JOHANNES GUTENBERG
UNIVERSITÄT MAINZ

KI-Einsatz bei der Verteilung von staatlichen Sozialleistungen

Prof. Dr. Petra Ahrweiler
Lehrstuhl Technik- und Innovationssoziologie / Simulationsmethoden
TISSS Lab, Johannes Gutenberg-Universität Mainz

TISSS LAB
Technology & Innovation Sociology / Social Simulation Laboratory

JG|U



Das Projekt

AI FORA

artificial intelligence for assessment

VolkswagenStiftung

https://www.ai-fora.de

TISSS LAB

Technology & Innovation Sociology / Social Simulation Laboratory

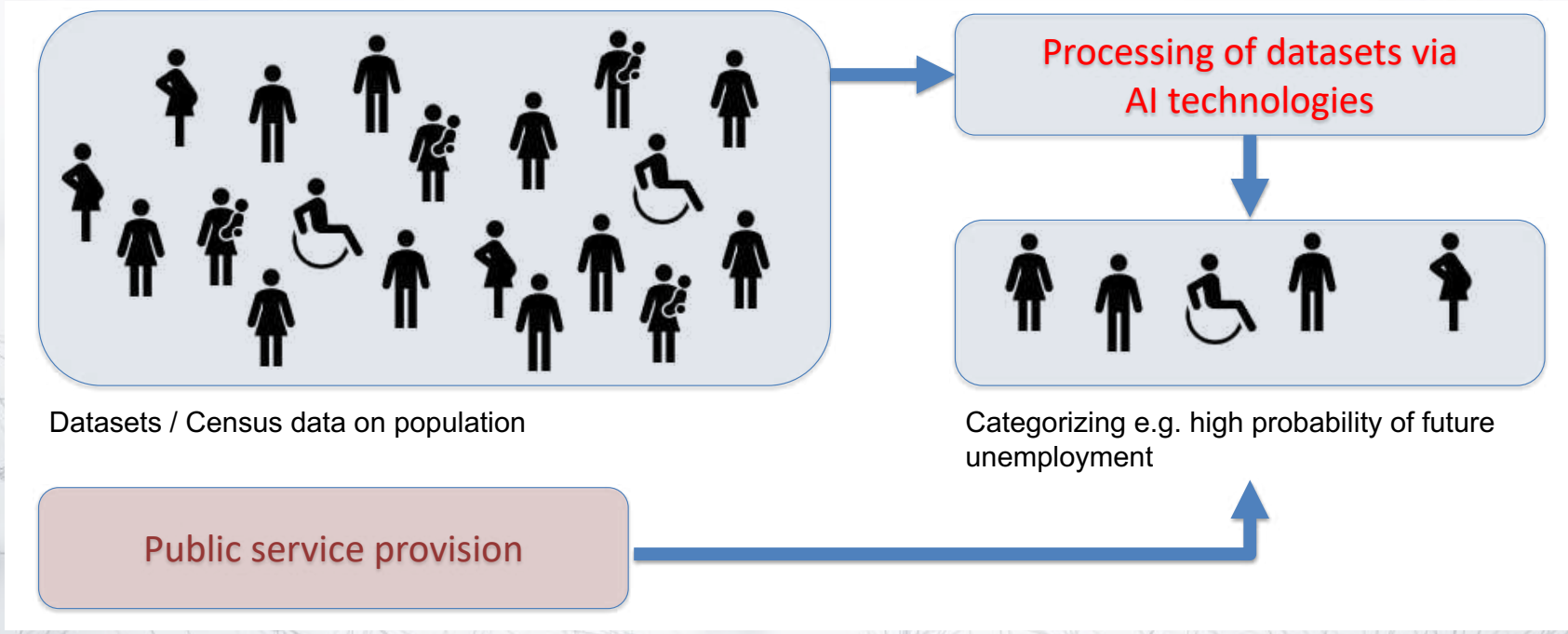
JG|U

TISSS LAB
Technology & Innovation Sociology / Social Simulation Laboratory

Das Start-Video kann auf der Webpage angeschaut werden

JG|U

In den öffentlichen Verwaltungen vieler Länder wird mittlerweile vermehrt Künstliche Intelligenz (KI) eingesetzt, um über die Verteilung von staatlichen Sozialdienstleistungen zu entscheiden.



