Rethinking Question Zero in AI

Prof. dr. Sennay Ghebreab

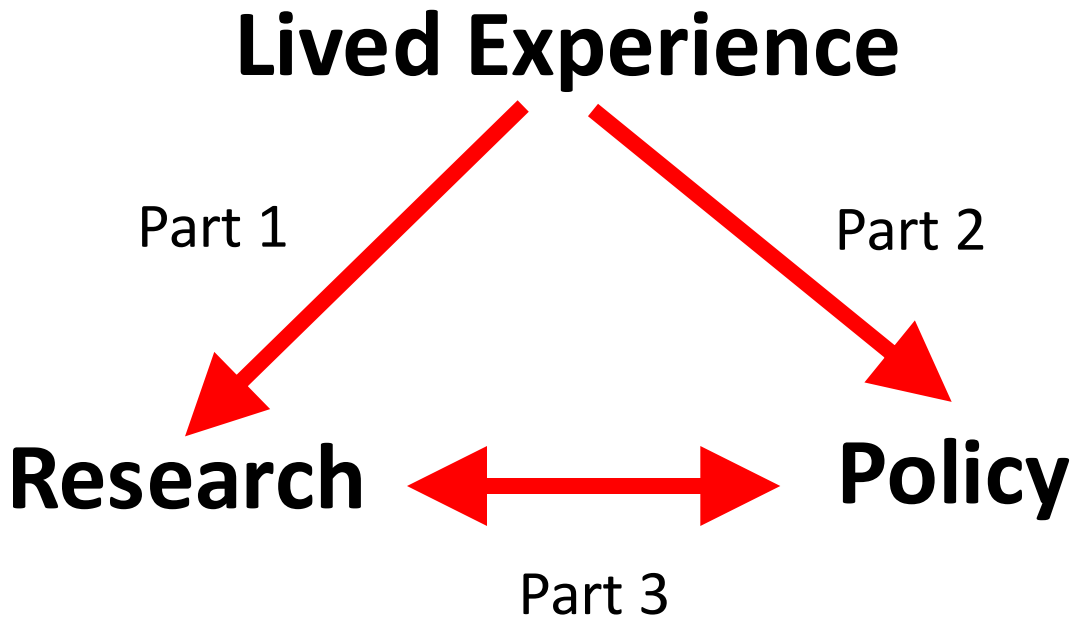
Socially Intelligent Artificial Systems Group / Civic AI Lab

Universiteit van Amsterdam

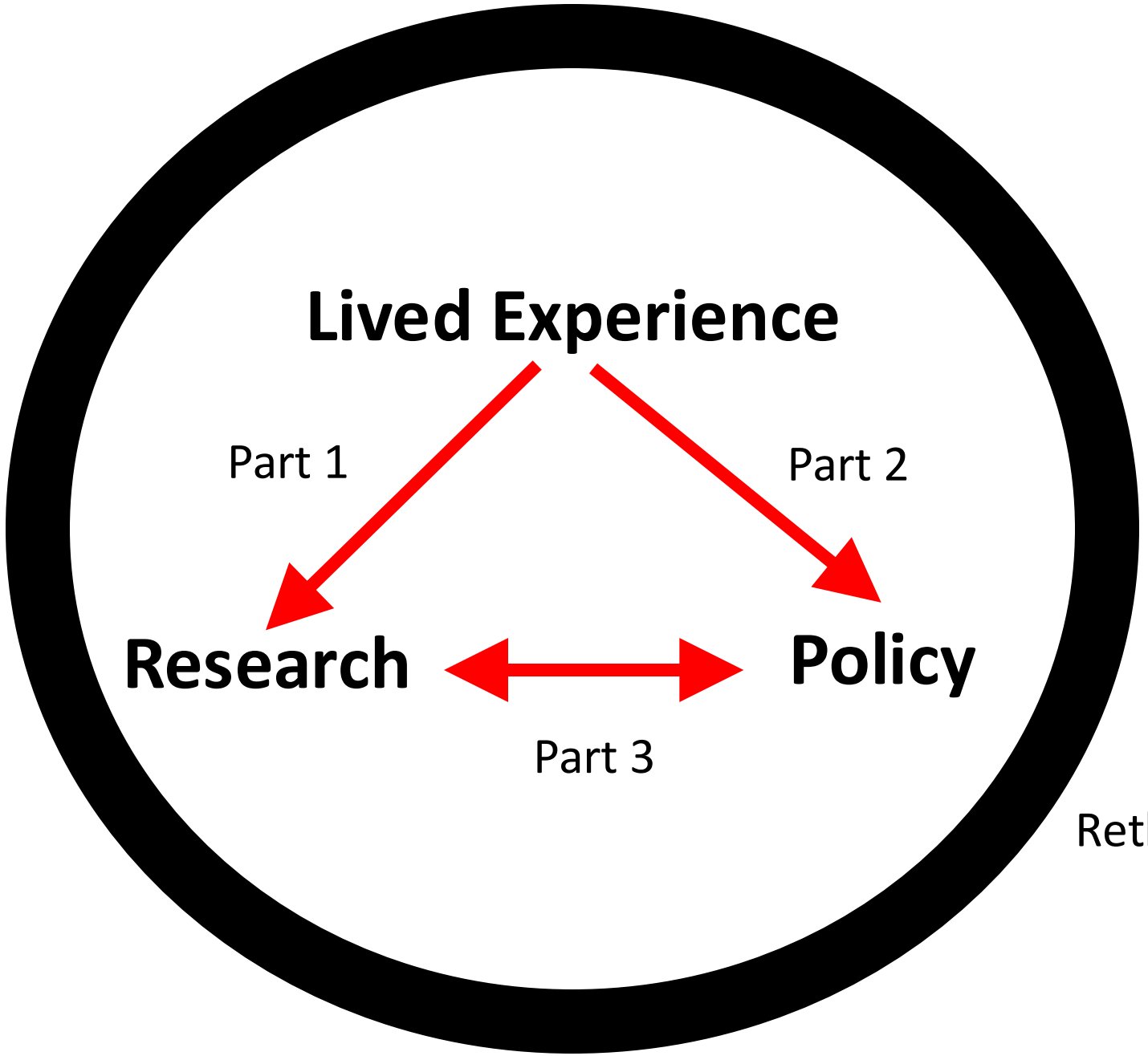
UNIVERSITEIT VAN AMSTERDAM

Socially Intelligent Artificial Systems
GROUP

Outline

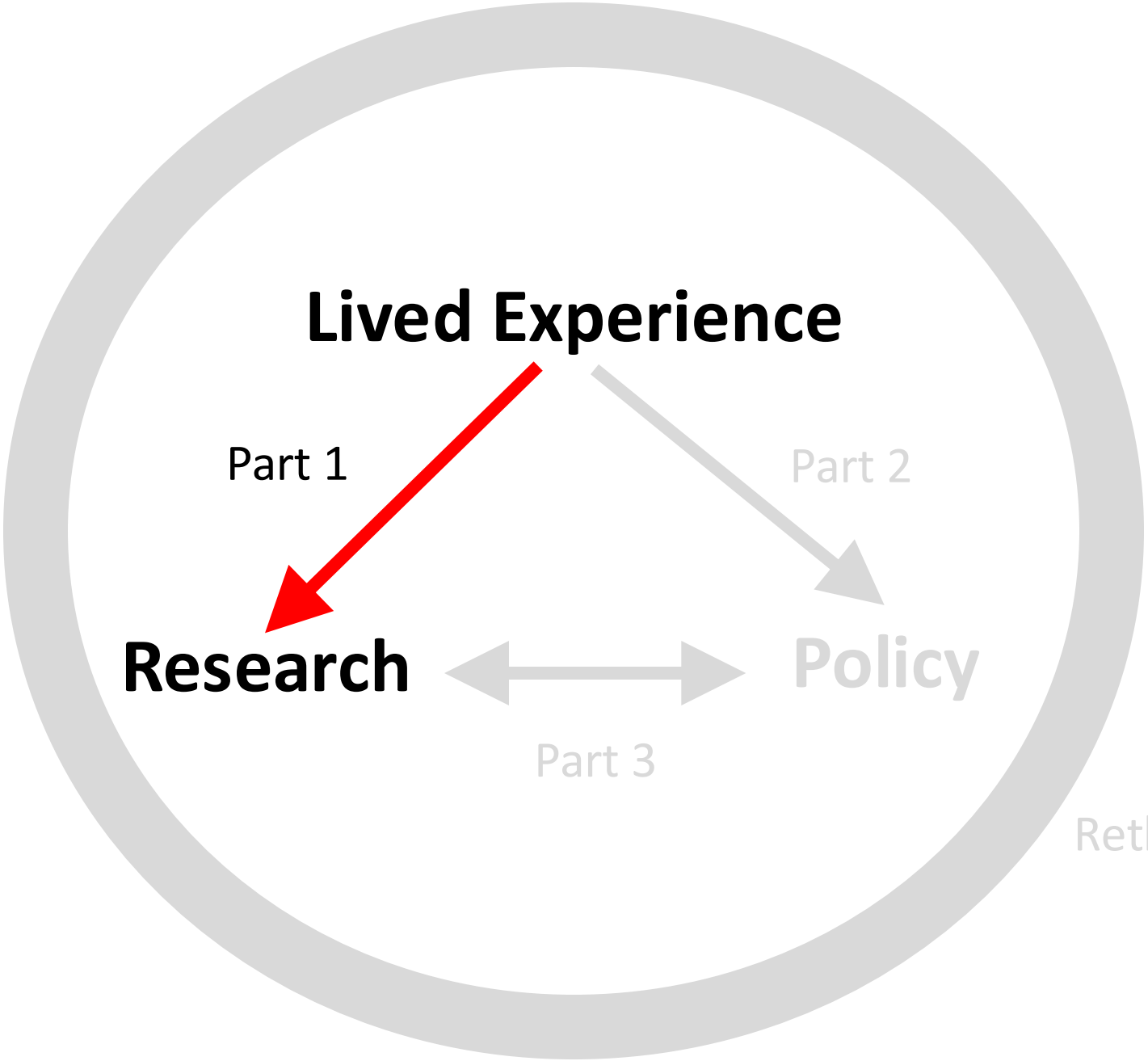


Outline



Part 4
Rethinking Question
Zero in AI

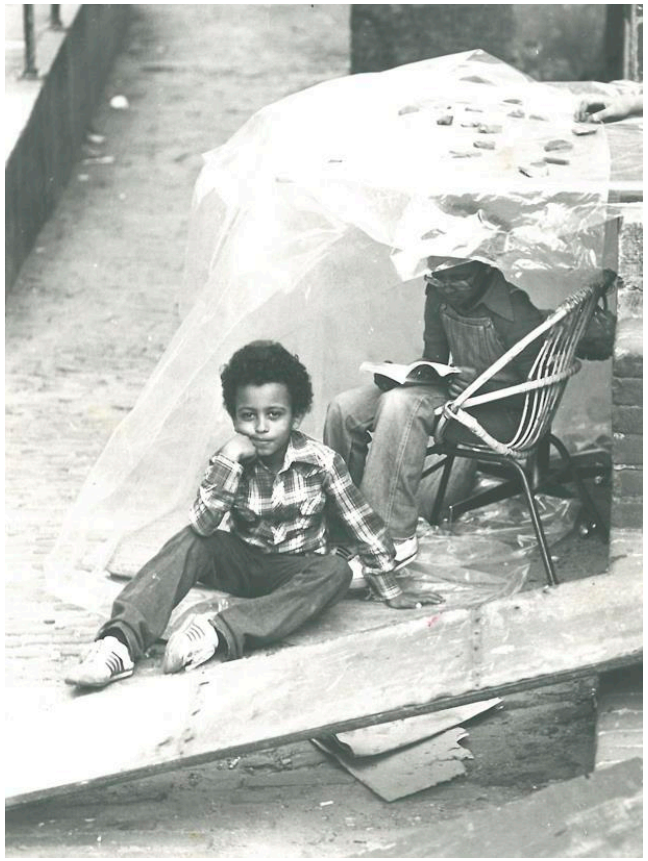
Part 1



Part 4
Rethinkling Question
Zero in AI

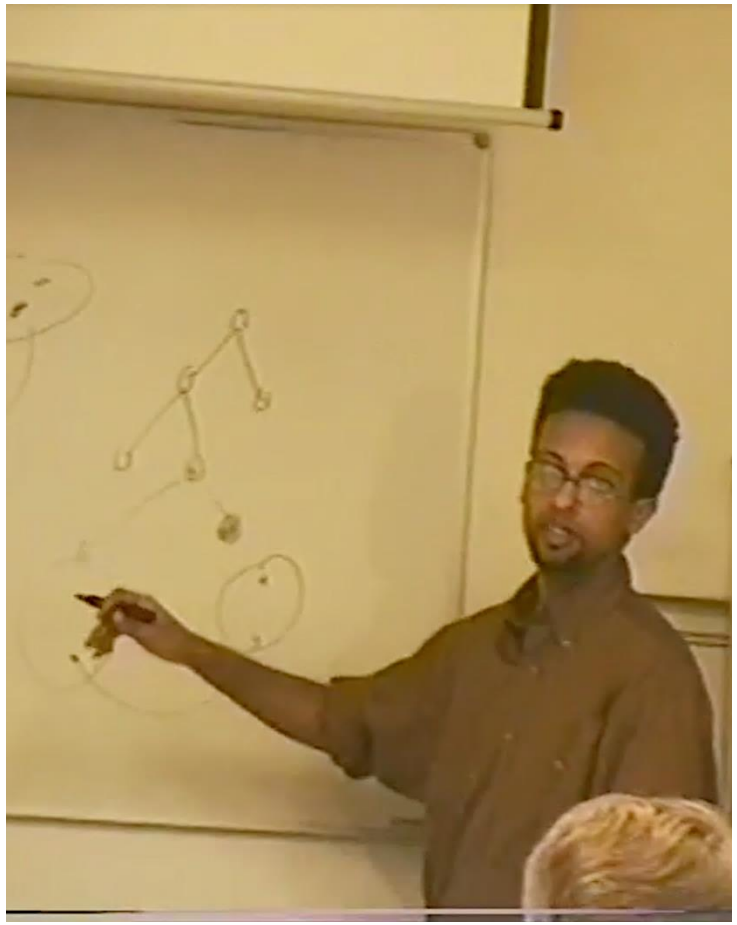
Early years

Sahel, Eritrea, 1979



Schiedam, NL, 1980

Computer science



Amsterdam, 1996

SCHEMed: a visual database tool for definition and entry of medical image data

S. Ghebreab* M. Worring
Intelligent Sensory Information Systems, University of Amsterdam
Kruislaan 403, 1098 SJ Amsterdam, The Netherlands

H.D. Tagare C.C. Jaffe
Image Processing and Analysis Group, Yale University

E-mail: worring@wins.uva.nl

Abstract

Image information systems, particularly in the medical field, require methods for defining and entering the geometric entities and their properties into the database system. As the geometric entities can be defined in different ways, where definitions can also be implicit, traditional databases and their associated schemes do not apply. In this contribution we argue why object-oriented systems are most appropriate and show how schemes should be defined to meet the demands (medical) image information systems pose. It will turn out that a good way is to divide the database scheme in a semantic and a computational part, based on a geometric algebra. We have developed a visual tool for defining and database. This tool facilitates data entry, in the mean time enforcing the consistency of the database.

keywords: *visual data modeling, visual support environment, feature extraction, visual data entry, image interaction, database schemes*

1. Introduction

Image information systems are playing an increasingly important role in medical practice. The main application being for the purpose of comparison and statistics. For example in the field of cardiology, we can expect comparison queries like "Give me all patients which have a similarly shaped left ventricle" and statistics queries like "What is the average width of the septum for normal patients".

