

Artificial Intelligence for Assessment

Petra Ahrweiler

ICTP 2022



VolkswagenStiftung

TISS LAB
Technology & Innovation Sociology / Social Simulation Laboratory

JOHANNES GUTENBERG
UNIVERSITÄT MAINZ



JGU

AI use in policy and public administration: Some initial examples

- [The Netherlands](#): System “SYRI” (System Risk Identification)
- [France](#) (Targeting by Numbers. The Uses of Statistics for Monitoring French Welfare Benefit Recipients)
 - These systems detect fraud in social service provisions (especially unemployment benefits) and sort recipients into risk groups for future fraud.
- [Belgium](#): VDAP – Algorithmic Activation
 - System analyses and assesses search behavior of unemployed people on websites of job agencies
- [Austria](#) (similar to [Poland](#) and [Estonia](#) with system RITA: Machine learning and AI-powered public service delivery) since 2019: Scoring unemployed people into three groups
 - Close to job market (No support necessary)
 - Employable (Training programmes, professional development)
 - Unemployable (Financial support but no training)

[USA](#): Algorithms detect child abuse (Allegheny County in the US uses Allegheny Family Screening Tool): Scoring risk cases

- System is supposed to help social workers in early detection
- Firms develop standard software in these are

Source: Algorithmwatch 2019

The project idea

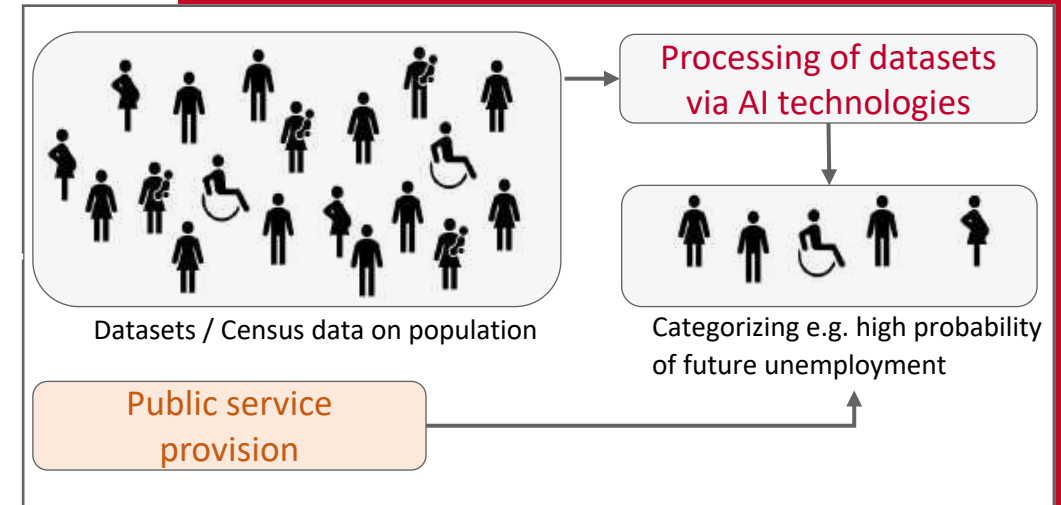
- **Domain:** Artificial Intelligence assessing people for receiving public goods
- **What the project is about:** How social justice is processed by technology
 - in different value contexts worldwide aiming at global fairness
- **Main hypothesis:** Discriminatory practices to distinguish beneficiaries from non-beneficiaries depend on the beliefs, norms and values in place as reference frameworks for public resource distribution in different societal contexts.
- Bias due to value judgements concerning fair distribution is prevalent in every selection system because they all work with inclusion/exclusion dichotomies, which are subject to cultural interpretation, social change and multi-actor negotiation.
- **Research:** Analysing cultural and social dynamics for providing social goods with and without AI in these countries allows to check the chances for a technology-triggered cohesive model of global justice integrating highly contextualized national value cultures

Assigning scarce resources in public services is always discriminatory and involves many issues

- Working with growing populations having to cover large demands
 - Dealing with public criticism
 - Working with social change and policy reforms
 - Balancing the margin of discretion in decision making
 - Dealing with vulnerable groups
 - Addressing bias and discrimination
 - Addressing corruption and fraud
 - Managing the relationship with “back-up systems”
 - etc. etc.
- **How can AI assist (becoming more efficient, more objective)?**

AI TECHNOLOGIES USED FOR SOCIAL ASSESSMENT IN PUBLIC SERVICE PROVISION OF NATIONAL WELFARE SYSTEMS

- In more and more countries, public administrations increasingly use Artificial Intelligence (AI) algorithms to decide on social service provisions and state benefits among their citizens.
- In our research context, AI means machine learning techniques for the profiling and scoring of individuals, accompanied by automated decision-making.
- Decisionmaking is highly value-laden: Present and future human behaviour needs to be categorised on scales such as legal recipients/fraudulent recipients, deserving/non-deserving, needy/non-needy, high-performing/low-performing, desirable/non-desirable, or acceptable/not-acceptable.