KI zur Entscheidung über öffentliche Dienstleistungen nach Bürgerprofilen
Weitreichende Folgen für die betroffenen Personen

Kategorisierung und Bewertung

Diese Entscheidungen zur Verteilung knapper öffentlicher Güter basieren auf der Evaluation und Bewertung von Bürgerprofilen entlang von Kriterien zur Empfangsberechtigung wie legal/betrügerisch, bedürftig/nicht-bedürftig oder verdient/nicht-verdient.

Dies sind Wert-Entscheidungen zur sozialen Gerechtigkeit

Gesellschaftlicher Diskurs

Obgleich die Art und das Ausmaß von KI-Implementierungen im Bereich sozialer Bewertung von Menschen für Zwecke staatlicher Ressourcenentscheidungen zwischen Ländern stark variiert, führt die Delegation der diesbezüglich notwendigen Werteentscheidungen an Maschinen überall zu wichtigen Fragen nach Ethik, Gerechtigkeit, Qualität, Verantwortung, Zurechenbarkeit und Transparenz der Entscheidungsgrundlagen.

In der Stadt Luoyang sind die ersten Polizisten mit Brillen zur Gesichtserkennung unterwegs – sie sollen sofort herausfinden können, was über Passanten in Datenbanken gespeichert ist. Noch ist das mehr Propaganda als Alltag. © Reuters

Wir sehen Dich!
(DIE ZEIT, 10.01.2019)

Gesellschaftliche Werte und gesellschaftlicher Kontext

Wahrnehmungen, Einstellungen und Akzeptanz bezüglich des Einsatzes von KI unterscheiden sich nun zwischen Ländern ganz erheblich, was unter anderem auf kontextspezifische Normen und Werte, Technologiestände, Wirtschaftsformen, zivilgesellschaftliche Haltungen, legislative, exekutive und judikative Charakteristika zurückzuführen ist.

TISSS LAB
Technology & Innovation Sociology / Social Simulation Laboratory

Das AI FORA-Projekt

Das Ziel des inter- und transdisziplinären Forschungsprojekts AI FORA ist es,

- den gegenwärtigen Status Quo und die zukünftigen Optionen für KI-basierte soziale Bewertung im Bereich der Verteilung öffentlicher Dienstleistungen zu verstehen, um bessere KI-Technologie für soziale Wohlfahrtsysteme zu gestalten.
- Basierend auf Daten aus elf empirischen Länderfallstudien zu Deutschland, Estland, Spanien, Indien, China, Nigeria, Iran, Ukraine, Italien, Mexico und den USA werden experimentelle Szenariosimulationen entwickelt, die das partizipatorische Co-Design unter Beteiligung gesellschaftlicher Stakeholder in einem Better-AI Lab anleiten.