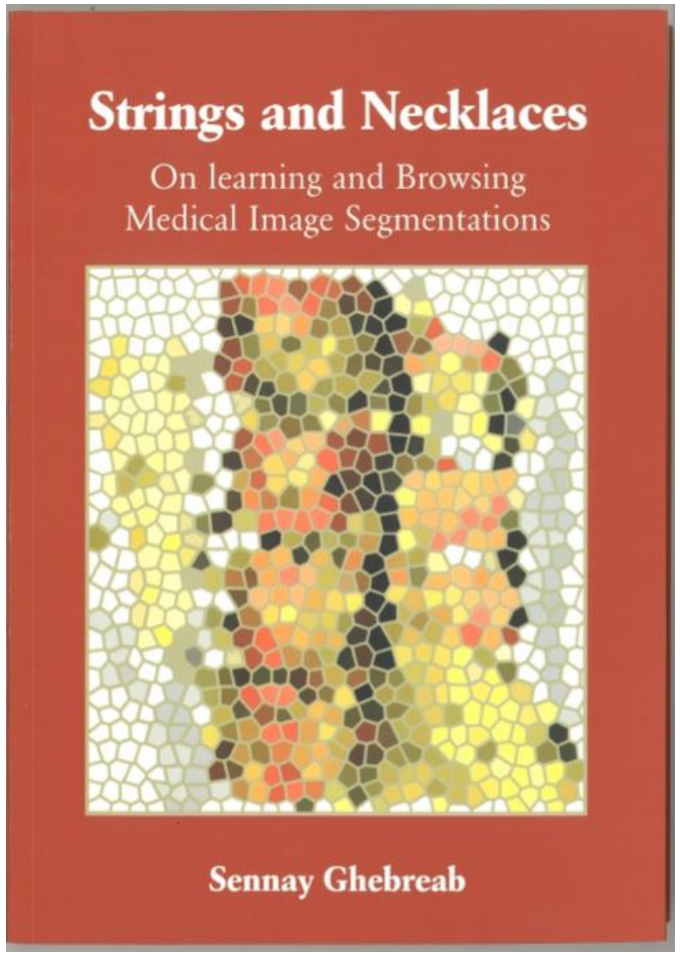
Medical image information systems as compared to other image information systems like [6][12] are special, due to a number of the characteristics of the information used [9].

Firstly, much of the relevant information is *geometric*, relating to properties of the individual entities (organs, bones, muscles etc.) in the image.

Secondly, the extensive domain knowledge is defined in vague semantic terms i.e. no fixed rules can be given to define entities and properties. As a consequence, the definitions are often *imprecise*. To improve on this, the semantic information should be translated to proper geometric infor-

VIS Conference, 1997

Computer vision



PhD thesis, 2001

Strings: Variational Deformable Models of Multivariate Continuous Boundary Features

Sennay Ghebreab and Arnold W.M. Smeulders, *Senior Member, IEEE*

Abstract—We propose a new image segmentation technique called strings. A string is a variational deformable model that is learned from a collection of example objects rather than built from a priori analytical or geometrical knowledge. As opposed to existing approaches, an object boundary is represented by a one-dimensional multivariate curve in functional space, a feature function, rather than by a point in vector space. In the learning phase, feature functions are defined by extraction of multiple shape and image features along continuous object boundaries in a given learning set. The feature functions are aligned, then subjected to functional principal components analysis and functional principal regression to summarize the feature space and to model its content, respectively. Also, a Mahalanobis distance model is constructed for evaluation of boundaries in terms of their feature functions, taking into account the natural variations seen in the learning set. In the segmentation phase, an object boundary in a new image is searched for with help of a curve. The curve gives rise to a feature function, a string, that is weighted by the regression model and evaluated by the Mahalanobis model. The curve is deformed in an iterative procedure to produce feature functions with minimal Mahalanobis distance. Strings have been compared with active shape models on 145 vertebra images, showing that strings produce better results when initialized close to the target boundary, and comparable results otherwise.

Population-Based Incremental Interactive Concept Learning for Image Retrieval by Stochastic String Segmentations

Sennay Ghebreab*, C. Carl Jaffe, and Arnold W. M. Smeulders

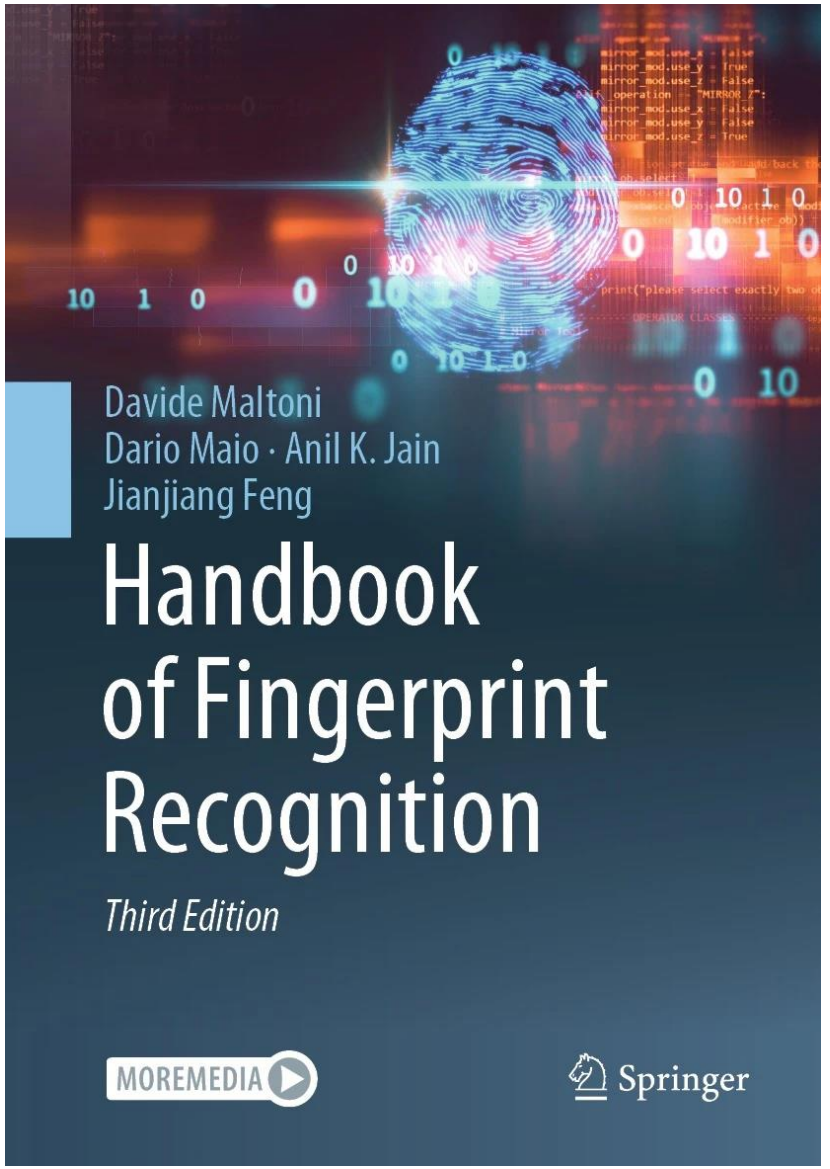
***Abstract*—We propose a method for concept-based medical image retrieval that is a superset of existing semantic-based image retrieval methods. We conceive of a concept as an incremental and interactive formalization of the user's conception of an object in an image. The premise is that such a concept is closely related to a user's specific preferences and subjectivity and, thus, allows to deal with the complexity and content-dependency of medical image content. We describe an object in terms of multiple the stochastic characteristics of an object population. A population-based incrementally learning technique, in combination with relevance feedback, is then used for concept customization. The user determines the speed and direction of concept customization using a single parameter that defines the degree of exploration and exploitation of the search space. Images are retrieved from a database in a limited number of steps based upon the customized**

therefore, it almost always disregards subtle visual patterns. When all information of the written record is exhausted can we still utilize subtle visual patterns to relate or discover small phenomena? In this paper, we strive to design an adequate content-based image retrieval method that takes into account the demands created by medical images.

Tagare *et al.* [1] postulate that 1) medical images need to be dexed by object rather than by image features, 2) object fea- res need to be formalized iteratively rather than at only one stance in order to deal with the complex and imprecise con- nt of medical images, and 3) object features need to be for- alized interactively rather than automatically to deal with the subjectivity and context-relatedness of medical image content.

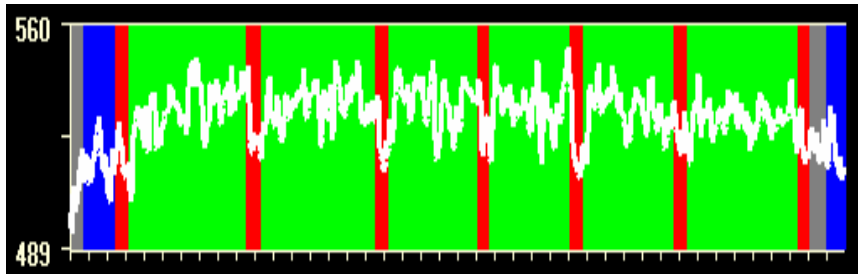
Vision competitions

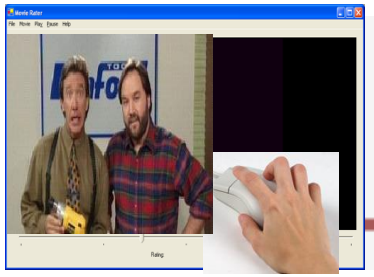
FVC Competition, 2000

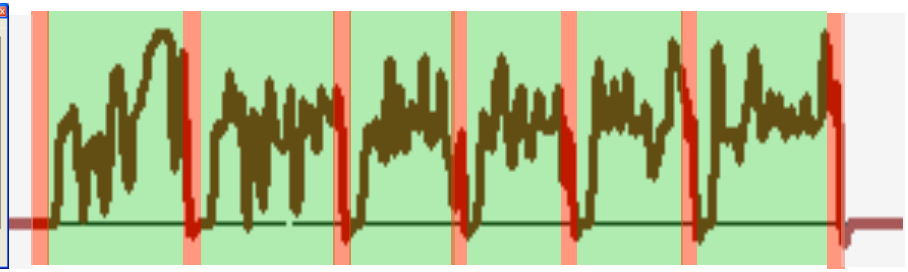


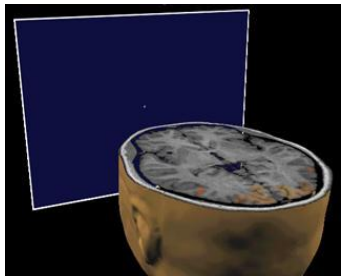
Computer & human vision

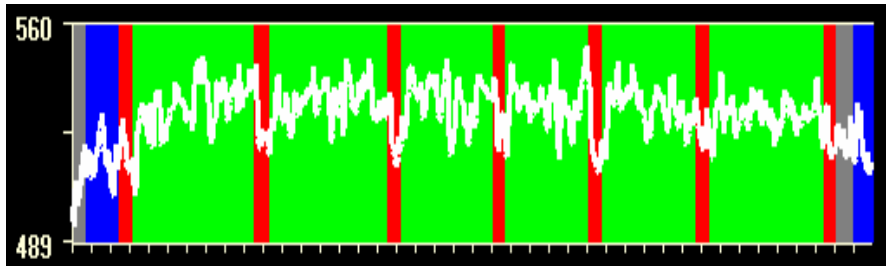
2006 Pittsburgh Brain Activity Interpretation Competition











EDITORIAL

nature neuroscience

What's on your mind?

From stage magicians to the Vulcan mind meld, mind reading has been the stuff of magic and science fiction. Recent developments in neuroimaging might be bringing it one step closer to reality, however, as increasingly sophisticated analysis techniques move toward the decoding of mental states from functional imaging data in humans. Two companies even offer fMRI-based lie detectors[1]. Although legal applications are still premature, these pattern classification techniques represent a new way of looking at neuroimaging data[2] and may extend the power of functional imaging substantially.

Conventional neuroimaging analysis correlates external regressors such as task condition with activity in specific brain areas. Pattern classification inverts this methodology and instead predicts the external stimulus based on neuroimaging data. Unlike conventional analyses, these pattern-based analyses take into account the full spatial pattern of brain activity rather than concentrating on specific regions. Thus, even if activity at a particular voxel does not distinguish different cognitive states, the pattern of activity distributed over many regions can do so, increasing sensitivity. This multivariate approach generates pattern vectors corresponding to specific cognitive states, and a classifier is trained to discriminate between these states. This classifier can then be used to predict the cognitive state on the basis of brain activity alone. Such approaches have been used to predict what percept is dominant in a binocular rivalry protocol[3] or what orientation subjects are viewing[4], even when they are not consciously aware of the stimulus.

These techniques make it easier to evaluate responses to naturalistic stimuli, as pattern classification algorithms are designed to analyze activity over the whole brain without attempting to localize function. In a recent competition at the University of Pittsburgh (http://www.ebc.pitt.edu/competition.html), participants were given fMRI data and subjective ratings from observers as they viewed two short film clips. Competitors then had to produce an algorithm that could predict what the subjects were seeing based on a third fMRI data set. The winning entries achieved correlations as high as 0.86 for basic features such as the presence of music. Uri Hasson from New York University, one of the researchers who judged the competition, says, "I am much more optimistic as to the power of fMRI to read and predict human experience. Many of the participants

as language, music, emotion), as well as the specific observers who coded the movie. Moreover, a few groups managed to predict the identity of the actor being seen or the location the subject is watching."

The power of this approach extends beyond just predicting cognitive

clues about how this information is processed as well. For example, this technique shows that object categories with shared image-based attributes have shared neural representation, even when multiple views of objects are included or when line drawings without much detail are used[5]. This kind of information would be difficult to uncover using a conventional analysis technique. James Haxby of Princeton University says, "I find myself working with a whole new community of people: computer scientists, electrical engineers, scientists who come from an applied physics and mathematics background. The fact that we are finally looking for patterns of activity across the whole population rather than areas with different function means that we are taking a more information-based approach. This is a fundamental shift."