→ “Values are key”

How good is the 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 females in their fifties with a history as part-time employees due to family break are difficult to bring into the job market again when they fall unemployed: Think of your profile, your score, and a machine making an important decision due to this score (maybe not giving you the job that would help you and work out in your special case)
- **Bias and discrimination** will be transported into the future and cemented from training data, to profiling, to scoring, to decision making...

Social service provision is about (culturally dependent) value decisions

AI-based social assessment technologies for public service provision categorise present and future human behaviour on scales such as

- legal recipients/fraudulent recipients,
- deserving/non-deserving,
- needy/non-needy,
- high-performing/low-performing,
- desirable/non-desirable,
- or acceptable/not-acceptable
- Etc.



What is a fair decision?

Knowing that there is bias is not sufficient, of course there is!

The screenshot shows the IBM Research Trusted AI website. The navigation bar includes links for Home, Demo, Resources (highlighted), Events, Videos, and Community. The main heading is "AI Fairness 360 - Resources". Below this are buttons for "API Docs" and "Get Code". A secondary navigation bar includes "Overview" (selected), "Tutorials", "Guidance", "Glossary", and "Trusted AI Technologies". The main content area starts with "Welcome to AI Fairness 360" and a paragraph: "We hope you will use it and contribute to it to help engender trust in AI and make the world more equitable for all." This is followed by a paragraph explaining that machine learning models are increasingly used for high-stakes decisions, but they can be biased due to statistical discrimination. A third paragraph describes the AI Fairness 360 Python package, which includes metrics, explanations, and algorithms to mitigate bias. A fourth paragraph notes that the package is comprehensive and provides guidance material. A fifth paragraph encourages community contributions. Finally, a list of papers is provided, including "Optimized Pre-Processing for Discrimination Prevention" (2017) and "Classification with Fairness Constraints: A Meta-Algorithm with Provable Guarantees" (2018). A "Cookie Preference" button is visible in the bottom right corner of the page content.

IBM fairness machine



Case studies

AI FORA

TISS LAB

Technology & Innovation Sociology / Social Simulation Laboratory

JOHANNES GUTENBERG
UNIVERSITÄT MAINZ

JG|U



VolkswagenStiftung

TISS LAB

Technology & Innovation Sociology / Social Simulation Laboratory

JG|U

NO “ONE-FITS-ALL” APPROACH

Values are context-bound: What is **perceived as „social justice“** in a national welfare system largely **depends on cultural context**

- What is considered as „fair“ in one cultural context, might be considered as severe discrimination practice in another.
- Example India
- Example Germany/UK
- Also, **acceptance of AI use** for public policy vary between countries

Attitudes not only vary between countries but also within countries between societal groups.



→ “Context is key”



9 case studies to investigate how cultural values affect AI-based social assessment for public service provision

- Fairness concept of public welfare policy in a case study country based on a current categorisation system for social assessment
- Current implementation of fairness concept in administrative practices
- Welfare gaps and scenarios of desired system as identified by social context (assessment of welfare indices, dealing with vulnerable groups, addressing bias and discrimination, addressing corruption and fraud, what back-up systems in place, what says public discourse/acceptance, are there policy reforms, what is the space of leeway etc.)
- (Policy advise on) Future implementation of fairness concept based on a desired categorisation system for assessment
- **Value-sensitive AI - Contextualised AI - Participatory AI – Dynamic AI**

Bringing AI and societies together

- All country contexts contain issues of bias and discrimination of and against diverse populations - these are transported/enforced in AI systems
- To identify critical issues and improve AI systems, the dynamic value system of societies needs to be brought close to technology production
- AI FORA analyses value/context dependency in AI-based social assessment comparing **eleven countries** as case studies,
- identifies technology/society gaps and
- develops chances for improving policies for **contextualized AI systems that are responsive to value dynamics in societies**

The project in a nutshell

How can **better AI technology** be created that engages with societal norms and values of stakeholders, and that is responsive to socio-cultural settings and societal needs?

AI FORA analyses value and context dependency in AI-based social assessment for social service provision (fairness concepts) comparing eleven countries as case studies, identifying welfare gaps in current systems, and informing policy on **contextualised AI systems that are responsive to value dynamics in societies**



So how to come to fair decisions
(and train the algorithm on it)?

WHO CAN IMPROVE CURRENT SYSTEMS?