AI FORA
artificial intelligence for assessment

VolkswagenStiftung

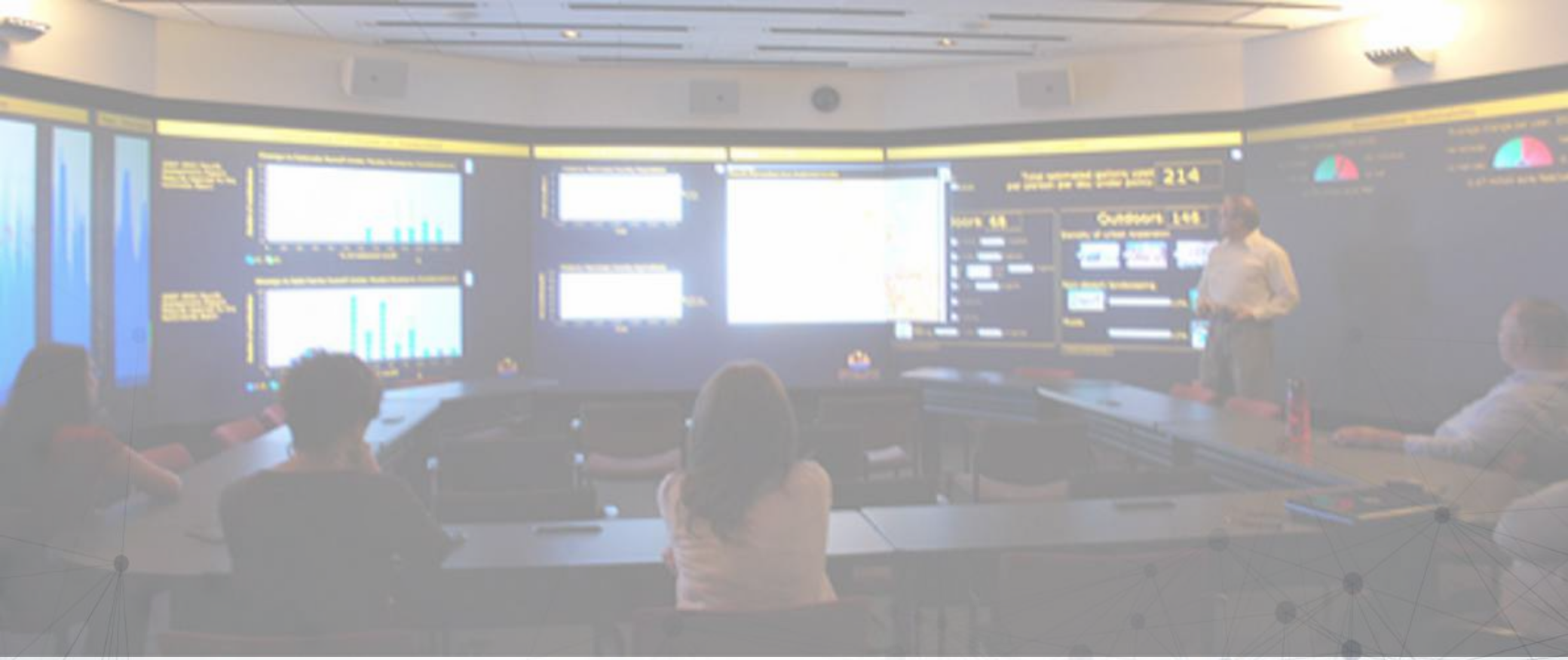
https://www.ai-fora.de

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

JG|U

Fallstudien mit ihren Themen

Land	Thema
Deutschland	Geflüchtete und Migration
USA	Arbeitslosigkeit und Bildung
China	Social Credit System / One Stop
Indien	PDS
Estland	Arbeitslosigkeit
Nigeria	Agrarsubventionen
Mexico	Agrarsubventionen
Spanien	Alle Leistungen
Italien	Alle Leistungen
Ukraine	Alle Leistungen
Iran	Alle Leistungen



Wie funktioniert das?

JG|U

TISSS LAB
Technology & Innovation Sociology / Social Simulation Laboratory

TISSS LAB
Technology & Innovation Sociology / Social Simulation Laboratory

KI: Mustererkennung

Intelligente Algorithmen, die auf leistungsfähigen Rechnern
BigData mit Maschinenlernen verbinden
und riesige Datenmengen
nach Mustern und Regelmäßigkeiten durchsuchen,
um neues Wissen zu generieren

Maschinenlernen ist höchst soziologisch

TISSS LAB
Technology & Innovation Sociology / Social Simulation Laboratory

Statistische Modellierung und Sozialprognose

- Statistische Vorhersagen: Zusammenhänge zwischen Variablen und deren Wahrscheinlichkeit
- Statistische Modellierung: Verfahren zur Bestimmung der Zukunft von bestimmten systemischen Variablenzusammenhängen

- Beispiel: wo hohe Schulabschlüsse, da
  hohes Einkommen

- Keine Garantie für den Einzelfall, aber...

- Signifikanter Zusammenhang
  - Die Wahrscheinlichkeit, mit einem hohen Schulabschluss ein hohes Einkommen zu erreichen, ist deutlich erhöht, weshalb Eltern gemeinhin versuchen, ihren Kindern eine gute Bildung zu ermöglichen.
  - Sie verlassen sich also zu Recht auf die Aussagekraft der Statistik für Zukunftsprognosen.