Pattern classification can also overcome another criticism of functional imaging: that it lacks the temporal resolution of other whole-brain imaging techniques such as EEG. Pattern classification has been used to identify distributed patterns of activity associated with different categories of objects[6]. During later free recall, the patterns of activity associated with the specific object category reappear several seconds before the verbal recall of the object. The greater sensitivity of pattern classification techniques compared to conventional imaging analysis results in a temporal resolution approaching that of EEG.

As yet, however, these techniques cannot generate the sort of brain maps that we all know from conventional neuroimaging studies, showing spots of localized activity associated with particular functions. Efforts are underway to develop such maps based on pattern classification techniques, but conventional image analysis is likely to remain the preferred method for generating activity maps to understand where in the brain a process is occurring. Localization via conventional brain imaging complements the information about how a process is occurring that can be provided by pattern classification techniques, so both approaches will continue to be useful.

Researchers are attempting to use pattern classification techniques to predict brain states in real-world applications, such as lie detection, but this endeavor seems much less promising. One problem is that activity is more likely to be variable but, more importantly, an fMRI lie detector would rely crucially on the compliance of its subjects. To train a classifier to categorize lying and truth telling, a suspect would essentially be asked to calibrate the instrument for his own conviction, potentially violating the fifth amendment to the United States Constitution, which protects people from being forced to incriminate themselves. Even though the accuracy of the technology is likely to improve, it is unclear if such an fMRI lie detector can surpass conventional polygraph and EEG lie detectors, with which it is likely to share drawbacks such as oise introduced by arousal or emotional responses.

Although the applicability of pattern classification techniques to lie etection is uncertain, their influence on basic research is likely to be mportant. Neuroimaging's obsession with localization has often led to ccusations that it is little more than phrenology. By using population sponses across the whole brain to ask how rather than where informa- tion is processed, neuroimaging may be starting to come of age. ■

1. Pearson, H. *Nature* **441**, 918–919 (2006).
2. Haynes, J.-D. & Rees, G. *Nat. Rev. Neurosci* **7**, 523–534 (2006).
3. Haynes, J.-D. & Rees, G. *Nat. Neurosci.* **8**, 686–691 (2005).
4. Kamitani, Y. & Tong, F. *Nat. Neurosci.* **8**, 679–685 (2005).
5. O'Toole, A.J., Jiang, F., Abdi, H. & Haxby, J.V. *J. Cogn. Neurosci.* **17**, 580–590 (2005).
6. Polyn, S.M., Natu, V.S., Cohen, J.D. & Norman, K.A. *Science* **310**, 1963–1966 (2005).

Ethical concerns

Subject: important mail
Date: February 2007

Hi,

My name is Sridhar Putta and am writing from a place called kakinada, Andhrapradesh, India. I am certain that the research you are doing (brain scan-reading peoples intention etc) is illegally being tested or used on me. I don't know of other research people involved in this type of research. Please inquire and ask them to stop it immediately as I do not want it to happen to any other human being.

best regards,
Sridhar.

The New York Times

India's Novel Use of Brain Scans in Courts Is Debated

Share full article

By Anand Giridharadas
Sept. 14, 2008

MUMBAI, India The new technology is, to its critics, Orwellian. Others view it as a silver bullet against terrorism that could render waterboarding and other harsh interrogation methods obsolete. Some scientists predict the end of lying as we know it.

Now, well before any consensus on the technology's readiness, India has become the first country to convict someone of a crime relying on evidence from this controversial machine: a brain scanner that produces images of the human mind in action and is said to reveal signs that a suspect remembers details of the crime in question.

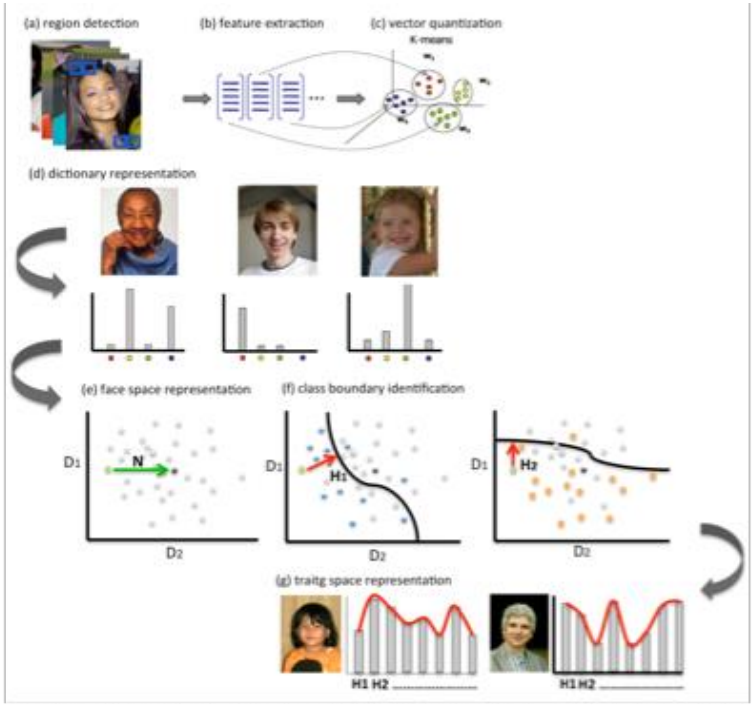
For years, scientists have peered into the brain and sought to identify deception. They have shot infrared beams through liars' heads, placed them in giant magnetic resonance imaging machines and used scanners to track their eyeballs. Since the Sept. 11 attacks, the United States has plowed money into brain-based lie detection in the hope of producing more fruitful counterterrorism investigations.

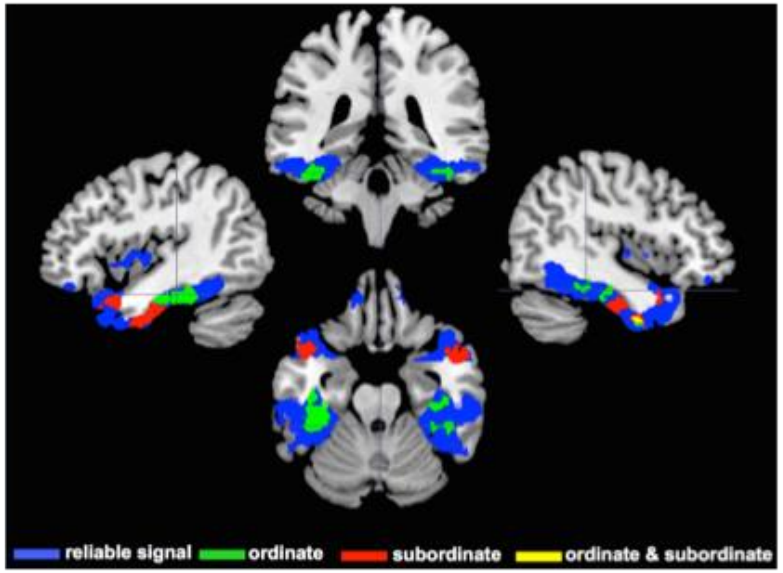
Face recognition

Title:
Computational evidence for the neural face perception system

Authors
S. Ghebreab[1,2], C.G.M. Snoek[1], V.A.F. Lamme[2], A.W.M. Smeulders[1] , H.S. Scholte[2]

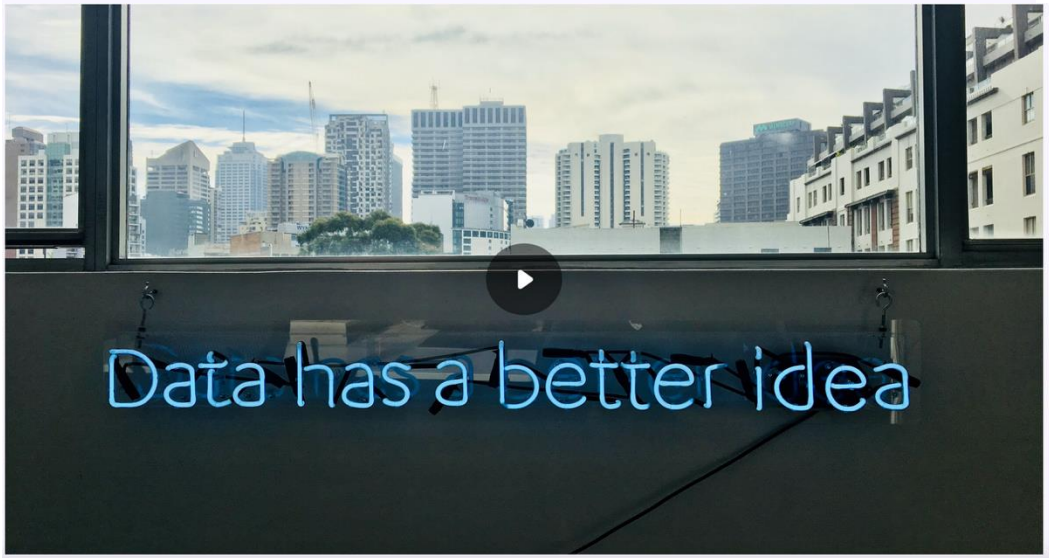
Abstract
Neuroimaging evidence has shown that a network of face sensitive brain regions underlies the ability of humans to detect and identify faces effortlessly. The network differentiates between core and extended brain regions, assumed to process low and high-level face features respectively. Various neural models have been proposed to account for the two networks, but hitherto none has been successful. Here we resort to computer vision models based on visual words, which are learned intermediate features. Images are represented as bags of visual words by counting the occurrence of each of the thousands different visual words. We studied whether ordinate and subordinate face representations based on visual words account for fMRI responses in face sensitive brain regions. Our results reveal that ordinate representations explain fMRI activity in the core face network, while subordinate representations account for activity mainly in the extended face network. Furthermore, ordinate and subordinate representations overlap only in the right anterior Inferior Temporal Gyrus, suggesting this is a candidate locus of face recognition. A single computational model thus accounts for two distinct neural face perception systems, and is possible applicable to generic visual object recognition.





Social implications

Zo bevooroordeeld kunnen algoritmes zijn



Sennay Ghebreab • 12 februari 2019

In Brainwash Talks van Human delen invloedrijke denkers, schrijvers, kunstenaars en wetenschappers verrassende ideeën voor persoonlijke en maatschappelijke problemen. Deze keer neuro-informaticus Sennay Ghebreab.

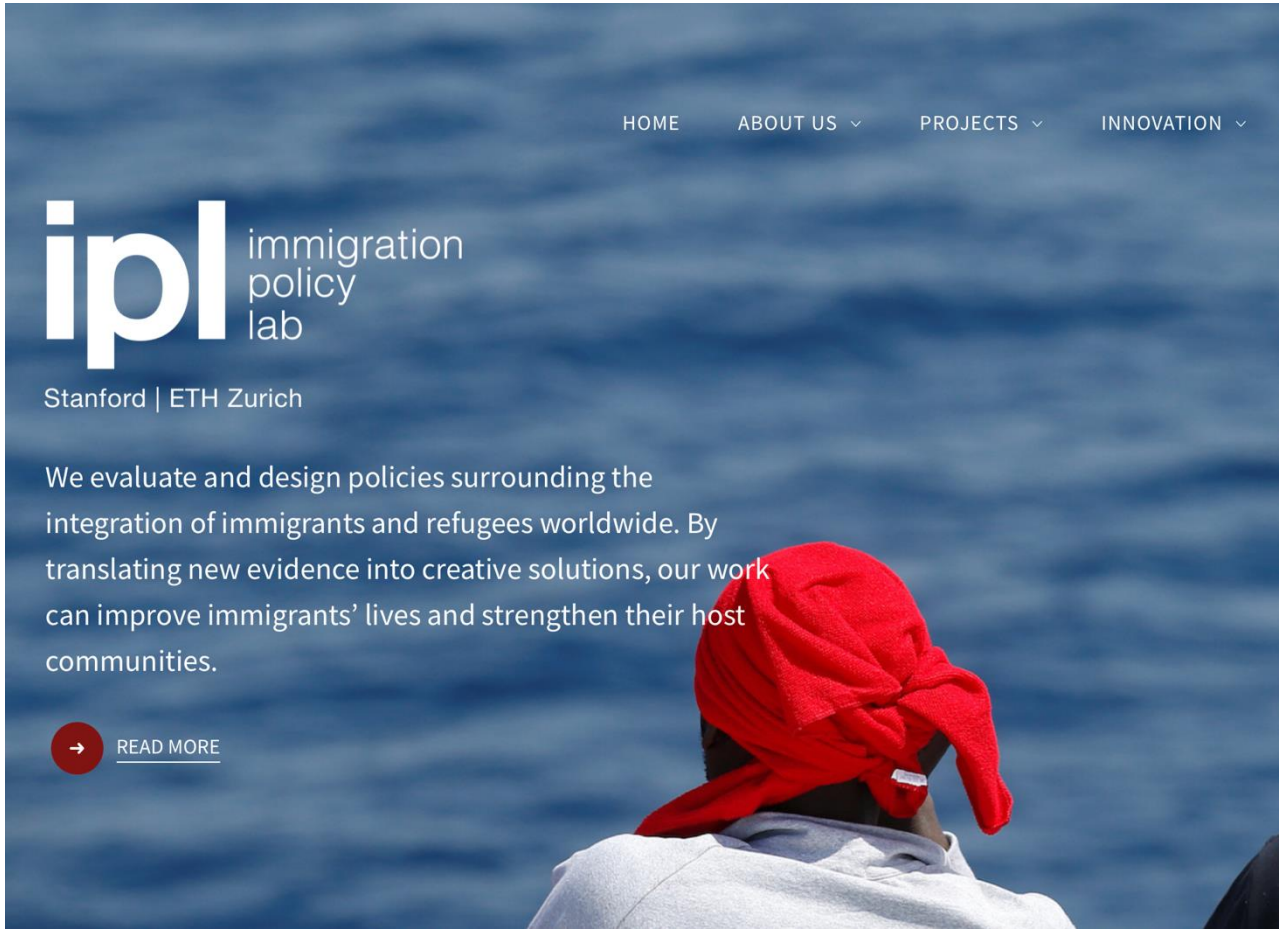
Na in 2006 een hele dag op de universiteit hebben gewerkt, besloten een aantal collega's en ik om naar huis te gaan. Het was na zessen en de balie was gesloten. Daarom moesten wij het gebouw verlaten via een kleine, automatische draaideur. Mijn collega's gingen één voor één naar buiten. Toen ik de deur naderde, stopte hij met draaien. Ik dacht: 'Er is iets mis met de draaideur.' Maar toen een andere collega het probeerde, liep ook hij moeiteloos naar buiten. Mijn collega's moesten erom lachen. Ik kreeg een ongemakkelijk gevoel van de draaideur. Een gevoel dat al snel plaatsmaakte voor nieuwsgierigheid. Waarom draaide de deur wel voor mijn collega's, maar niet voor mij? Waarom werden mijn collega's wel gezien, maar ik niet?