Improve issues around bias and discrimination trying to make AI systems more fair

Develop AI in relation to cultural context closing the gap between technology production and societal value propositions

Develop better AI in co-design

- This can not be done by one individual or one societal subsystem, but requires expertise and perspectives of all parts of society
- Multiple Stakeholders, including vulnerable populations who need empowerment

→ “Participation is key”



AI FORA'S PARTICIPATORY APPROACH

...to improve issues around bias and discrimination for fairer AI systems.

This requires

....the inclusion of all societal stakeholders, especially of those the most affected by bias and discrimination, who know the flaws of the system best

- Inter- and transdisciplinary multi-stakeholder workshops
- Workshops for dedicated target groups (e.g. vulnerable groups, policy makers, public administration staff; NGOs)

....a „Safe Space“ – a place for multi-stakeholder learning, where people can meet on eye-level, communicate freely about their different perspectives, and develop co-created solutions with the help of participatory methods

.....a toolbox of low-barrier participatory methods

- Participatory Systems Mapping (PSM)
- Gamification



Participatory approach

- Working with the expertise and perspectives of **all societal stakeholders** including vulnerable populations, to develop a joint „product“ in co-design
 - Inter- and transdisciplinary expert workshops
 - Workshops for dedicated target groups (e.g. policy workshops)
 - Multi-Stakeholder Workshops
 - Safe Spaces and Ethical Observatory (in the next session)
- Developing a **toolbox of low-barrier participatory methods** mostly implemented in multi-stakeholder workshops
 - Participatory Systems Mapping (PSM)
 - Gamification
 - Companion Modelling

AI FORA MIXED-METHODS APPROACH (SOCIAL & TECHNICAL SCIENCES)

Social Empirical Research

- Qualitative and quantitative research methods on “AI cultures”
- Identification of stakeholders’ experiences, perspectives and values

Social Simulation

- Empirically-informed participatory scenario modelling considering human behaviour, values, and context
- Modelling the interaction dynamics between society and technology for anticipating, monitoring, evaluating and improving the societal impact of AI-based assessment

Better-AI Lab

- Based on stakeholder-driven input from empirical social research and social simulation
- Co-designing an approach/methodology to create “better AI”

EMPIRICAL RESEARCH

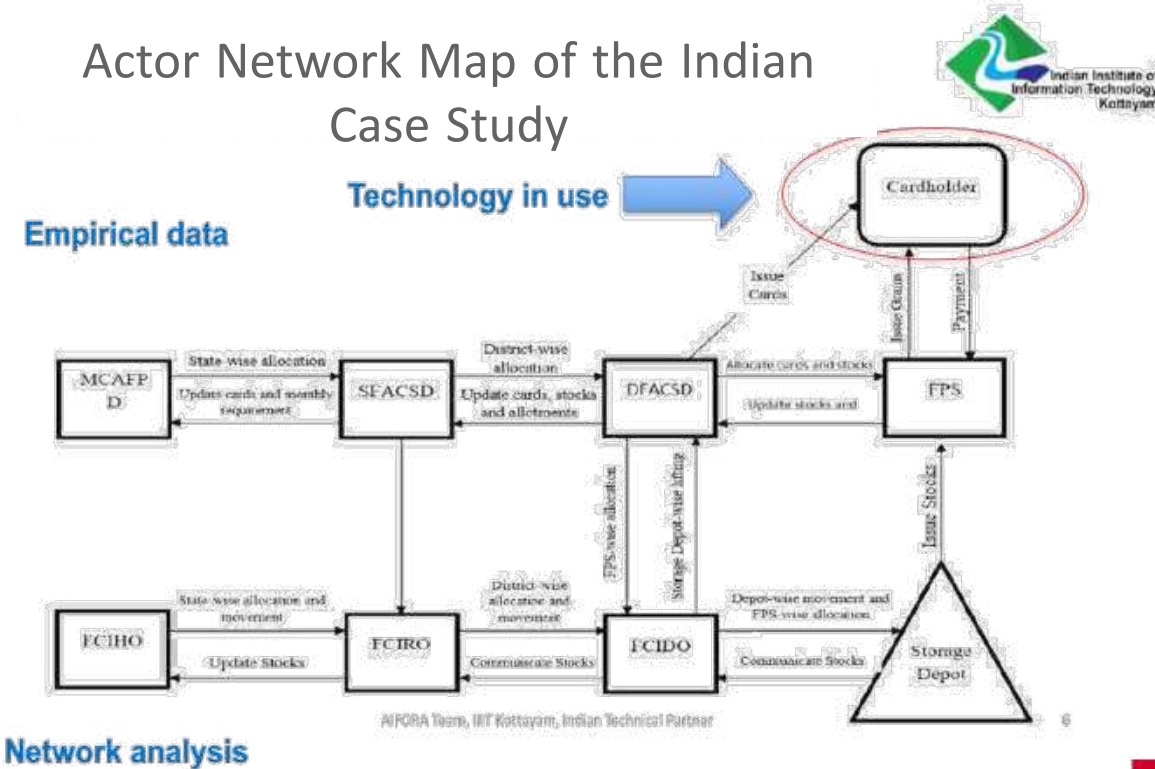
Workflow: Analysing...

