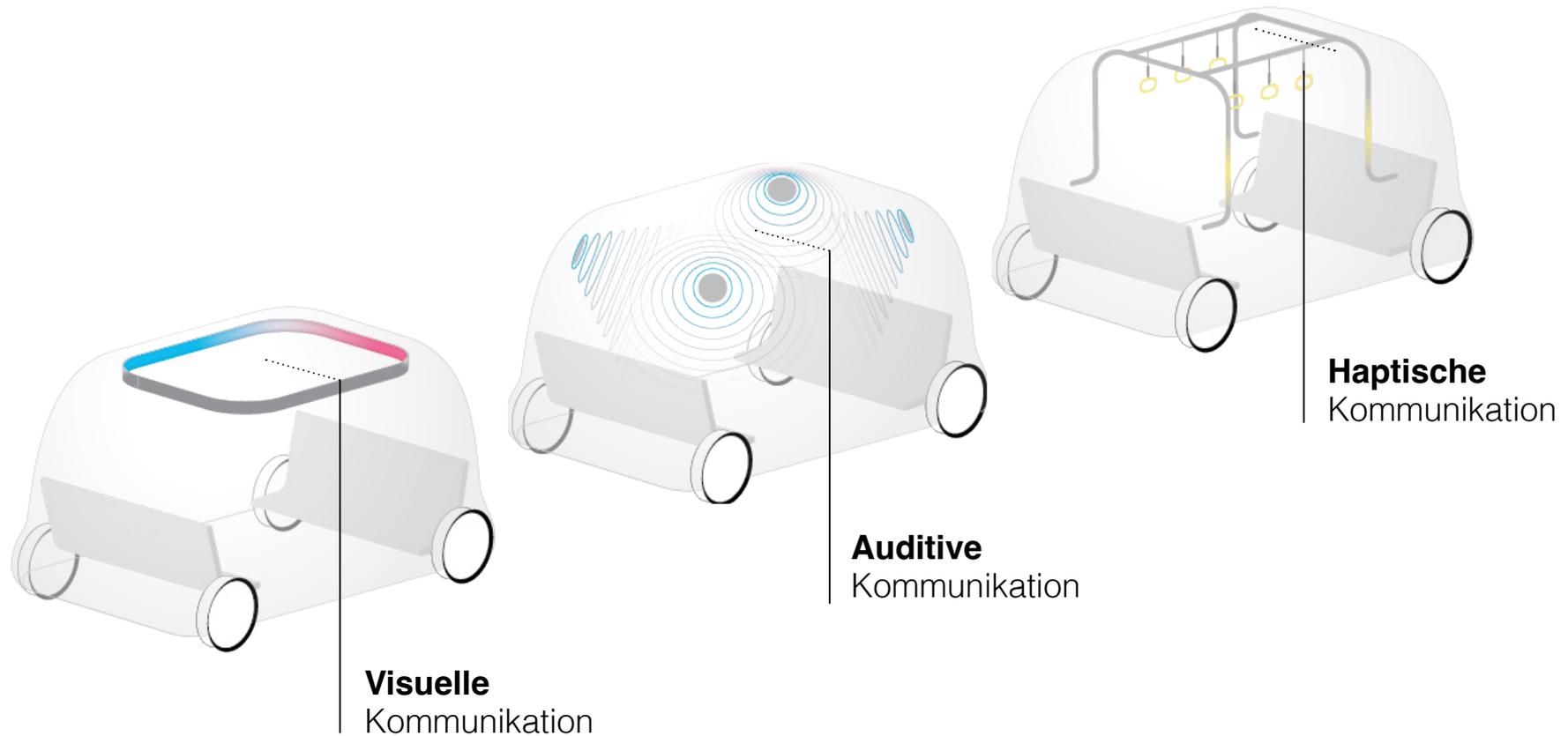
## ↗ Versuchsaufbau

Festlegung auf ein spezifisches Szenario