Neural, social and computational bias

Brain and Technology Lab
2010 (Education lab)

Civic AI Lab
2020 (Research Lab)

Algorthimic relocation



SOCIAL SCIENCE

Improving refugee integration through data-driven algorithmic assignment

Kirk Bansak,[1,3*] Jeremy Ferwerda,[2,5*] Jens Hainmueller,[1,2,4*†] Andrea Dillon,[3] Dominik Hangartner,[2,3,6] Duncan Lawrence,[2] Jeremy Weinstein[1,2]

Developed democracies are settling an increased number of refugees, many of whom face challenges integrating into host societies. We developed a flexible data-driven algorithm that assigns refugees across resettlement locations to improve integration outcomes. The algorithm uses a combination of supervised machine learning and optimal matching to discover and leverage synergies between refugee characteristics and resettlement sites. The algorithm was tested on historical registry data from two countries with different assignment regimes and refugee populations, the United States and Switzerland. Our approach led to gains of roughly 40 to 70%, on average, in refugees' employment outcomes relative to current assignment practices. This approach can provide governments with a practical and cost-efficient policy tool that can be immediately implemented within existing institutional structures.

Refugees are among the world's most vulnerable populations (*1*, *2*). After experiencing war, violence, and years of living in overcrowded refugee camps, refugees arrive in a new country with few resources and must acclimate to an unfamiliar local language, economy, and culture. Refugees frequently remain economically marginalized, with low levels of employment in the years following their arrival (*3*–*5*).

The assignment of refugees to different resettlement locations within a host country is one of the first policy decisions made during the resettlement process (*6*). It is also one of the most consequential in maximizing refugees' economic integration and self-sufficiency as a first step toward a more comprehensive integration into society (*7*–*9*). Three sets of factors affect refugee integration: geographical context, personal characteristics, and synergies between geography and personal characteristics (Fig. 1 and fig. S1). For instance, some resettlement locations in the United States offer better economic and social opportunities that can result in higher levels of refugee employment (Fig. 1A). In addition, refugees with certain characteristics, such as language and educational skills, are more likely to succeed economically regardless of the resettlement location to which they are sent (Fig. 1B). Finally, the expected employment returns associated with personal characteristics can vary across different resettlement locations (Fig. 1C). This indicates that there are synergies between places and people; certain characteristics will make a refugee a better match for a particular location. In Switzerland, for example, we find that the ability to speak French (i.e., among French-speaking African refugees) results in a larger payoff for refugees assigned to French-speaking cantons than for those assigned to German-speaking cantons (fig. S2).

Host countries' current procedures for determining how to allocate refugees across domestic resettlement sites do not fully leverage synergies between refugees and geographic locations. For instance, in the United States, refugees without existing U.S. ties are primarily assigned to resettlement locations according to the capacity of local resettlement offices at the time of arrival, without a systematic assessment of the local employment rate for refugees of similar profiles. In Switzerland, where most refugees initially enter as asylum seekers, the federal government attempts to reduce fiscal and social strain on individual localities by making assignment random and proportional across regions.

Prior research has proposed different schemes for refugee assignment both across countries (*10*, *11*) and within countries (*12*, *13*). These proposals include two-sided matching markets in which an optimized assignment is determined on the basis of match efficiency and/or the preferences of refugees and host locations (*14*). Although these approaches are theoretically appealing, there are practical barriers to their implementation, including a lack of systematic data on refugee preferences and the need for extensive political coordination.

We have developed a data-driven approach that, in contrast, can be immediately implemented by using existing data to optimize integration outcomes. Our algorithm has three stages: modeling, mapping, and matching. The modeling stage involves a supervised machine learning process that predicts the expected success for any quantifiable metric—for example, early employment—of new refugee arrivals across all possible resettlement locations. We designated historical resettlement data for model training, in which the unit of observation was a single refugee and which contained information on the refugees' background characteristics (e.g., country of origin, language skills, gender, age, etc.), time of arrival, assigned location, and measured employment success. These training data were then used to build a bundle of supervised learning models that predicted refugees' expected employment success as a function of their background characteristics. A separate model was fit for subgroups of refugees assigned to each location, thus yielding different models for each location and allowing for the discovery of refugee/location synergies. These fitted models were then applied to new, out-of-sample refugee arrival data to predict the expected employment success of each new arrival at each possible resettlement location.

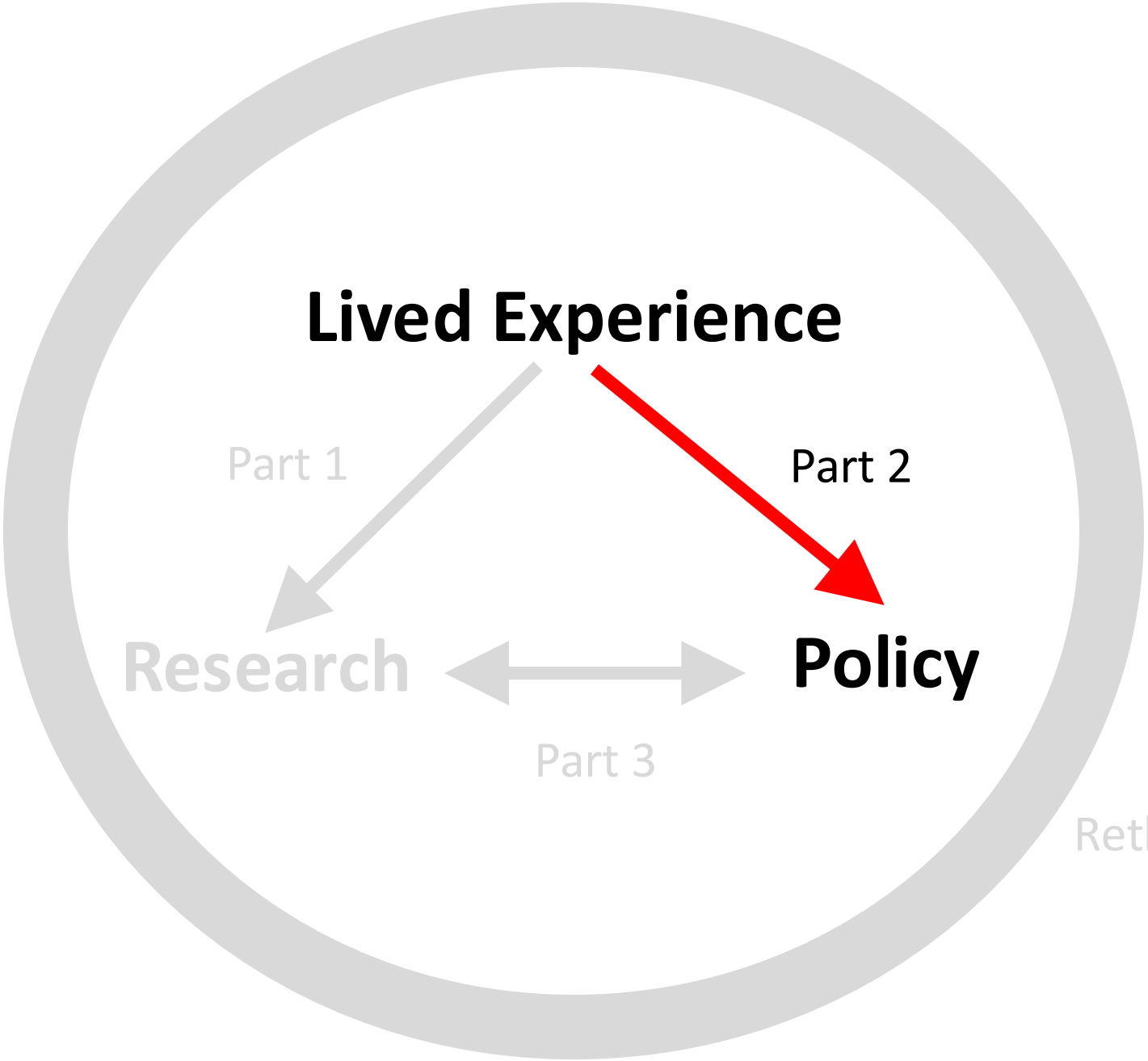
The mapping stage involves transforming the refugee-level predictions from the modeling stage to a case-level metric. Mapping to a case-level metric is necessary because refugees are often not assigned to locations on an individual basis, but rather on a case-level basis, with cases most often being family units. Various mapping functions can be used. Our preferred case-level metric was the predicted probability that at least one refugee in the case would find employment at the location in question. This metric uses a simplifying assumption that the probabilities of employment for refugees within a case are independent, although we also tested alternative mapping functions—namely the mean, maximum, and minimum predicted probability of employment within each case—that do not require this assumption (*15*).

Finally, the matching stage involves assigning each case to a specific location to fulfill a chosen optimality criterion subject to constraints. Our algorithm is flexible and can accommodate multiple criteria and constraints. The optimality criterion we used in our applications was to maximize the average of the case-level metric (i.e., the global average of the probability that at least one refugee in each family gains employment). We also imposed constraints that represent real-world assignment restrictions, such as how many cases can be sent to different locations. To solve this constrained optimization problem, we used an optimal matching procedure with the RELAX-IV minimum cost flow solver (*16*, *17*); see supplementary materials and figs. S3 to S5 for details of the algorithm, data, measures, and statistical analysis (including out-of-sample classification accuracy and probability calibration).

For the algorithm to obtain reliable predictions, it is important that the historical assignment process not be determined by unobserved refugee characteristics. This criterion is currently met in many countries that assign refugees either randomly (according to burden-sharing constraints) or according to premeasured refugee characteristics that would serve as feature inputs into the algorithm. We assessed the performance of the

[1]Department of Political Science, Stanford University, Stanford, CA 94305, USA. [2]Immigration Policy Lab, Stanford University, Stanford, CA 94305, USA, and ETH Zurich, 8092 Zurich, Switzerland. [3]Department of Government, Dartmouth College, Hanover, NH 03755, USA. [4]Graduate School of Business, Stanford University, Stanford, CA 94305, USA. [5]Center for Comparative and International Studies, ETH Zurich, 8092 Zurich, Switzerland. [6]Department of Government, London School of Economics and Political Science, London WC2A 2AE, UK.
*These authors contributed equally to this work.
†Corresponding author. Email: jhain@stanford.edu