- Actors, interactions, resource flows, networks (Actor-network maps)
- Cultural values, economy, regulations, policies
- Distribution algorithms in place (AI or else)
- Problems, gaps, barriers of the current system from a stakeholders' point of view
- Desired scenarios from stakeholders' point of view
- Recommendations for AI development and policy for "better AI"



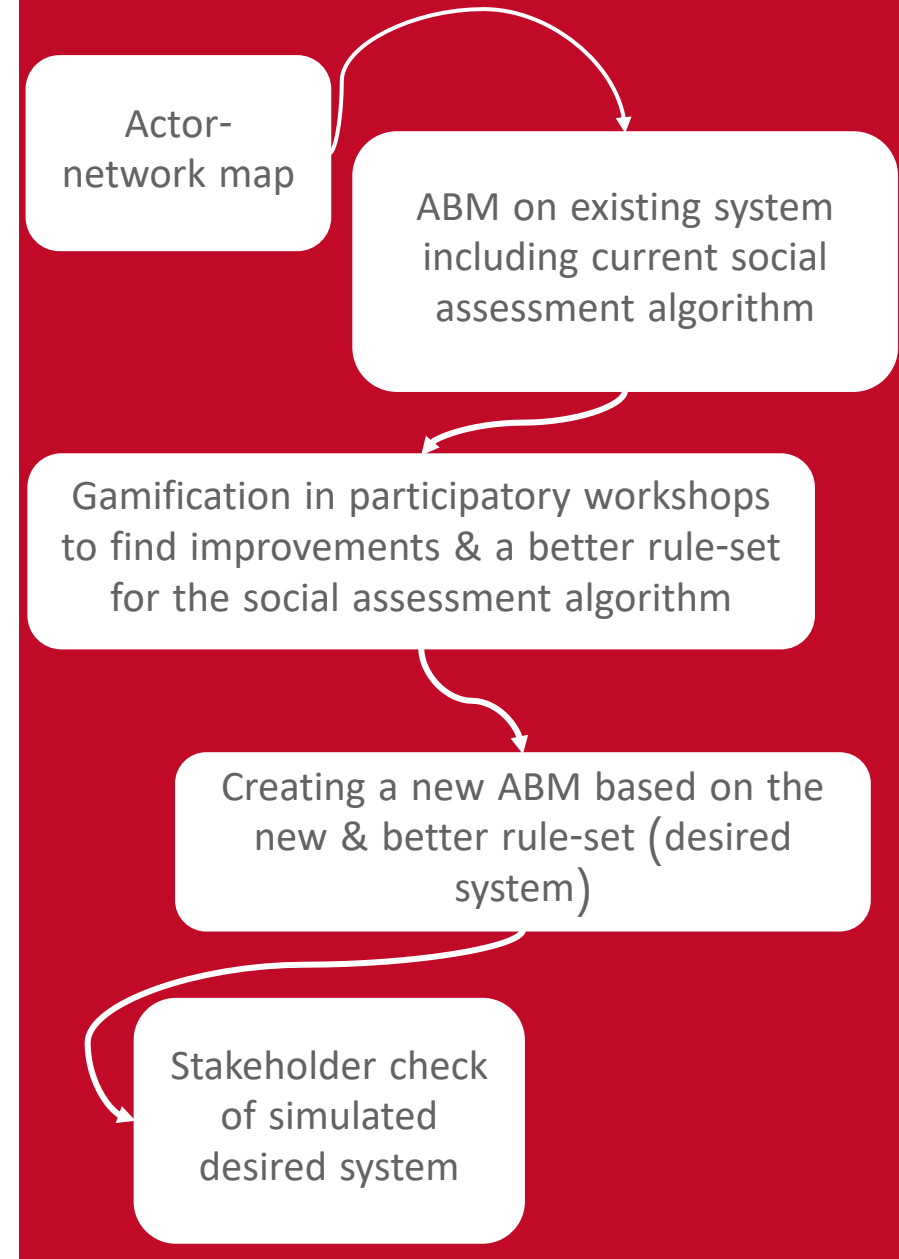
Aadhaar Card

Actor Network Map of the Indian Case Study



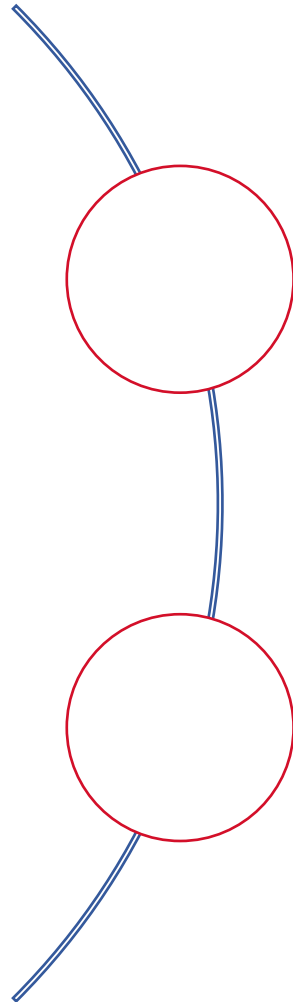
SOCIAL SIMULATION WITHIN AI FORA

- Scenario simulations to experiment with assessment algorithms in use and to model future alternatives
- Companion modelling approach: Engaging implicit and explicit knowledge of stakeholders / non-scientists to create computational models of the system of interest
- Use of Agent-Based Models (ABM) to simulate actions, interactions and interventions of agents (players, stakeholders)



- Mimick/play the system
 - Learn to deal with the complexities in the system
 - Change the algorithm
 - Simulate the system
 - Simulate policies for the system
- Training data is bad: biased
 - We will simulate unbiased training data according to the value discussions of participants