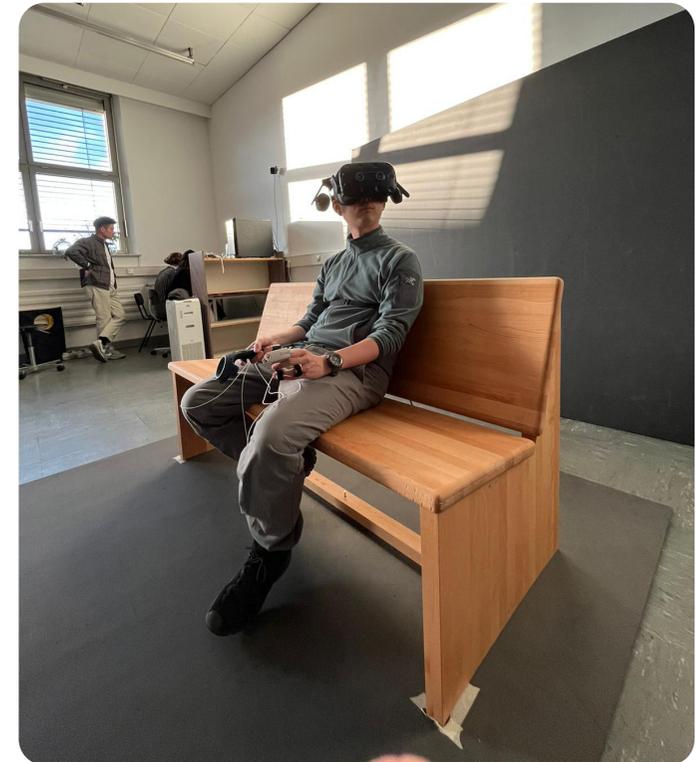
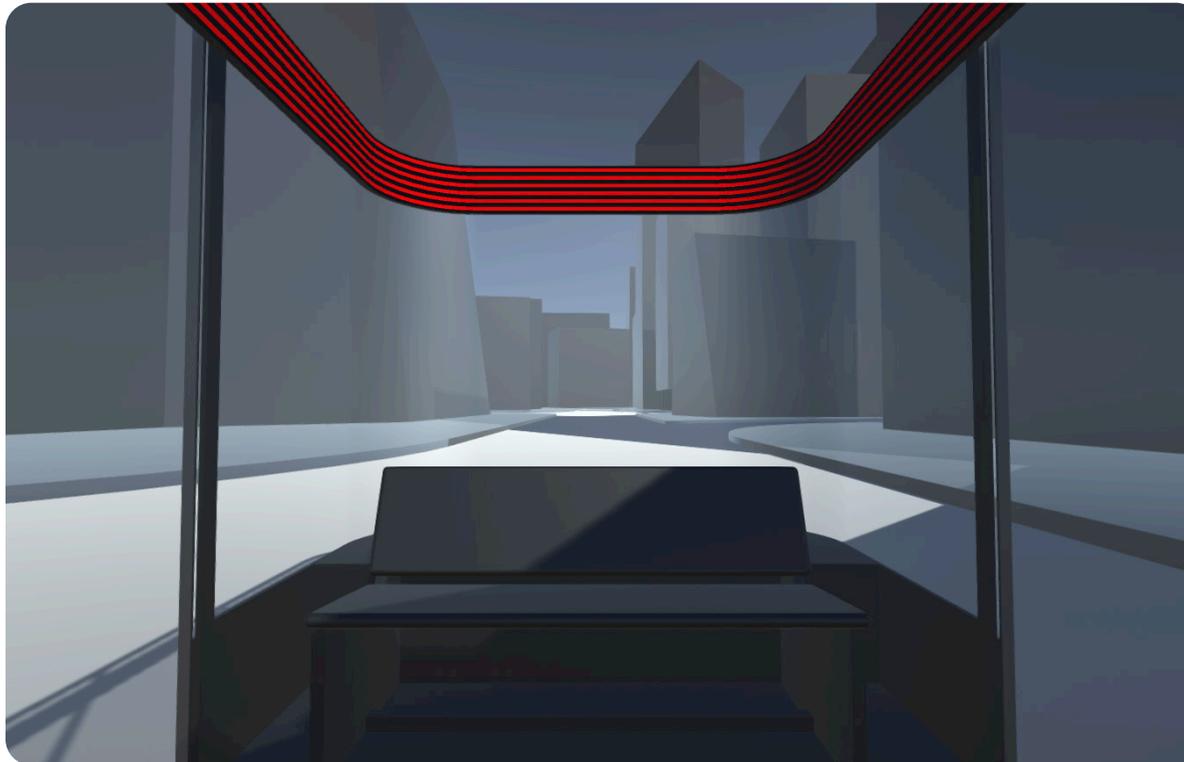
## ➤ Versuchsaufbau