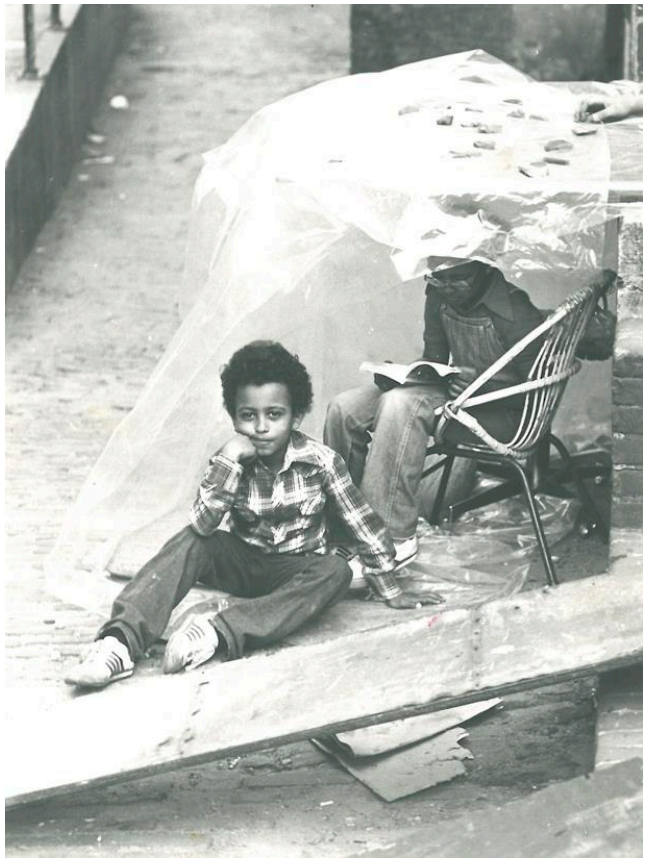
Part 2



Part 4
Rethinkling Question
Zero in AI

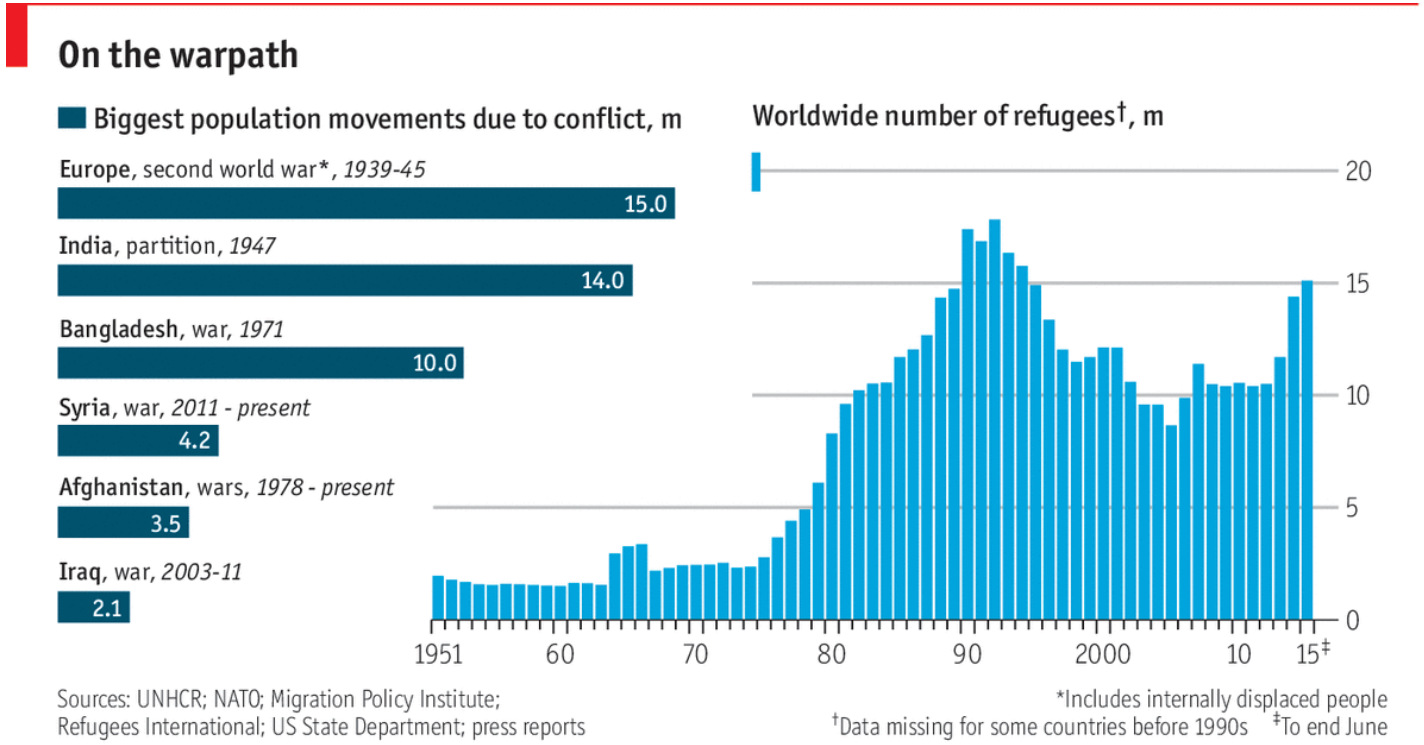
Early years

Sahel, Eritrea, 1979

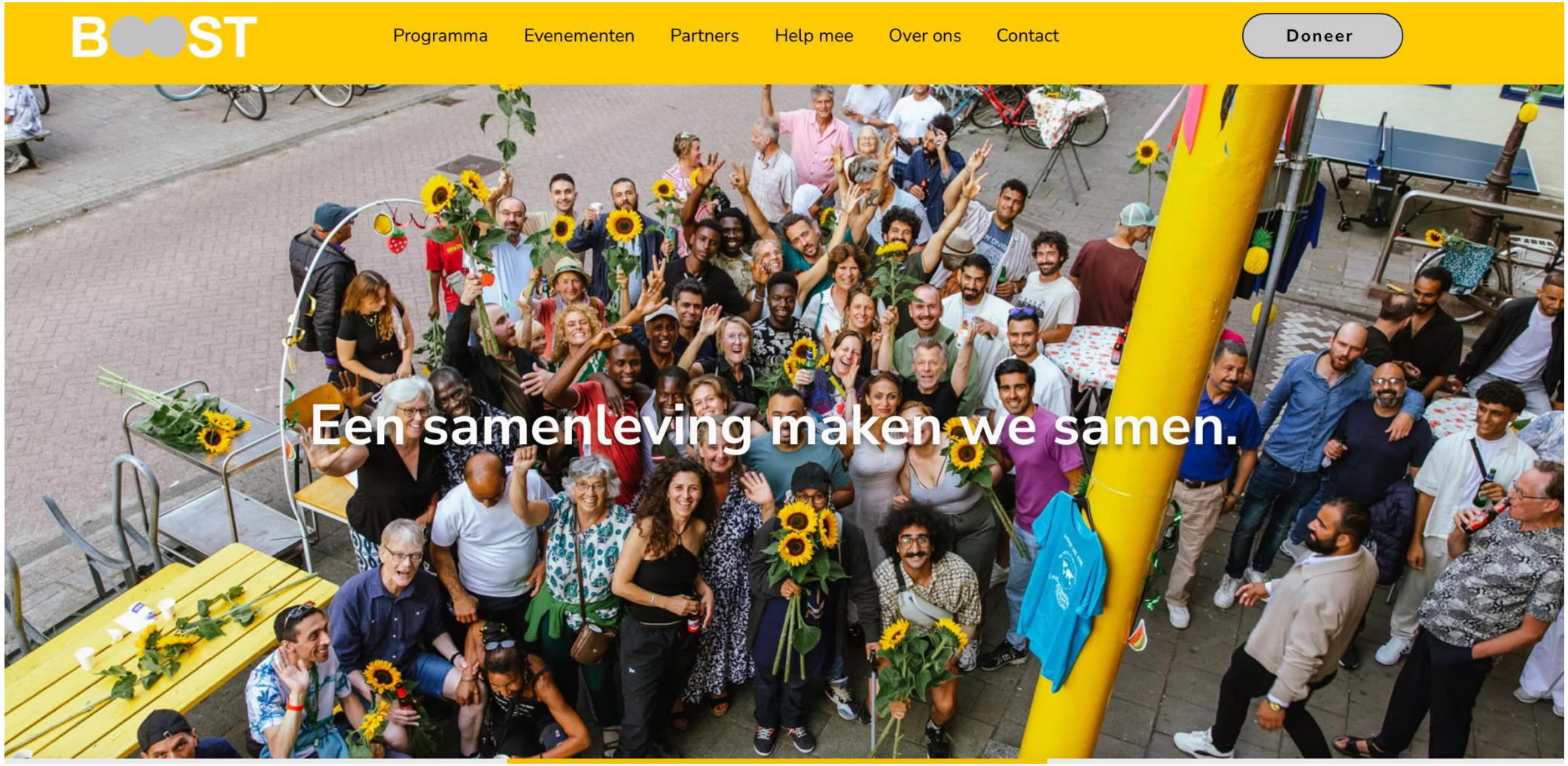


Schiedam, NL, 1980

Refugee peaks



Refugee support



Refugee reality

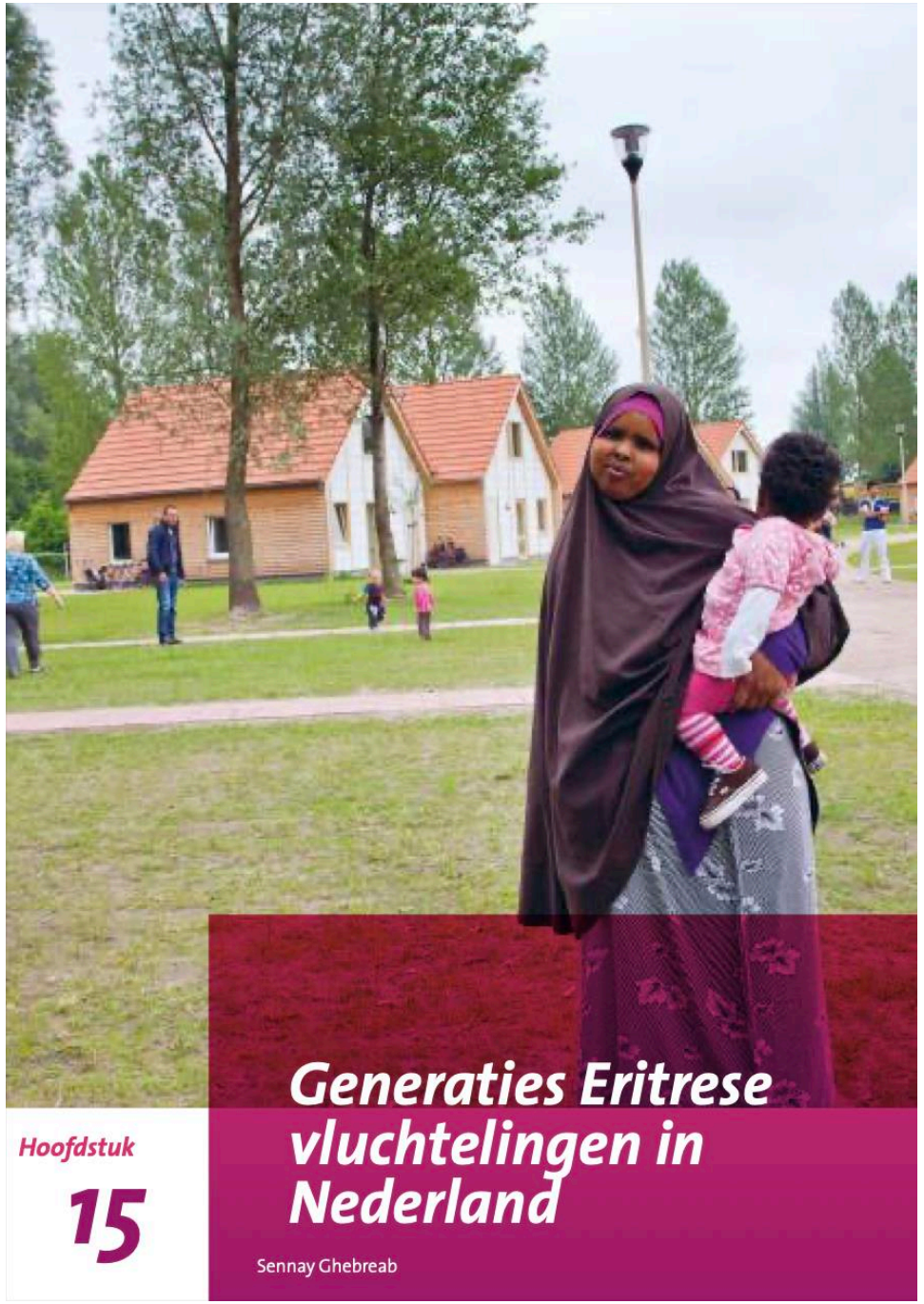
De complexe realiteit van Eritreeërs in Nederland

Dr. Sennay Ghebreab, Universiteit van Amsterdam

Sennay Ghebreab dankt het ministerie van Sociale Zaken en Werkgelegenheid, en in het bijzonder ESS, voor het organiseren van deze bijeenkomst en hij dankt gemeenten en de gemeenschap voor het samenwerken. De collectieve aanwezigheid en inspanning maakt deze unieke bijeenkomst hopelijk het begin van iets waardevols.

Persoonlijke achtergrond

Hij is geboren in Addis Abeba, de hoofdstad van Ethiopië, waar zijn ouders destijds tijdelijk studeerden en werkten. In 1975 keerden zijn ouders permanent terug naar Asmara. Toen in 1979 uitkwam dat zij onderdeel waren van de Eritrese vrijheidsbeweging, moesten zij halsoverkop op de vlucht, samen met zijn twee zussen en hem. Via Soedan en Italië kwamen ze in 1979 in Nederland terecht, als een van de eerste Eritrese gezinnen. Zijn familie en hijzelf zijn altijd maatschappelijk betrokken geweest bij de Eritrese gemeenschap in Nederland, samen met veel anderen.



Integration Policy

Policy advice

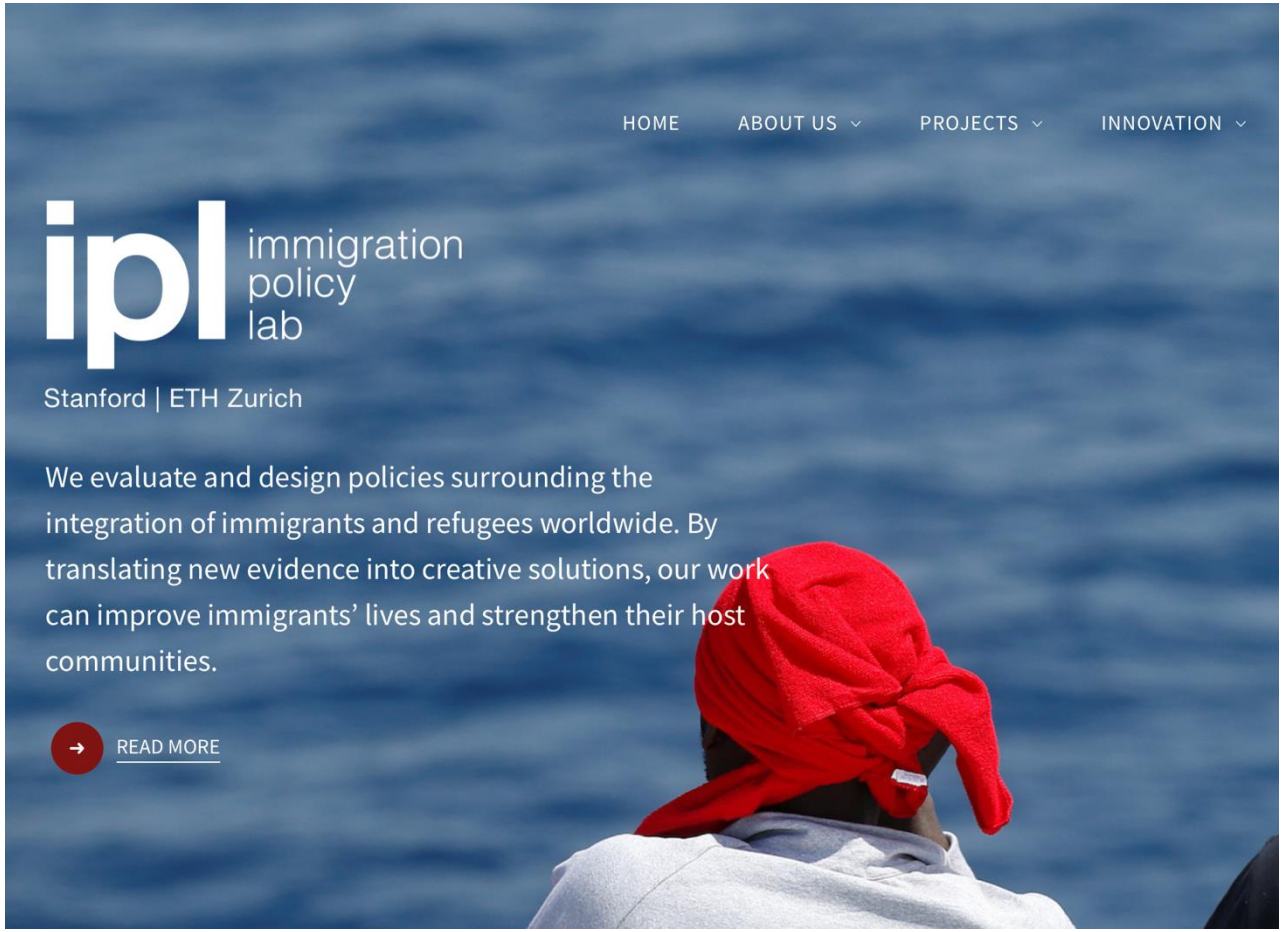
Immigration and Naturalisation Service
Ministry of Asylum and Migration

Expertise-unit Sociale Stabiliteit
Ministerie van Sociale Zaken en Werkgelegenheid

COA
Centraal Orgaan opvang asielzoekers

City of Amsterdam

Algorithmic relocation



SOCIAL SCIENCE

Improving refugee integration through data-driven algorithmic assignment

Kirk Bansak,[1,3*] Jeremy Ferwerda,[2,5*] Jens Hainmueller,[1,2,4*†] Andrea Dillon,[3] Dominik Hangartner,[2,3,6] Duncan Lawrence,[2] Jeremy Weinstein[1,2]

Developed democracies are settling an increased number of refugees, many of whom face challenges integrating into host societies. We developed a flexible data-driven algorithm that assigns refugees across resettlement locations to improve integration outcomes. The algorithm uses a combination of supervised machine learning and optimal matching to discover and leverage synergies between refugee characteristics and resettlement sites. The algorithm was tested on historical registry data from two countries with different assignment regimes and refugee populations, the United States and Switzerland. Our approach led to gains of roughly 40 to 70%, on average, in refugees' employment outcomes relative to current assignment practices. This approach can provide governments with a practical and cost-efficient policy tool that can be immediately implemented within existing institutional structures.

Refugees are among the world's most vulnerable populations (*1*, *2*). After experiencing war, violence, and years of living in overcrowded refugee camps, refugees arrive in a new country with few resources and must acclimate to an unfamiliar local language, economy, and culture. Refugees frequently remain economically marginalized, with low levels of employment in the years following their arrival (*3*–*5*).