Profile und Prognosen

Was wäre, wenn: Verfügbarkeit des Profils eines Menschen als Netz von Variablen mit exakten Daten auf Personenebene

- Nutzung der Prognosekraft von statistischer Modellierung mit solchen Profilen zur Vorhersage zukünftigen sozialen Verhaltens
- (Völlige?) Transparenz der Bürgerinnen und Bürger durch stetige, automatisierte Datenerhebung mit Echtzeitdaten zu Prognosezwecken
-



Zwischenfazit:

- Es geschieht nichts radikal Neues – statistische Modellierung zur Sozialprognose gibt es schon lange
- Aber es geschieht durch die Verfügbarkeit großer Datenmengen in bisher nicht möglichem Ausmaß mit bisher unerreichter Abdeckung und mit noch unbekannten Folgen

TISSS LAB
Technology & Innovation Sociology / Social Simulation Laboratory
JG|U

TISSS LAB
Technology & Innovation Sociology / Social Simulation Laboratory

Wie funktioniert KI bei der Verteilung von Sozialdienstleistungen?

- Es werden Profile aus vergangenen Daten generiert: z.B. wie sah der bisherige Top-Student, der es zu etwas gebracht hat, bisher aus?
  - Vielleicht: männlich, weiss, kein Migrationshintergrund, aus gebildetem Elternhaus, höhere Einkommensklasse der Eltern etc.
- Heutzutage haben wir riesige Datenbasen über die gesamte Bevölkerung auf Individualebene
- Jede/r einzelne bekommt dann einen Score (einen statistischen Wahrscheinlichkeitswert, wie wahrscheinlich es ist, in Zukunft zum Zielprofil zu gehören)
- Nach diesen Scores verteilt die Maschine dann Sozialleistungen, beispielsweise Staatsstipendien für einen Universitätsbesuch
- Warum macht das eine Maschine? Effizienzgewinne (ist billiger als Personal), “Objektivitätsgewinne”, “Rationalitätsgewinne”
- Überall ist Technologieproduktion Sache von Expertennetzwerken, für den technologischen Laien nicht transparent und vom gesamtgesellschaftlichen Diskurs abgekoppelt

JG|U

AI assessing people - Data is key

- You need sufficient amounts of good training data
- This data might be sensitive
- You need a database on all individuals of interest
- This is past data
  - Maybe, the training data set shows that mostly married white males from rich areas and non-vulnerable groups with high interpersonal skills have been recruited for technical jobs in big companies in the past: think of your profile, your score, and somebody making an important decision due to this score (maybe not giving you the job you are qualified for)
- Bias will be transported into the future and cemented from training data, to profiling, to scoring, to decision making
- AI will continue with this bias that is in your country's database

Here, the algorithm scores you as:
0 (no match)
0
0
1 (match)
0
1
With 2 out of 6 you are out

Weltanschauliche, ethische und soziale Aspekte

- “Bias”: Hier wird diskriminiert und ausgegrenzt
  - Zementierung der Vergangenheit: Was mal vorrangig war, wird über den Bias verfestigt
  - Stell Dir vor, Du bist nicht-weiss, weiblich, hast Migrationshintergrund…
- Werte sind in den KI-Algorithmen nicht implementiert, selbst wenn so etwas wie Bias Thema im gesellschaftlichen Diskurs ist: es kommt in der Technologieproduktion nicht an
- Die Zukunft ist abgeschlossen (Minority Report): Aber es ist doch nur eine Wahrscheinlichkeit, keine Gewissheit…
- Es gibt Gewinner und Verlierer: die Verlierer haben keine Stimme

Was müsste passieren?