Designvorschlag - Entscheidung für das Lichtband



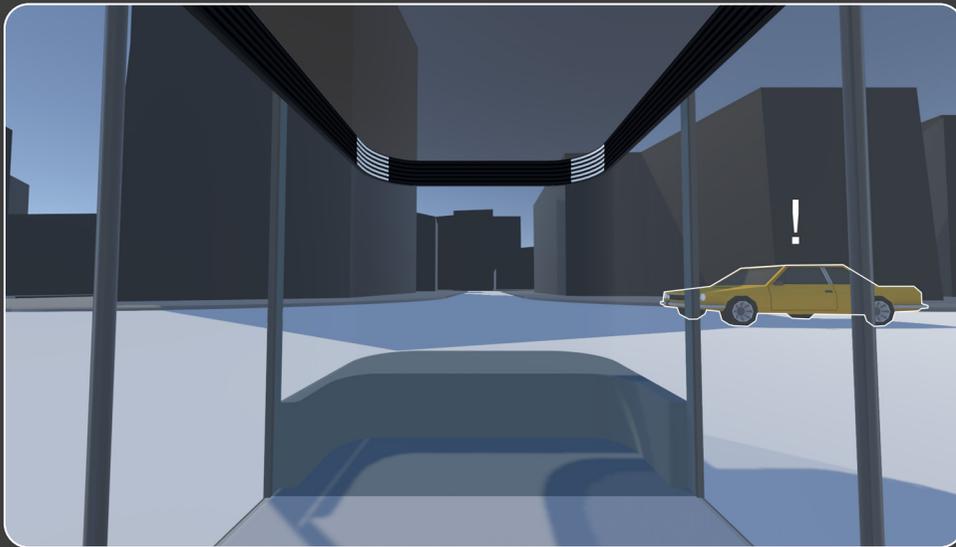
## ➤ Versuchsaufbau

Designvorschlag - Entscheidung für das Lichtband



## ↗ Versuchsaufbau

Designvorschlag - Entscheidung für das Lichtband



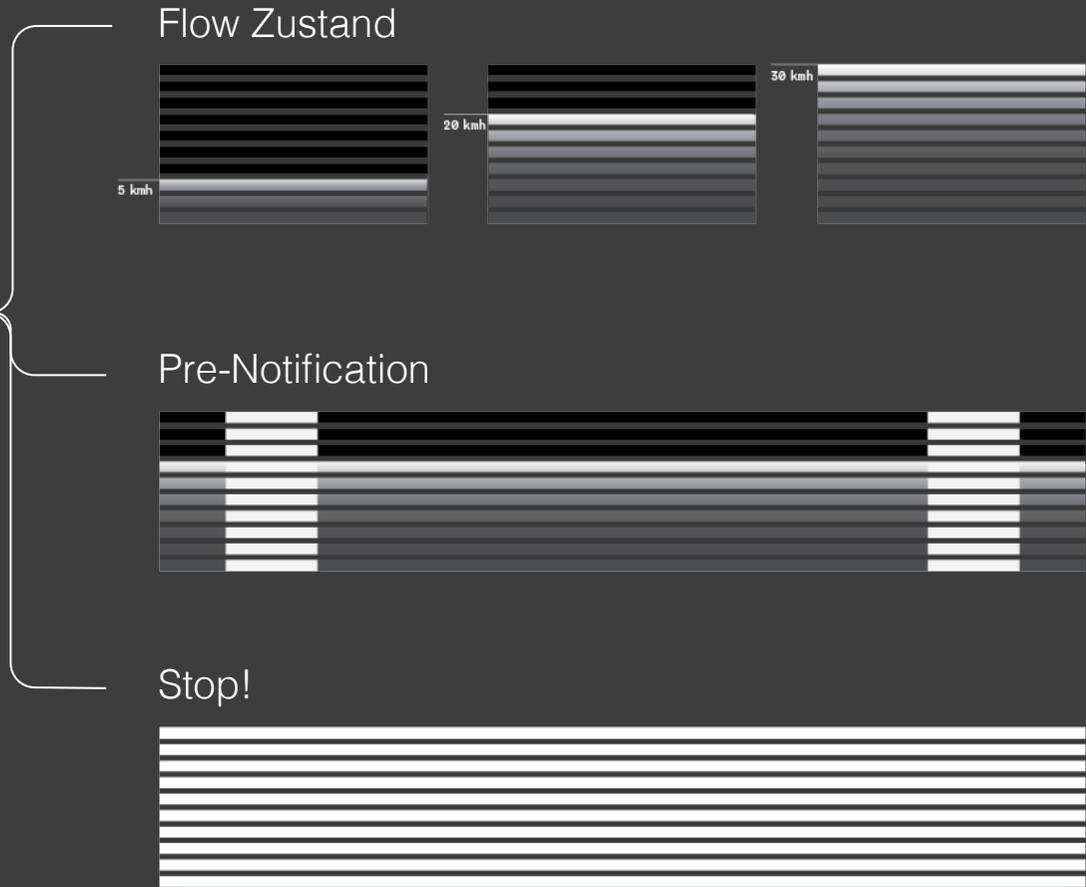
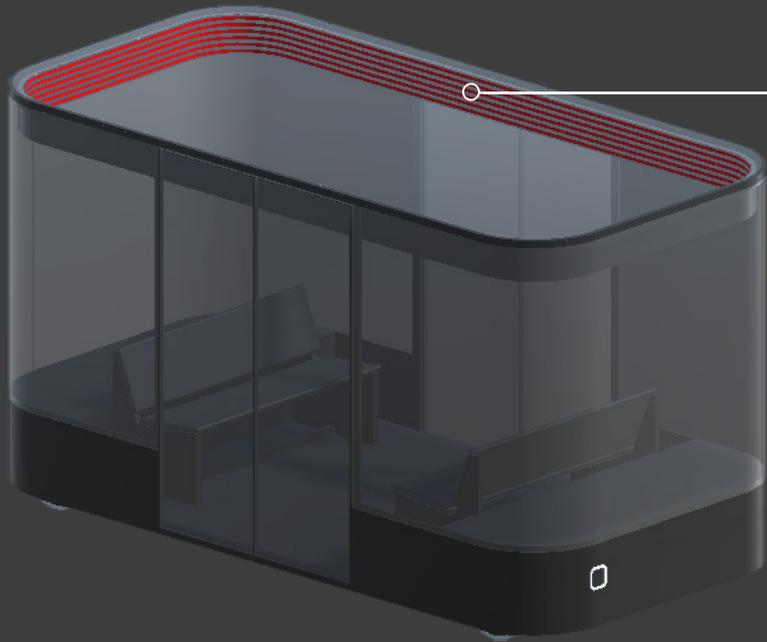
Pre-Notification



Stop!