The assignment of refugees to different resettlement locations within a host country is one of the first policy decisions made during the resettlement process (*6*). It is also one of the most consequential in maximizing refugees' economic integration and self-sufficiency as a first step toward a more comprehensive integration into society (*7*–*9*). Three sets of factors affect refugee integration: geographical context, personal characteristics, and synergies between geography and personal characteristics (Fig. 1 and fig. S1). For instance, some resettlement locations in the United States offer better economic and social opportunities that can result in higher levels of refugee employment (Fig. 1A). In addition, refugees with certain characteristics, such as language and educational skills, are more likely to succeed economically regardless of the resettlement location to which they are sent (Fig. 1B). Finally, the expected employment returns associated with personal characteristics can vary across different resettlement locations (Fig. 1C). This indicates that there are synergies between places and people; certain characteristics will make a refugee a better match for a particular location. In Switzerland, for example, we find that the ability to speak French (i.e., among French-speaking African refugees) results in a larger payoff for refugees assigned to French-speaking cantons than for those assigned to German-speaking cantons (fig. S2).

Host countries' current procedures for determining how to allocate refugees across domestic resettlement sites do not fully leverage synergies between refugees and geographic locations. For instance, in the United States, refugees without existing U.S. ties are primarily assigned to resettlement locations according to the capacity of local resettlement offices at the time of arrival, without a systematic assessment of the local employment rate for refugees of similar profiles. In Switzerland, where most refugees initially enter as asylum seekers, the federal government attempts to reduce fiscal and social strain on individual localities by making assignment random and proportional across regions.

Prior research has proposed different schemes for refugee assignment both across countries (*10*, *11*) and within countries (*12*, *13*). These proposals include two-sided matching markets in which an optimized assignment is determined on the basis of match efficiency and/or the preferences of refugees and host locations (*14*). Although these approaches are theoretically appealing, there are practical barriers to their implementation, including a lack of systematic data on refugee preferences and the need for extensive political coordination.

We have developed a data-driven approach that, in contrast, can be immediately implemented by using existing data to optimize integration outcomes. Our algorithm has three stages: modeling, mapping, and matching. The modeling stage involves a supervised machine learning process that predicts the expected success for any quantifiable metric—for example, early employment—of new refugee arrivals across all possible resettlement locations. We designated historical resettlement data for model training, in which the unit of observation was a single refugee and which contained information on the refugees' background characteristics (e.g., country of origin, language skills, gender, age, etc.), time of arrival, assigned location, and measured employment success. These training data were then used to build a bundle of supervised learning models that predicted refugees' expected employment success as a function of their background characteristics. A separate model was fit for subgroups of refugees assigned to each location, thus yielding different models for each location and allowing for the discovery of refugee/location synergies. These fitted models were then applied to new, out-of-sample refugee arrival data to predict the expected employment success of each new arrival at each possible resettlement location.

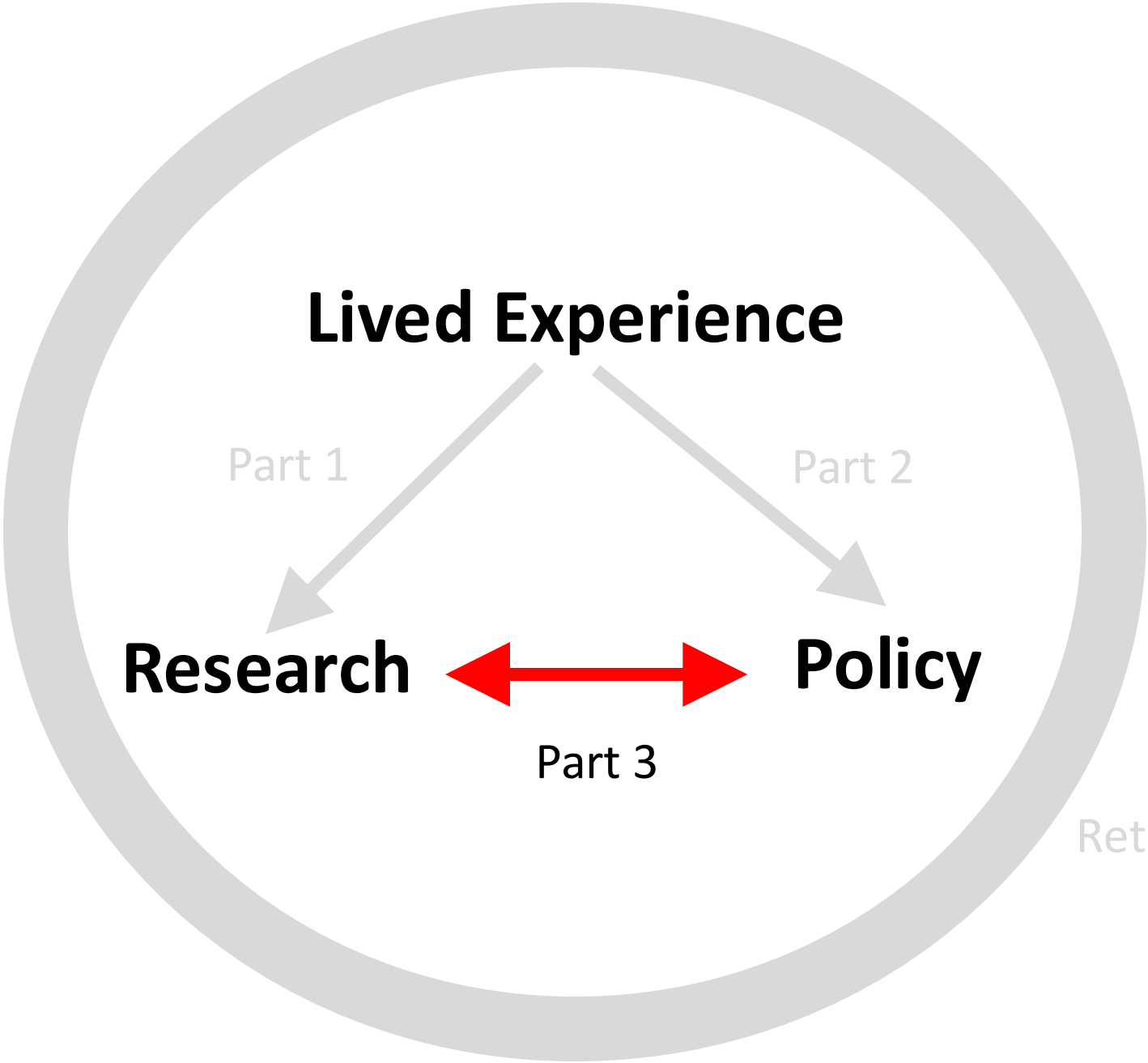
The mapping stage involves transforming the refugee-level predictions from the modeling stage to a case-level metric. Mapping to a case-level metric is necessary because refugees are often not assigned to locations on an individual basis, but rather on a case-level basis, with cases most often being family units. Various mapping functions can be used. Our preferred case-level metric was the predicted probability that at least one refugee in the case would find employment at the location in question. This metric uses a simplifying assumption that the probabilities of employment for refugees within a case are independent, although we also tested alternative mapping functions—namely the mean, maximum, and minimum predicted probability of employment within each case—that do not require this assumption (*15*).

Finally, the matching stage involves assigning each case to a specific location to fulfill a chosen optimality criterion subject to constraints. Our algorithm is flexible and can accommodate multiple criteria and constraints. The optimality criterion we used in our applications was to maximize the average of the case-level metric (i.e., the global average of the probability that at least one refugee in each family gains employment). We also imposed constraints that represent real-world assignment restrictions, such as how many cases can be sent to different locations. To solve this constrained optimization problem, we used an optimal matching procedure with the RELAX-IV minimum cost flow solver (*16*, *17*); see supplementary materials and figs. S3 to S5 for details of the algorithm, data, measures, and statistical analysis (including out-of-sample classification accuracy and probability calibration).

For the algorithm to obtain reliable predictions, it is important that the historical assignment process not be determined by unobserved refugee characteristics. This criterion is currently met in many countries that assign refugees either randomly (according to burden-sharing constraints) or according to premeasured refugee characteristics that would serve as feature inputs into the algorithm. We assessed the performance of the

[1]Department of Political Science, Stanford University, Stanford, CA 94305, USA. [2]Immigration Policy Lab, Stanford University, Stanford, CA 94305, USA, and ETH Zurich, 8092 Zurich, Switzerland. [3]Department of Government, Dartmouth College, Hanover, NH 03755, USA. [4]Graduate School of Business, Stanford University, Stanford, CA 94305, USA. [5]Center for Comparative and International Studies, ETH Zurich, 8092 Zurich, Switzerland. [6]Department of Government, London School of Economics and Political Science, London WC2A 2AE, UK.
*These authors contributed equally to this work.
†Corresponding author. Email: jhain@stanford.edu

Part 3



Part 4
Rethinkling Question
Zero in AI

COA

Centraal Orgaan opvang asielzoekers

- Dutch government agency responsible for incoming refugees
- Provides shelter in various reception centers
- Offers support and guidance to integrate or prepare for return.
- Ensures access to necessities like food (or allowance) and other care
- Works with other organizations like the Immigration and Naturalisation Service, and Repatriation and Departure Service

COA policy

- Government assigns each municipality a housing quota
- COA conducts screening interviews to assess personal/educational/medical criteria
- Status holders are placed in regions that best support future work and integration
- Municipalities provide suitable housing, coordinating with COA;

Promising match with municipality

Every half year, the central government determines the number of status holders a municipality must house: the municipal task. The COA matches status holders to a municipality. When we do this, we look at where they have the best opportunities to build a new life and contribute to society.

Screening interview

In the process reception centre and asylum seekers' centre, we have a screening interview with the refugee. During this interview we inquire after their education and work experience in the country of origin, ambitions, and social network. These are also called the 'soft criteria'. In addition, we look at hard criteria, such as first-degree relatives, medical particulars and current work and education in the Netherlands.

Placement in a suitable region for work

After the screening interview, we select the best reception centre to place the status holder. We try to accommodate him close to the place where he will be living and working later. This improves their integration and participation. Next, the supervisor matches the status holder within 2 weeks to the municipality which is best for him to go and live in.

COA practice

- Staffing challenges
- Expertise shortage
- Housing shortage in NL
- Political pressure
- Social discourse in NL
- Media framing
- Geopolitical dynamics
- …..

COA result

- Ad-hoc relocation of refugees
- Often based on capacity of places
- Circular relocation
- Emergency housing
- Refugee needs, talents etc unknown
- Needs relocation place unmet
- …….

Menu | nrc

Luister naar

Deze crisis zag het COA al jaren aankomen

Asielopvang Al heel lang waarschuwde het COA dat steeds weer op- en afschalen van opvangcapaciteit tot grote problemen zou leiden. Maar er veranderde niets, nu is de chaos compleet.

Martin Kuiper & Romy van der Poel
29 juli 2022 om 16:40 Leestijd 11 minuten

Algorithmic solution?

‘Zet kunstmatige intelligentie in om geschikte plek asielzoeker te vinden’

Jan Fred van Wijnen

Sennay Ghebreab vluchtte als kind vanuit Eritrea naar Schiedam. Hij ontwikkelt nu een computersysteem dat helpt om asielzoekers een passende plek in de samenleving te...

Ministry of Social Affairs and Employment

Ministry of Justice and Security

COA

Centraal Orgaan opvang asielzoekers

Potential benefits

SOCIAL SCIENCE

Improving refugee integration through data-driven algorithmic assignment

Kirk Bansak,[1,2]* Jeremy Ferwerda,[3,5]* Jens Hainmueller,[1,3,4]*† Andrea Dillon,[3] Dominik Hangartner,[2,3,6] Duncan Lawrence,[2] Jeremy Weinstein[1,2]

Developed democracies are settling an increased number of refugees, many of whom face challenges integrating into host societies. We developed a flexible data-driven algorithm that assigns refugees across resettlement locations to improve integration outcomes. The algorithm uses a combination of supervised machine learning and optimal matching to discover and leverage synergies between refugee characteristics and resettlement sites. The algorithm was tested on historical registry data from two countries with different assignment regimes and refugee populations, the United States and Switzerland. Our approach led to gains of roughly 40 to 70%, on average, in refugees' employment outcomes relative to current assignment practices. This approach can provide governments with a practical and cost-efficient policy tool that can be immediately implemented within existing institutional structures.

Refugees are among the world's most vulnerable populations (*1*, *2*). After experiencing war, violence, and years of living in overcrowded refugee camps, refugees arrive in a new country with few resources and must acclimate to an unfamiliar local language, economy, and culture. Refugees frequently remain economically marginalized, with low levels of employment in the years following their arrival (*3*–*5*).

The assignment of refugees to different resettlement locations within a host country is one of the first policy decisions made during the resettlement process (*6*). It is also one of the most consequential in maximizing refugees' economic integration and self-sufficiency as a first step toward a more comprehensive integration into society (*7*–*9*). Three sets of factors affect refugee integration: geographical context, personal characteristics, and synergies between geography and personal characteristics (Fig. 1 and fig. S1). For instance, some resettlement locations in the United States offer better economic and social opportunities that can result in higher levels of refugee employment (Fig. 1A). In addition, refugees with certain characteristics, such as language and educational skills, are more likely to succeed economically regardless of the resettlement location to which they are sent (Fig. 1B). Finally, the expected employment returns associated with personal characteristics can vary across different resettlement locations (Fig. 1C). This indicates that there are synergies between places and people; certain characteristics will make a refugee a better match for a particular location. In Switzerland, for example, we find that the ability to speak French (i.e., among French-speaking African refugees) results in a larger payoff for refugees assigned to French-speaking cantons than for those assigned to German-speaking cantons (fig. S2).

Host countries' current procedures for determining how to allocate refugees across domestic resettlement sites do not fully leverage synergies between refugees and geographic locations. For instance, in the United States, refugees without existing U.S. ties are primarily assigned to resettlement locations according to the capacity of local resettlement offices at the time of arrival, without a systematic assessment of the local employment rate for refugees of similar profiles. In Switzerland, where most refugees initially enter as asylum seekers, the federal government attempts to reduce fiscal and social strain on individual localities by making assignment random and proportional across regions.

Prior research has proposed different schemes for refugee assignment both across countries (*10*, *11*) and within countries (*12*, *13*). These proposals include two-sided matching markets in which an optimized assignment is determined on the basis of match efficiency and/or the preferences of refugees and host locations (*14*). Although these approaches are theoretically appealing, there are practical barriers to their implementation, including a lack of systematic data on refugee preferences and the need for extensive political coordination.