- Zentrale gesellschaftliche Werte sind berührt: Gerechtigkeit, Gleichheit, Fairness, Minderheitenschutz, Schutz des Schwächeren, oder auch einfach das Recht, es (in Zukunft) gut/besser machen zu dürfen..
- Dies /der gesellschaftliche Wertediskurs müsste an die Technologieproduktion angeschlossen werden

Der gesellschaftliche Wertediskurs ist stark kulturabhängig

- '**Kultur**': "the unwritten rules of the social game" (Hofstede et al. 2010), Kultur als Eigenschaft eines Sozialzusammenhangs, die von allen Mitgliedern geteilt wird
  - Beispiel: eine Gesellschaft mit strenger Kategorisierung, ausgeprägten Hierarchien und Machtaysmmetrien vs. eine Gesellschaft, die Integration, Netzwerke und Kommunikation fördert
  - der Unterschied wird sich auf die Art und Akzeptanz der Verwendung von KI in sozialen Bewertungssystemen auswirken
  - Andere Vorstellungen von sozialer Gerechtigkeit, andere Bewertung von Gleichheit/Ungleichheit etc.

AI FORAs Forschungsfragen: KI, Gesellschaft, Werte

- Wie und von wem waren die Verteilungspraktiken vor KI-Einsatz in der jeweiligen Gesellschaft organisiert? Wer waren die Gewinner, wer die Verlierer?
- Welchen gesellschaftlichen Werten folgten die Verteilungspraktiken vor dem Einsatz von KI?
- Wie sind die Verteilungspraktiken jetzt mit KI organisiert? Wer sind die Gewinner und Verlierer?
- Sind die gesellschaftlichen Werte in den KI-Algorithmen auffindbar?
- Gibt es einen gesellschaftlichen Wertediskurs? Wie sieht der gegenwärtige Wertediskurs zu den (KI-basierten) Verteilungspraktiken aus? Gibt es einen Wertewandel? Wie wird die "gute Gesellschaft" als Zielvorstellung im Wertediskurs thematisiert?
- Sind alle gesellschaftlichen Akteure am Wertediskurs beteiligt, die von den Verteilungspraktiken betroffen sind (Stakeholder)?
- Wie kann man in der jeweiligen Gesellschaft die Technologieproduktion an den Wertediskurs ankoppeln?

Methoden

- Empirische Forschung zur Beantwortung der Forschungsfragen in den Fallstudienländern

- Modellierung und Simulation zu den möglichen Zukünften der jeweiligen Gesellschaften

- Co-Design Labs unter Beteiligung aller Stakeholder zur gesamtgesellschaftlichen Technologieproduktion

Fallstudienpartner

Sozialwissenschaft und Informatik/Technologieproduktion

- Sozialwissenschaften
  - Gesellschaftlicher Wertediskurs
  - Machtverhältnisse, Wirtschaft, Kultur
  - Akteure und Netzwerke
  - Stakeholderpartizipation
- Informatik
  - KI-Algorithmen im Einsatz
  - Daten und Programmierung
  - Co-Creation Labs

TISSS LAB
Technology & Innovation Sociology / Social Simulation Laboratory

Minderheiten und benachteiligte Bevölkerungsgruppen

- Verlierer der gegenwärtigen Verteilungspraktiken sind
  - Menschen “ohne Stimme”
  - Marginalisierte, ausgegrenzte, manchmal bekämpfte Communities
  - Benachteiligte Gruppen an den gesellschaftlichen Rändern
- Barrrieren der Beteiligung an Entscheidungsprozessen sind hoch
  - Niedriger Bildungsstand
  - Schwaches Einkommen
  - Keine Diskurserfahrung
  - Kein Vertrauen in Institutionen
  - Kein Selbstvertrauen
  - Keine Motivation

Wie also lokale Stakeholder-Gruppen erfolgreich and der Technologieproduktion beteiligen?