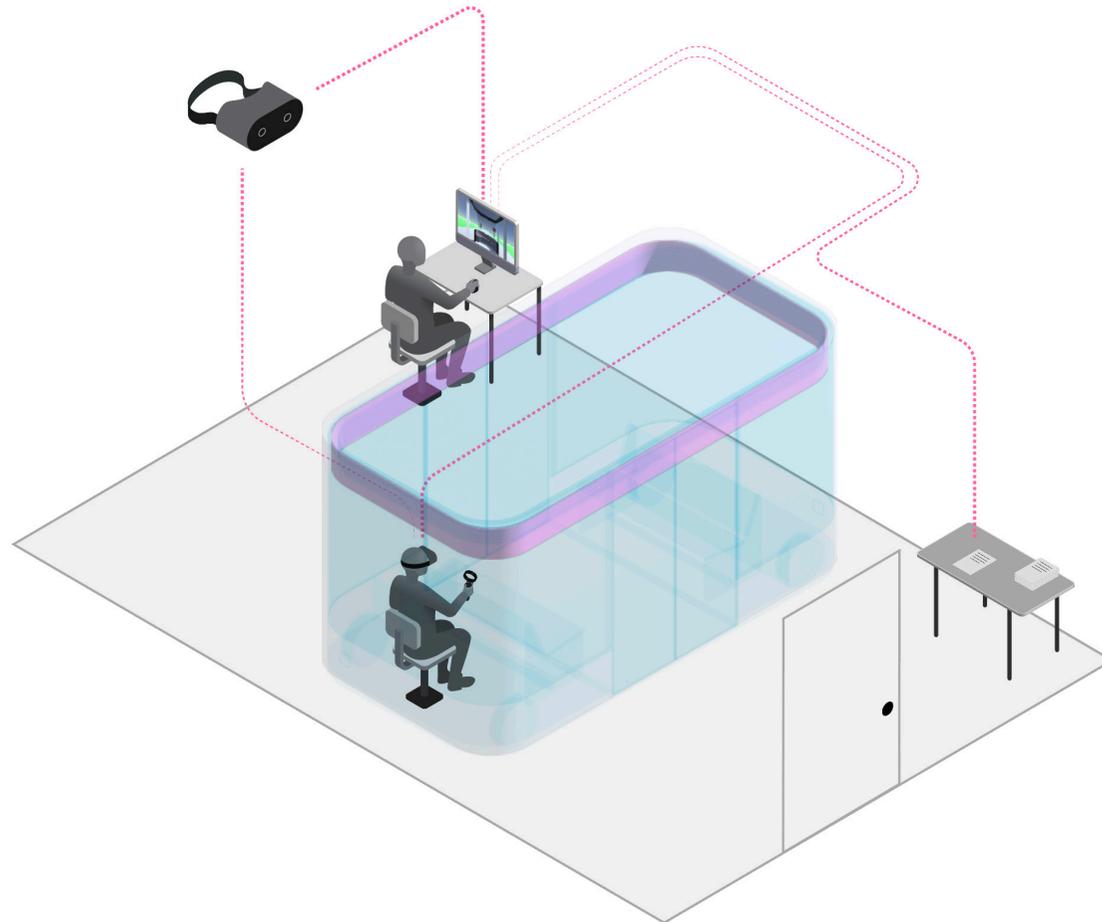
# ➤ Versuchsaufbau

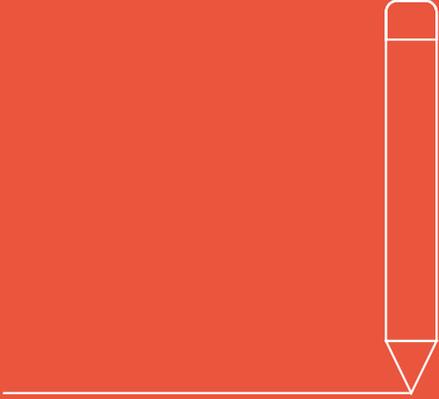
Designvorschlag - Entscheidung für das Lichtband



## ↗ Versuchsaufbau

Formaler Aufbau des gesamten Experiments





**Teil 4**  
Erkenntnisse

## ➤ Erkenntnisse

### Auswertung von Vertrauensbildung nach Aspekten der ISO-Norm 9241

#### **Erlernbarkeit**

Die VR-Simulation zeigte, dass das neuartige **Ring-Display zur Informationsvermittlung** von den Probanden leicht verstanden wurde, ohne eine ausführliche Erklärung oder längere Eingewöhnungszeit zu erfordern. Die Passagiere fühlten sich durch den Informationsgehalt des Displays **nicht überfordert**, sondern wünschten sich **sogar mehr Feedback**. Es wurde jedoch darauf hingewiesen, dass es wichtig ist, den Punkt zwischen dem Wunsch nach zusätzlichem Feedback und der potenziellen Überforderung zu finden.

#### **Selbstbeschreibungsfähigkeit**

Die meisten Probanden empfanden die verschiedenen **Zustände des autonomen Shuttles** (Erkennung der Gefahr, Bremsvorgang, Stillstand) als **klar verständlich**. Dennoch gab es Momente, in denen der Systemzustand nicht allen Probanden klar war. Zusätzlich äußerten sie den Wunsch nach mehr Feedback, insbesondere bei der zeitlichen Abstimmung von visuellen Benachrichtigungen.

## ➤ Erkenntnisse

### Vertrauensbildung nach Aspekten der ISO-Norm 9241

#### Erwartungskonformität

Die Passagiere empfanden das **Verhalten** des autonomen Shuttles als **konsistent und aufmerksam**. Die visuelle Information wurde als **nachvollziehbar** angesehen, und Transparenz bezüglich des Fahrzeugverhaltens und des Steuerungssystems wurde erreicht. Es wurde jedoch erneut der Wunsch nach mehr Feedback geäußert.