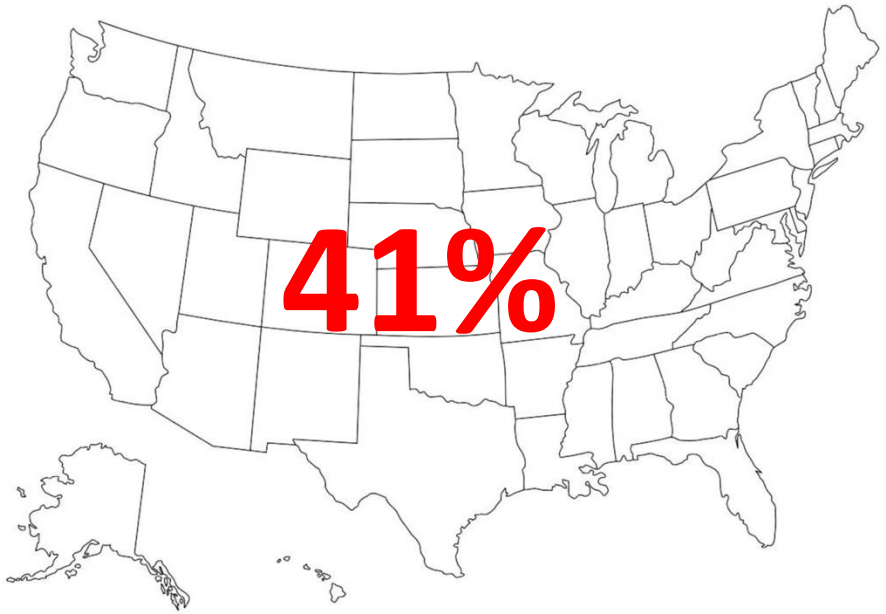
We have developed a data-driven approach that, in contrast, can be immediately implemented by using existing data to optimize integration outcomes. Our algorithm has three stages: modeling, mapping, and matching. The modeling stage involves a supervised machine learning process that predicts the expected success for any quantifiable metric—for example, early employment—of new refugee arrivals across all possible resettlement locations. We designated historical resettlement data for model training, in which the unit of observation was a single refugee and which contained information on the refugees' background characteristics (e.g., country of origin, language skills, gender, age, etc.), time of arrival, assigned location, and measured employment success. These training data were then used to build a bundle of supervised learning models that predicted refugees' expected employment success as a function of their background characteristics. A separate model was fit for subgroups of refugees assigned to each location, thus yielding different models for each location and allowing for the discovery of refugee/location synergies. These fitted models were then applied to new, out-of-sample refugee arrival data to predict the expected employment success of each new arrival at each possible resettlement location.

The mapping stage involves transforming the refugee-level predictions from the modeling stage to a case-level metric. Mapping to a case-level metric is necessary because refugees are often not assigned to locations on an individual basis, but rather on a case-level basis, with cases most often being family units. Various mapping functions can be used. Our preferred case-level metric was the predicted probability that at least one refugee in the case would find employment at the location in question. This metric uses a simplifying assumption that the probabilities of employment for refugees within a case are independent, although we also tested alternative mapping functions—namely the mean, maximum, and minimum predicted probability of employment within each case—that do not require this assumption (*15*).

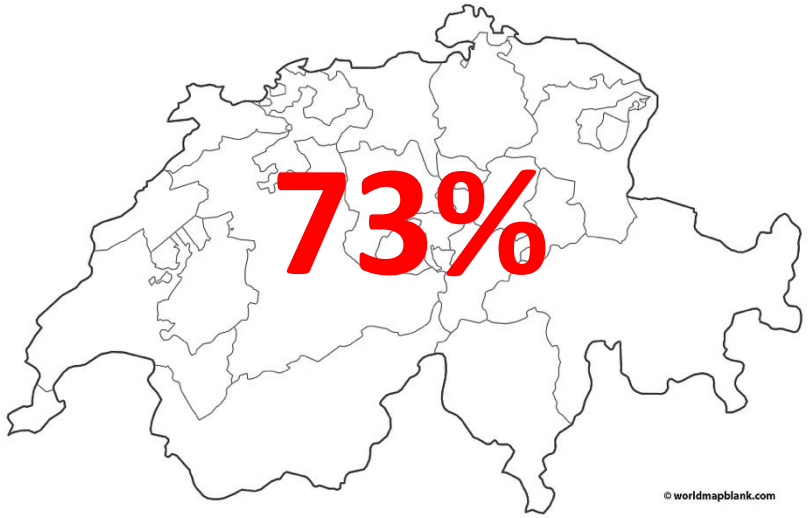
Finally, the matching stage involves assigning each case to a specific location to fulfill a chosen optimality criterion subject to constraints. Our algorithm is flexible and can accommodate multiple criteria and constraints. The optimality criterion we used in our applications was to maximize the average of the case-level metric (i.e., the global average of the probability that at least one refugee in each family gains employment). We also imposed constraints that represent real-world assignment restrictions, such as how many cases can be sent to different locations. To solve this constrained optimization problem, we used an optimal matching procedure with the RELAX-IV minimum cost flow solver (*16*, *17*); see supplementary materials and figs. S3 to S5 for details of the algorithm, data, measures, and statistical analysis (including out-of-sample classification accuracy and probability calibration).

For the algorithm to obtain reliable predictions, it is important that the historical assignment process not be determined by unobserved refugee characteristics. This criterion is currently met in many countries that assign refugees either randomly (according to burden-sharing constraints) or according to premeasured refugee characteristics that would serve as feature inputs into the algorithm. We assessed the performance of the

[1]Department of Political Science, Stanford University, Stanford, CA 94305, USA. [2]Immigration Policy Lab, Stanford University, Stanford, CA 94305, USA, and ETH Zurich, 8092 Zurich, Switzerland. [3]Department of Government, Dartmouth College, Hanover, NH 03755, USA. [4]Graduate School of Business, Stanford University, Stanford, CA 94305, USA. [5]Center for Comparative and International Studies, ETH Zurich, 8092 Zurich, Switzerland. [6]Department of Government, London School of Economics and Political Science, London WC2A 2AE, UK.
*These authors contributed equally to this work.
†Corresponding author. Email: jhain@stanford.edu

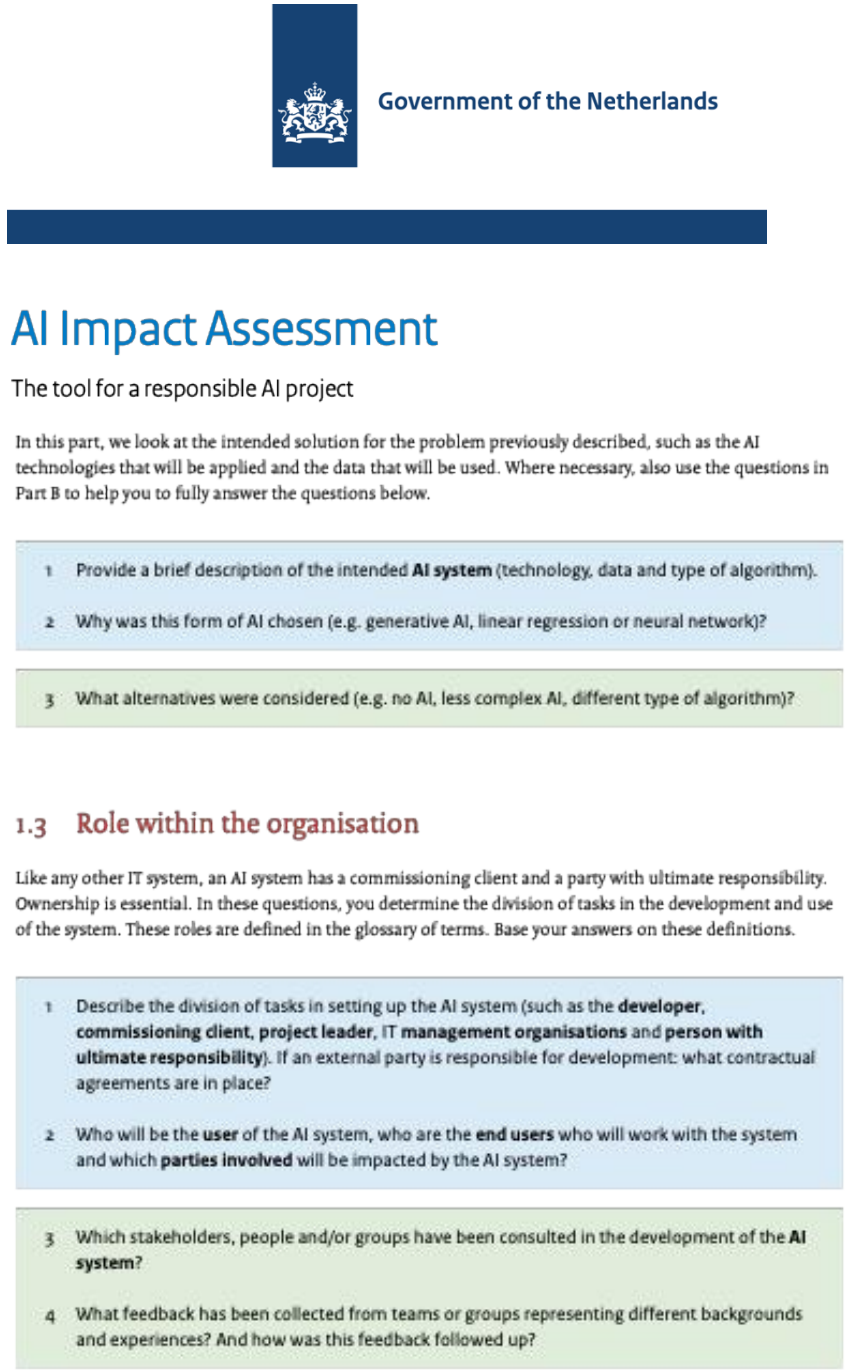


Increase in employability rates



COA questions

- COA turned out to already explore AR
- And to ask sensible questions
  - Why do we need an algorithmic solutions?
  - Can we not rely on humans only?
  - Who is responsible for relocation outcomes?
  - How do we ensure privacy of refugee data
  - Etc, etc
- In spirit of DP & AI assessment question



Government of the Netherlands

AI Impact Assessment

The tool for a responsible AI project

In this part, we look at the intended solution for the problem previously described, such as the AI technologies that will be applied and the data that will be used. Where necessary, also use the questions in Part B to help you to fully answer the questions below.

1 Provide a brief description of the intended **AI system** (technology, data and type of algorithm).

2 Why was this form of AI chosen (e.g. generative AI, linear regression or neural network)?

3 What alternatives were considered (e.g. no AI, less complex AI, different type of algorithm)?

1.3 Role within the organisation

Like any other IT system, an AI system has a commissioning client and a party with ultimate responsibility. Ownership is essential. In these questions, you determine the division of tasks in the development and use of the system. These roles are defined in the glossary of terms. Base your answers on these definitions.

1 Describe the division of tasks in setting up the AI system (such as the **developer**, **commissioning client**, **project leader**, IT **management organisations** and **person with ultimate responsibility**). If an external party is responsible for development: what contractual agreements are in place?

2 Who will be the **user** of the AI system, who are the **end users** who will work with the system and which **parties involved** will be impacted by the AI system?

3 Which stakeholders, people and/or groups have been consulted in the development of the **AI system**?

4 What feedback has been collected from teams or groups representing different backgrounds and experiences? And how was this feedback followed up?

Research suggestions

- Research-policy collaboration, across discilines and domains
- Build algorithmic solutions locally, not rely on
- Focus on long term, not short term political cycle
- Optimize the algorithm not only for employabiliy
- Center on wellbeing, education and other social aspects
- Consult and cooperate with refugees themselves

- Let’s start the Civic AI Lab

Follow-up

- No follow-up on research-policy collaboration
- COA decided to start a pilot independently
- Requested participation in advisory board
- Never approached for advice

COA Centraal Orgaan opvang asielzoekers

Opvang en begeleiding | Bijdrager

Procedures en regelgeving | Locaties

AI zorgt voor betere match statushouder aan gemeente

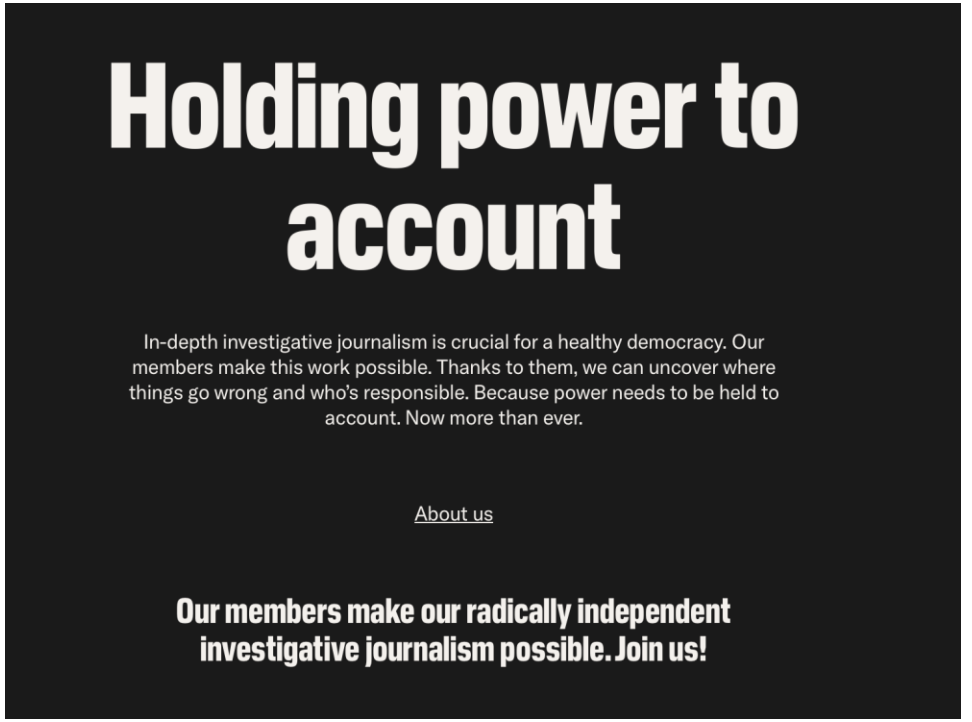
Geplaatst op 25 april 2023

Het bestaande project Kansrijke Koppeling is bedoeld om statushouders sneller en beter naar betaald werk te leiden. Welke rol kan artificiële intelligentie (AI) hierin spelen? Kunnen algoritmen COA-medewerkers helpen om statushouders beter te matchen met een regio waar de kans op werk groter is? Die vraag stond centraal in een uitgebreid en zorgvuldig onderzoek dat in 2022 is afgerond. In oktober ging de pilotfase van start.

Aanleiding voor het onderzoek was een aanbeveling van het Centraal Planbureau (CPB) in 2018 om kansrijke koppeling te versterken met data gedreven toewijzing en algoritme. Het CBP putte uit internationaal onderzoek van Stanford University en ETH Zurich.