 - Artificial population to train the new algorithm

**AI FORA's
input to
BETTER AI**



Better AI through better automated decision making

- Better training data for ML provided by the ABM that produces data on the desired system specific to context
 - Legitimised by the values of societal stakeholders
 - Easily adaptable to value dynamics and value change
- Proof-of-concept to create training data for ML by social simulation

Better AI through living lab experience

- Better-AI Lab that allows users to embody representatives of stakeholder groups in virtual social assessment scenarios



India

Social science



Computer science

- **Sumathi Srinivasalu**, Anthropology, University of Madras, India
- **Ebin Raj**, Indian Institute of Information Technology Kottayam, India



Actor-network maps

<https://prsm.uk>



Participatory System Mapper

[About](#) [Our Story](#) [Examples](#) [Help Centre](#) [Community](#)

[Start now](#)



VolkswagenStiftung

TISS LAB

Technology & Innovation Sociology / Social Simulation Laboratory





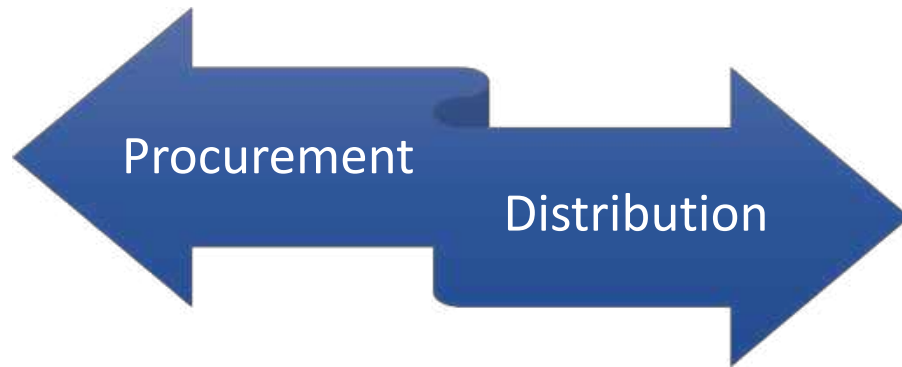
Procurement and Storage Process - Agents