#### Nutzendenbindung

Die Mehrheit der Probanden gab an, **weiterhin ein solches Fahrsystem nutzen zu wollen**, und empfand die Informationsvermittlung als gut in das Gesamtbild integriert. Sie **konnten sich** während der Fahrt **auf eigene Aktivitäten konzentrieren**, was auf ein hohes Maß an Vertrauen in das autonome Shuttle hinweist.

## ➤ Erkenntnisse

### Vertrauensbildung im Hinblick auf biosensorische Daten

#### Zeitliche Vorankündigung und Erregungswerte

Die zeitliche Vorankündigung von Bremsmanövern hatte **keinen signifikanten Einfluss** auf die **Stresswerte** der Probanden. Jedoch wurde festgestellt, dass **zu kurze Vorankündigungen** von den Passagieren **schwerer einzuschätzen** waren und als mögliche Systemfehler wahrgenommen wurden. Daher wird empfohlen, visuelle **Vorankündigungen zu vermeiden, wenn** die verbleibende **Zeit bis zur Aktion zu knapp** ist. Die Verwendung weiterer Sinneskanäle wie Audiosignale oder haptische Signale könnte die Vorankündigung ergänzen und sollte in zukünftigen Studien untersucht werden.

Insgesamt zeigen die Ergebnisse, dass das **Ring-Display zur Vertrauensbildung beiträgt**, indem es Informationen verständlich vermittelt und Nutzendenbindung fördert. Dennoch gibt es **Raum für Verbesserungen**, insbesondere in Bezug auf die **Menge** und den **Zeitpunkt des Feedbacks** sowie die Gestaltung von Vorankündigungen für Fahrmanöver.