Beispiel Indien

- Public Distribution System (PDS)
- Aadhaar-Registrierung mit Datenbasis und KI der Zentralregierung in Delhi
- Sozialdienstleistungen nach hinduistischer Kastenzugehörigkeit
- “Scheduled Cast” (8000 Kasten), Kastenzertifikate nach Abstammungsbäumen
- “Positive” Diskriminierung
- Viele Kastenlose
- Sehr viel Korruption (30% kommen an)
- Verfassung gegen Tradition
- Eine Gesellschaft im Umbruch
- Reformbestrebungen in Südindien mit den Regionalregierungen weit weg von Delhi

JOHANNES GUTENBERG UNIVERSITÄT MAINZ
JG|U

Beispiel Indien

- Aber: für Europäer sehr schwer zu verstehen
- Man kämpft nicht gegen Klassifizierung und KI (Kaste und KI-Verteilung), sondern für eine andere Klassifizierung (Tribes/Communities und KI-Verteilung), in der die Dalit besser wegkommen

- Gleichheit ist ein schwieriges Konzept, die “britisch eingepflegte” Verfassung mit allgemeinen Menschenrechten ist nicht richtig angekommen
- Es gibt viele Verlierer, viele Konflikte, viel Mißtrauen, viel Gewalt und Frustration
- Es gibt Angst vor Repressalien und davor, auch noch das wenige, was man hat, zu verlieren
- Selbst gutwillige Reformkräfte in Politik und Technikproduktion sind handlungsunfähig

Partizipation als Herausforderung

- Um ethisch und gesellschaftlich verantwortliche KI zu entwickeln, bedarf es der Interaktion zwischen heterogenen Akteuren aus den verschiedensten Bereichen, die "Stimme" brauchen, um auf Augenhöhe zu kommunizieren, ohne durch das Umfeld, in dem diese Begegnungen stattfinden, vorkonfiguriert und eingeschränkt zu sein.
- Risiko des Scheiterns: Erfolgreiche Einbeziehung lokaler Interessengruppen
- Einzelne Interessengruppen sprechen möglicherweise nicht gleichermaßen ihre Meinung aus und bringen ihre spezifische Perspektive und ihr Fachwissen ein, da sie von ihrer Umgebung gewarnt oder eingeschüchtert werden.
- Dieses Risiko besteht dann, wenn partizipative Formate und Veranstaltungsorte nicht neutral sind, sondern die Interessen eines beteiligten Akteurs vertreten, so dass eine horizontale und integrative Kommunikation nicht möglich ist.
- Wo können diese partizipativen und interaktiven Formate stattfinden, ohne dass einzelne Gruppen/Einzelpersonen vorkonfiguriert sind oder durch ihre Umgebung bevorzugt oder diskriminiert werden?
- Dies muss in einem "sicheren Raum" geschehen, d.h. an einem neutralen, aber unterstützenden Ort der gewaltfreien Kommunikation, wo gleiche Möglichkeiten für vertrauliche Meinungsäußerungen und partizipative Entscheidungsfindung gegeben sind.

“Safe Spaces”
Network Laboratories for innovating Societies dealing with Complexity

- Das AI FORA-Projekt verwendet Standorte von "Vermittlern" in jedem Fallstudienland als "Sichere Räume".
- Dies sind Netzwerkorganisationen, die auf interkulturelle und intergesellschaftliche Kommunikation spezialisiert sind.
- Die Rolle von Brücken, Vermittlern, Netzwerken, Grenzobjekten, Kommunikationsdisketten spielen
- Sie werden verwendet, um mit der Pluralität der Perspektiven und der Vielfalt des interkulturellen Kontexts umzugehen
- Sichere Räume bündeln methodische Ressourcen, die eine gemeinsame Problemdefinition und Problemlösung unter Beachtung hoher Differenzierungsgrade ermöglichen
- Gewaltfreie Kommunikationsmethoden (Marschall Rosenberg)
- Low-Tech-Konsultationsmethoden
- Partizipative Multi-Stakeholder-Workshop-Methoden
- Abbildung partizipatorischer Systeme
- Szenario- und Prognosemethoden
- Formate der Zusammenarbeit. World Café, Fish Bowl etc.
- Gamifizierung und ausdrucksstarke Gruppenaktivitäten

- Partizipative Modellierung erwünschter Zukünfte und gesellschaftlicher Szenarien

Vielen Dank für Ihre Aufmerksamkeit!

AI FORA-Projekt im Internet:

https://www.ai-fora.de

Besuchen Sie uns auf unserer Webpage
https://technologyandinnovation.sociology.uni-mainz.de/tisss-lab/

TISSS LAB
Technology & Innovation Sociology / Social Simulation Laboratory

JG|U