AIFORA Team, IIIT Kottayam, Indian Technical Partner

3

Mapping the PDS



AIFORA Team, IIIT Kottayam, Indian Technical Partner

2

Distribution - Agents Involved



AIFORA Team, IIIT Kottayam, Indian Technical Partner

5





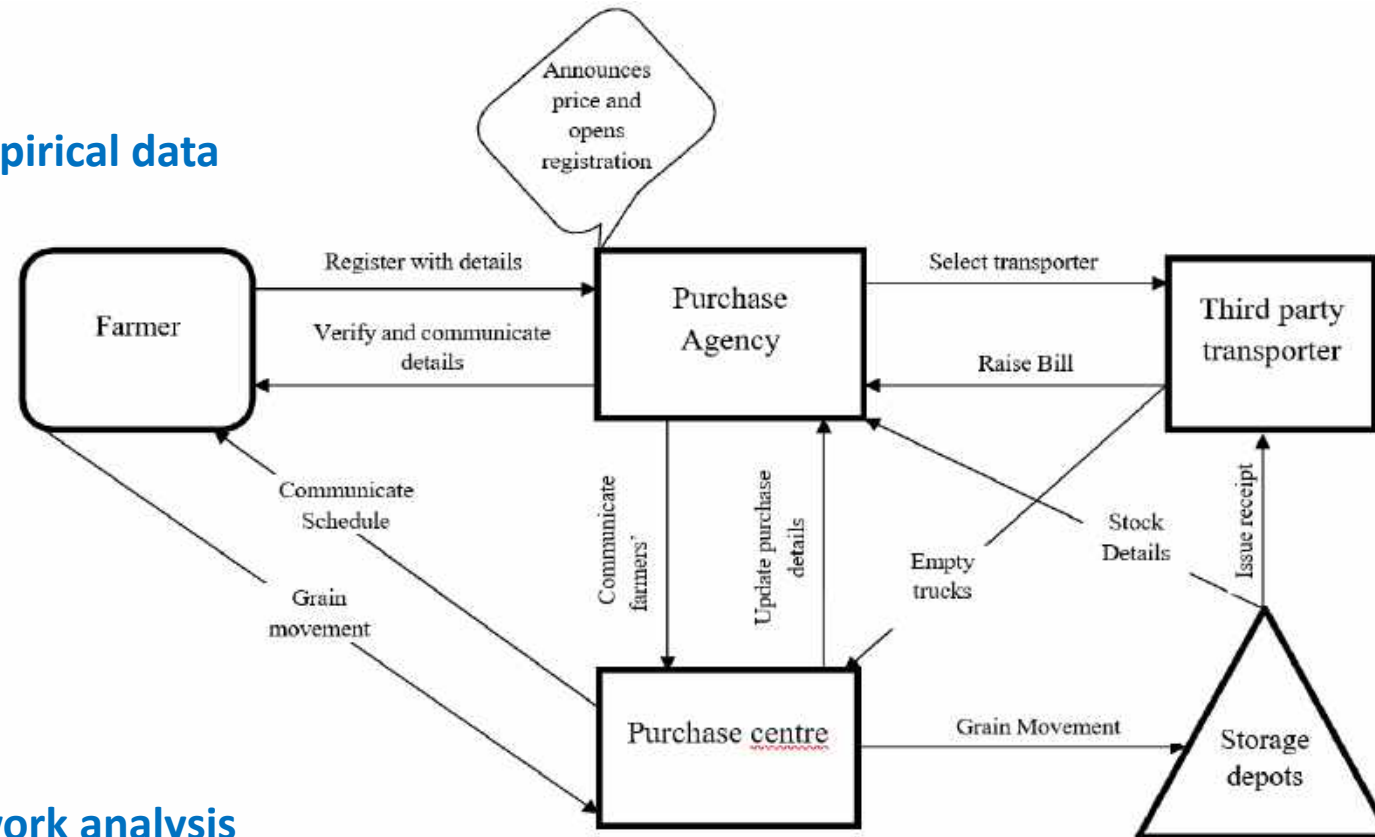
Agent interaction in PDS(Procurement)