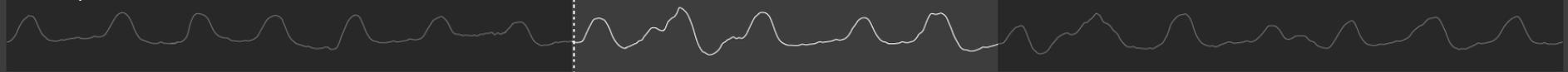
# ➤ Erkenntnisse

Auswertung des Experiments

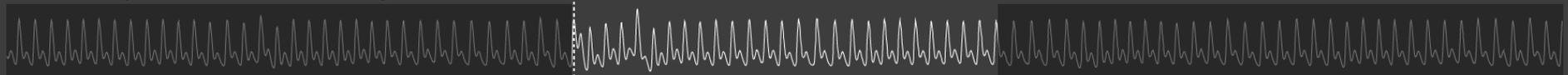


■ brake event

■ respiration



■ blood pressure variability



■ electrodermal activity



# Erkenntnisse

## Auswertung des Experiments

#XY31-45220403 | dispositional trust

open to experiences      extraversion



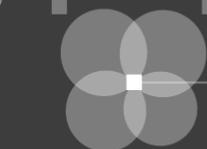
extraversion

agreeableness

#XY30-45220403 | situational trust

computer self efficacy

driving self efficacy



mood

subject matter

#XX45+220403 | learned trust

degree of retention

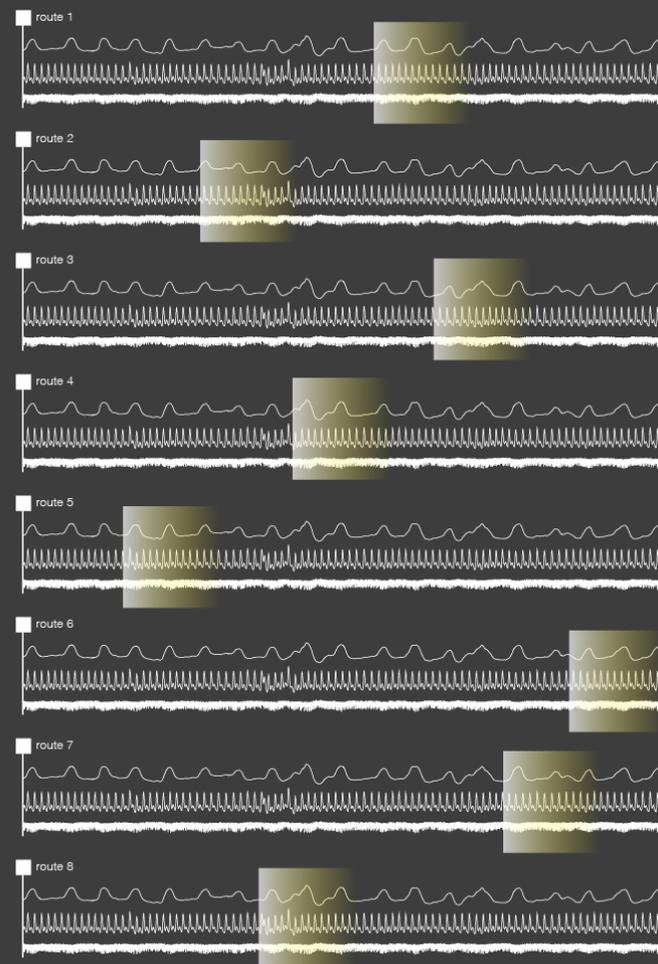
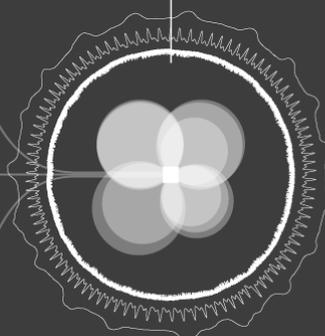
self descriptiveness