Empirical data

Empirical data

Network analysis

Network analysis



AIFORA Team, IIIT Kottayam, Indian Technical Partner

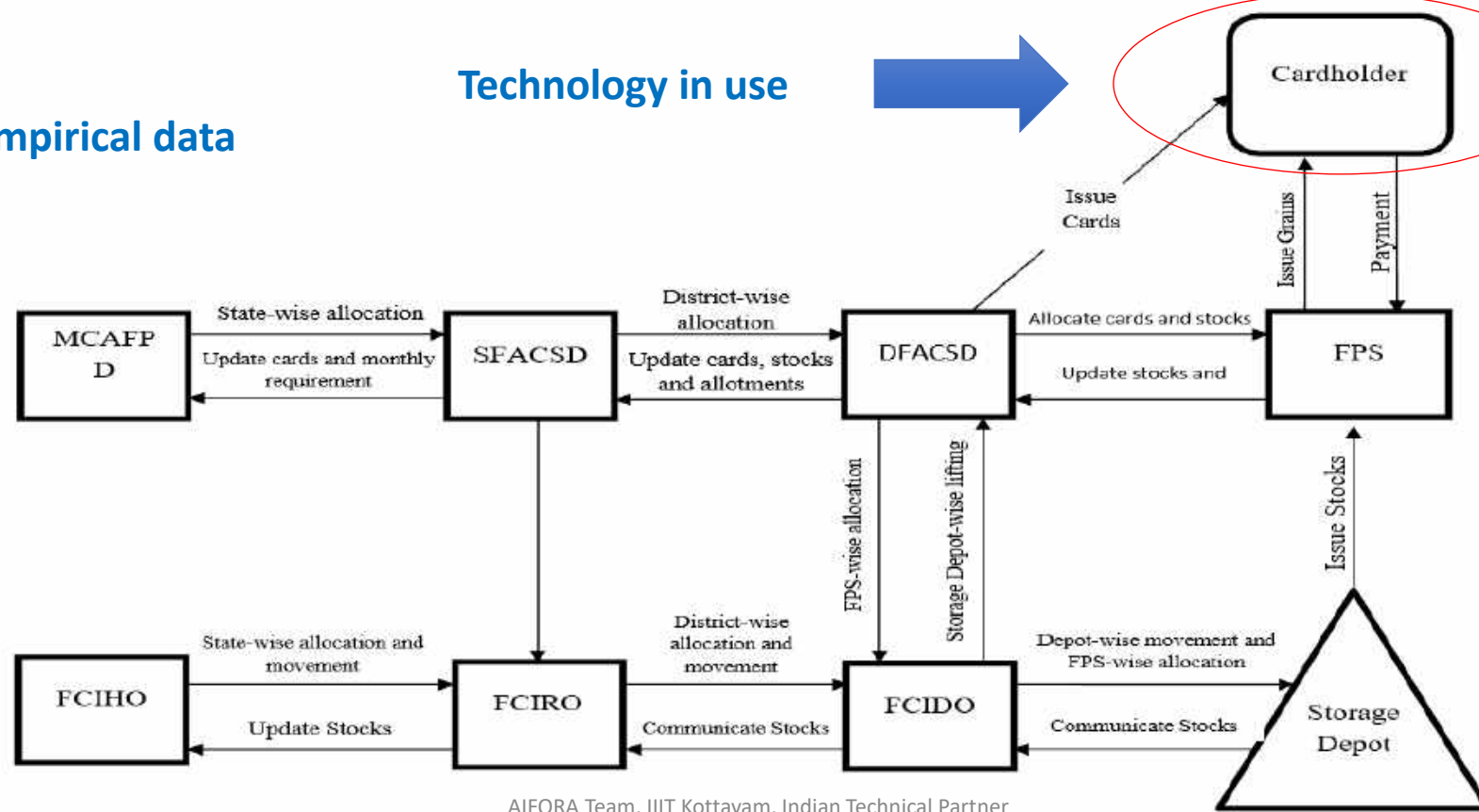
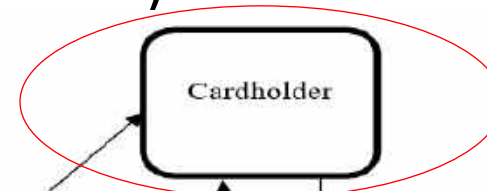




Agent interaction in PDS(Distribution)