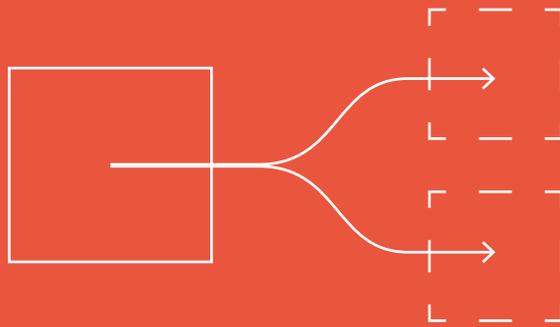


user engagement

user expectation

respiration  
blood pressure variability  
electrodermal activity





## Teil 5

Fazit und Ausblick

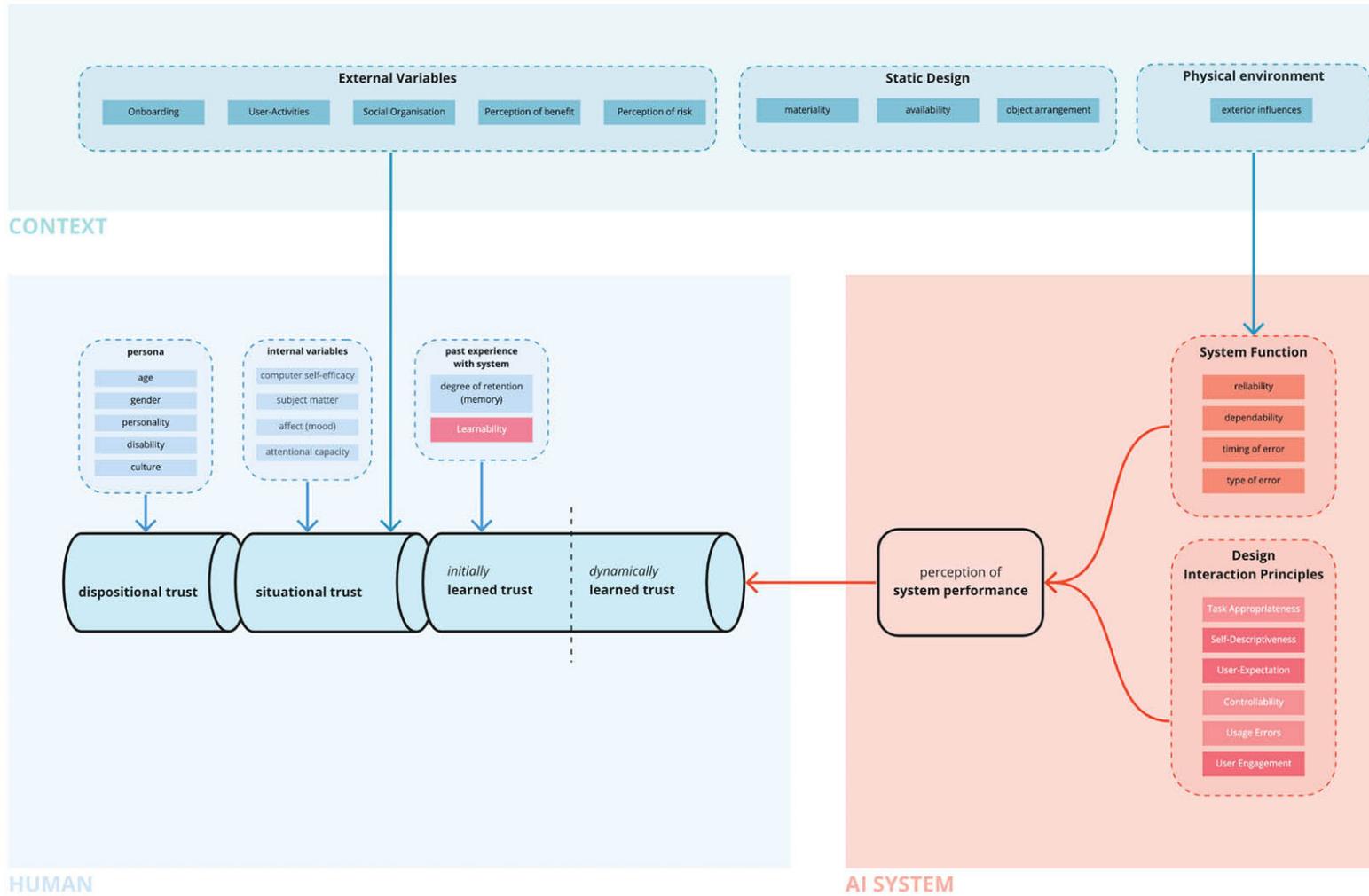
## ➤ Fazit und Ausblick

Aus der Studie geht hervor, dass es nicht allein die **Zuverlässigkeit** der Leistung des Systems ist, die Vertrauen bildet. Ein wesentlich **nachhaltigeres Vertrauen** in ein KI-System wird durch die **Nachvollziehbarkeit der „Motive“** (also der Entscheidungsprozesse) der KI erzeugt. Diese „Motive“ können nur durch eine **neue Form von Sprache** vermittelt werden, die den “Denkprozess“ des Systems verständlich macht. Hier ist durch die Gestaltung eines grafischen Interfaces im Rahmen dieses Projektes ein erster Schritt gemacht worden. Es konnte ein **Informationsfluss** bezüglich des Verhaltens der KI hergestellt werden, der **aufmerksam, konsistent und selbsterklärend** wahrgenommen wurde. Im Rahmen dieser Studie ist die **Aufgabe des Designs** im Aufbau eines solchen **transparenten Informationsflusses** innerhalb des Vertrauensbildungsprozesses deutlich geworden.