Empirical data

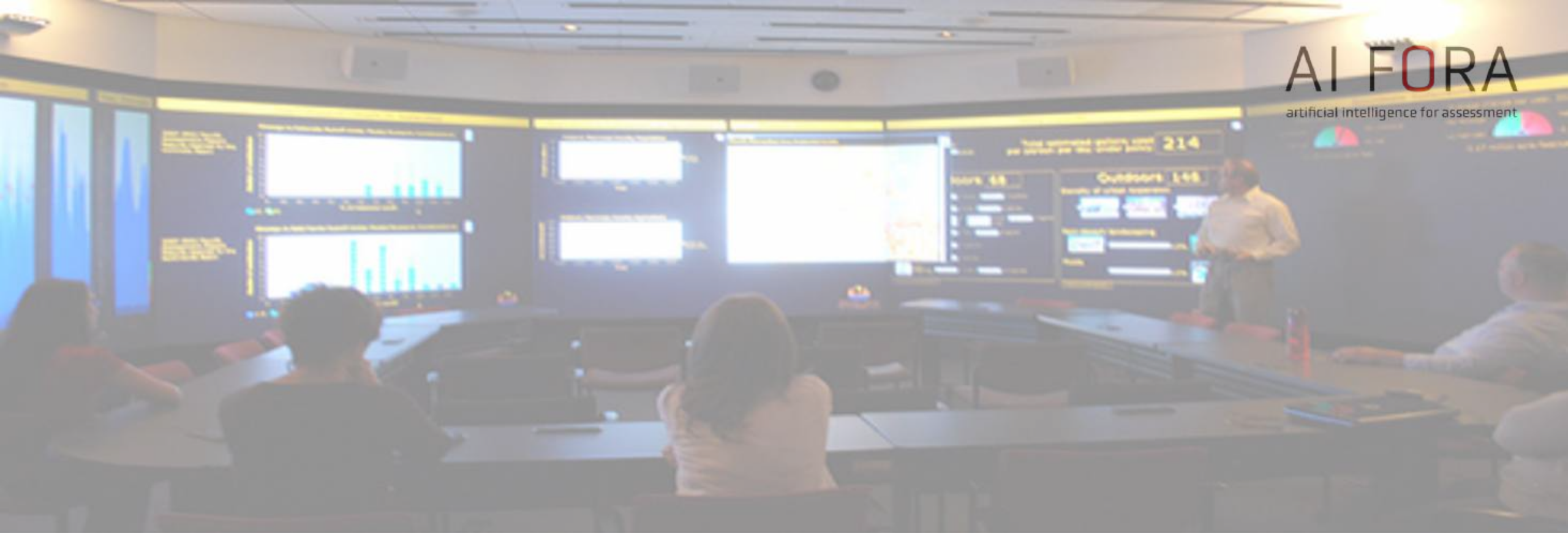
Technology in use



AIFORA Team, IIIT Kottayam, Indian Technical Partner

Network analysis





Technology-in-use





- Ration card, Aadhaar number, and electronic Points of Sale (ePoS). The system identifies a beneficiary through biometric authentication on ePoS devices at fair price shops
- The system runs with the support of two portals — Integrated Management of Public Distribution System (IM-PDS) (impds.nic.in) and Annavitran (annavitran.nic.in), which host all the relevant data
- Current research on distributions, biases etc. ongoing (**sampling bias!**)

Performance indicators



Problem	Performance Indicator
Leakages	<ul style="list-style-type: none"> • Average quantities purchased (take-off) per Cardholder • Average number of cardholders per FPS
Movement of foodgrain	<ul style="list-style-type: none"> • Average distance between the storage depot and the FPS • Average quantities moved between storage depots • Average distance between two storage depots • Average distance travelled by grain bags from the purchase centre to the storage depot • Number of trips made between the purchase centre and the storage depot
Poor off-take by states	<ul style="list-style-type: none"> • Average quantities purchased (take-off) per cardholder • Average number of days' inventory stocked at the FPS • Average number of cardholders per FPS • Average inventory stocked at the storage depots
Excess cost	<ul style="list-style-type: none"> • Average inventory stocked at the storage depots • Average distance between two storage depots • Average distance between the storage depot and the FPS • Average quantities purchased (take-off) per cardholder • Average number of cardholders per FPS • Average number of days' inventory stocked at the FPS • Average distance travelled by grain bags from the purchase centre to the storage depot • Number of trips made between the purchase centre and the storage depot • Average quantity procured per day
Speculation in the market	<ul style="list-style-type: none"> • Average quantity procured per day • Average number of farmers processed per day • Average number of days the purchasing centre works per purchase cycle • Average number of days' inventory stocked at the FPS • Average inventory stocked at the storage depots • Number of FPS served by a storage depot

AIFORA Team, IIT Kottayam, Indian Technical Partner





Gaps and barriers

Identifying gaps, barriers and fairness issues from a stakeholder perspective (Safe Spaces workshops: vulnerable groups, minorities)

- Scarcity, poverty (supplies for 20 days a month); rest supply by backup systems
- Family cards based on households discriminating e.g. against single mothers
- Caste membership and proof of identity
 - Scheduled castes
 - Positive discrimination
- Ethnic discrimination and tribes
- Corruption and identity fraud
- Local shops and migrant workers
- Food quality



- „Kitchen“-based cards
- Tribes instead of castes
- One nation– one ration



NetLogo — UnemploymentGame v1.2 (Users/petraahrweiler/Desktop)

Interface Info Code

normal speed | view updates [checked] | continuous [dropdown]

months: 101 | Settings...

The employment game Version 1.2

Current or last job:
Pink: unemployed
Blue: bad job
White: good job

Setup | Go once | Go

number-of-rounds: 100

number-of-agnts: 15

rent: 5

unemployment-benefit: 10

welfare-benefit: 5

disability-benefit: 10

voluntary-fund-rate: 2%

vol-fund-good-job-cost: 48

vol-fund-bad-job-cost: 50

good-jobs: 6

good-job-wage: 6

bad-jobs: 9

bad-job-wage: 4

bad-job-work-stress: 3

station-selection: random

assessment: random

holiday-cost: 5

holiday-delight: 4

Total wealth

Total happiness

Total graduates

Vacant good jobs: 13

Vacant bad jobs: 1

Amount in voluntary fund: 19

Command Center

THANK YOU FOR YOUR ATTENTION!



AI FORA CONSORTIUM, IN APRIL 2022