Es bleibt die Frage, wie die **Fähigkeit** intelligenter Systeme **zu lernen** und sich selbst **zu verbessern** in den **Validierungsprozess des entwickelten Vertrauensmodells** aufgenommen werden können. Hierfür müsste eine zuverlässige Methode des Monitorings von Stimmungen und des Verhaltens der Mitfahrenden integriert werden, sodass das Feedback über den Vertrauensgrad in Form von vergleichbaren Daten in das System zurückgeführt werden kann. Das entwickelte theoretische Modell bietet zumindest eine Grundlage für eine solche **dynamische Form von proaktiver Vertrauensbildung**, in der die Anpassungsfähigkeit der KI berücksichtigt wird.

# ➤ Ausblick

## Das Vertrauensmodell





# HessenAgentur

HA Hessen Agentur GmbH

**Dieses Projekt (HA-Projekt-Nr. 1014/21-16; 2021–2022) wurde aus Mitteln des Landes Hessen und der HOLM-Förderung im Rahmen der Maßnahme „Innovationen im Bereich Logistik und Mobilität“ des Hessischen Ministeriums für Wirtschaft, Energie, Verkehr und Wohnen gefördert.**

**AI·TRUST**

GOETHE  
UNIVERSITÄT  
FRANKFURT AM MAIN

**hfg** OF\_MAIN