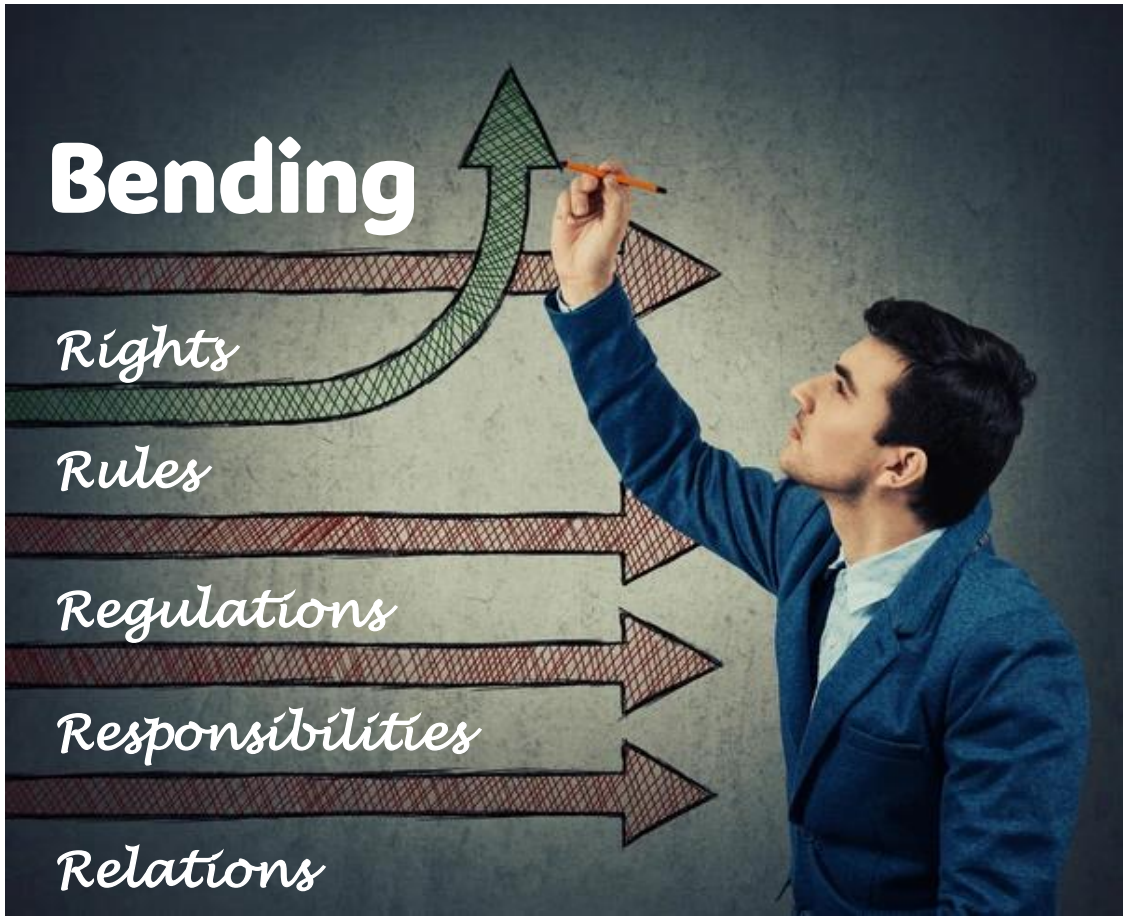
Recent developments

- Investigation FTM and UU researchers
- Freedom of Information data collection
- Critical close reading
  - Technical Documentation by IPL
  - Project Initiative Document by COA
  - Berenschot DP and AI Impact Assessment
  - Deloitte auditing report
  - ….



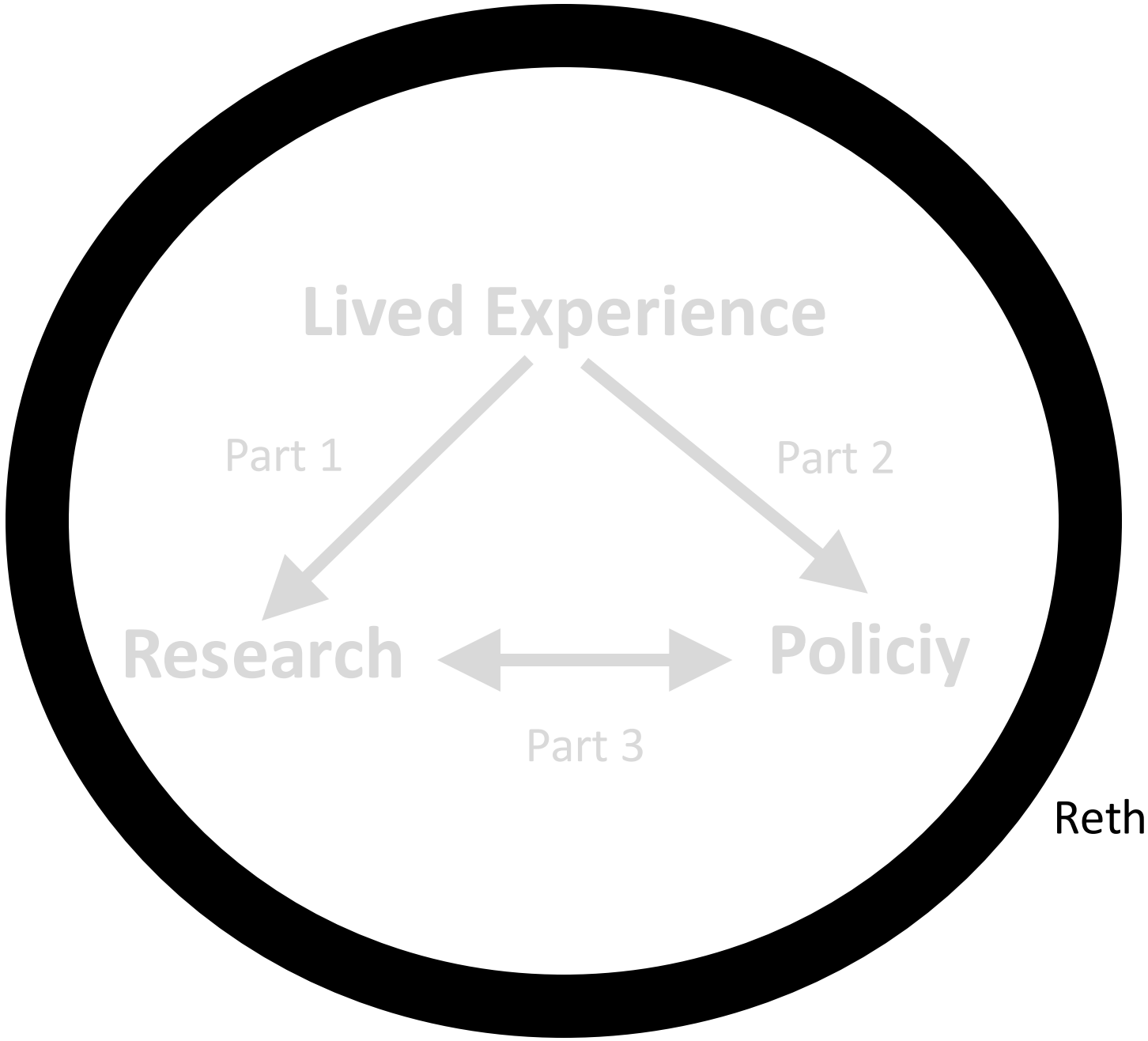
Current status

- Surgical analysis uncovers institutional logics:
  - Bureaucratic (COA)
  - Technical (IPL)
  - Managerial/auditing (Deloitte)
- More importantly:
  - Refugee data shared with IPL?
  - Optimization of economic values
  - Places - not people - centered
  - Ethnic and gender discrimination



AI systems - along with their critical assessment - are bended and embedded to serve institutions and the market rather than people and communitites and their wellbeing.

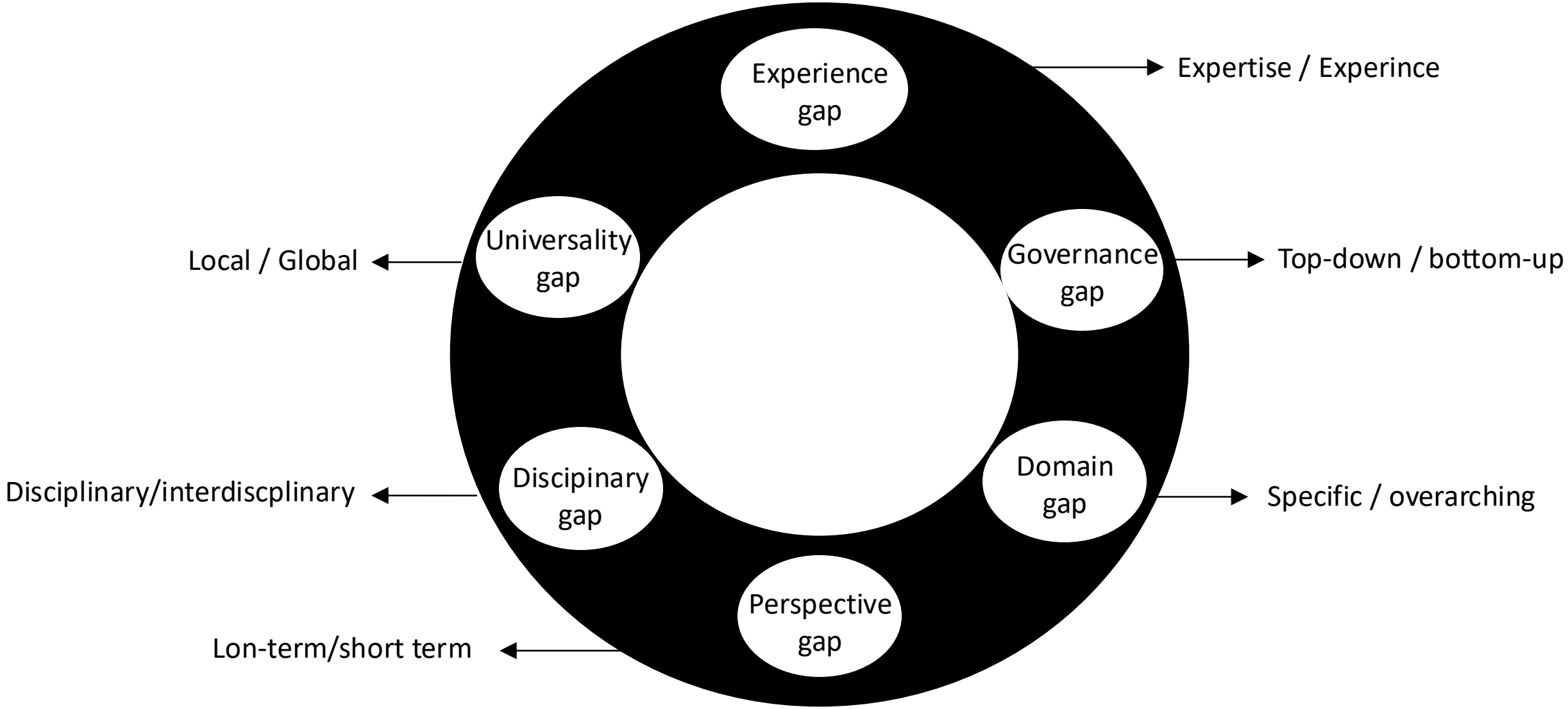
Part 1



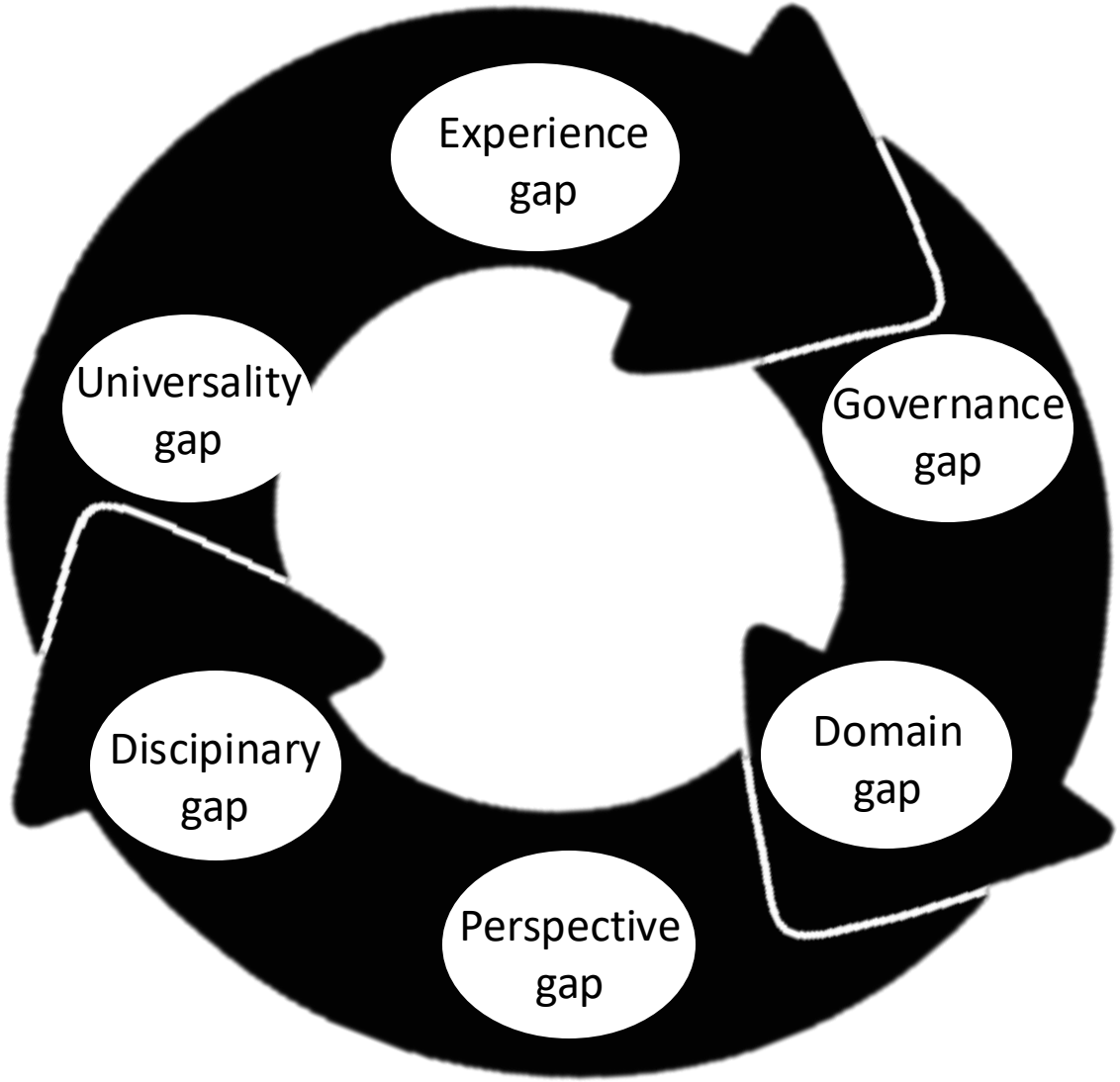
Part 4
Rethinkling Question
Zero in AI

Perpetual contraditions

Perpetual tensions

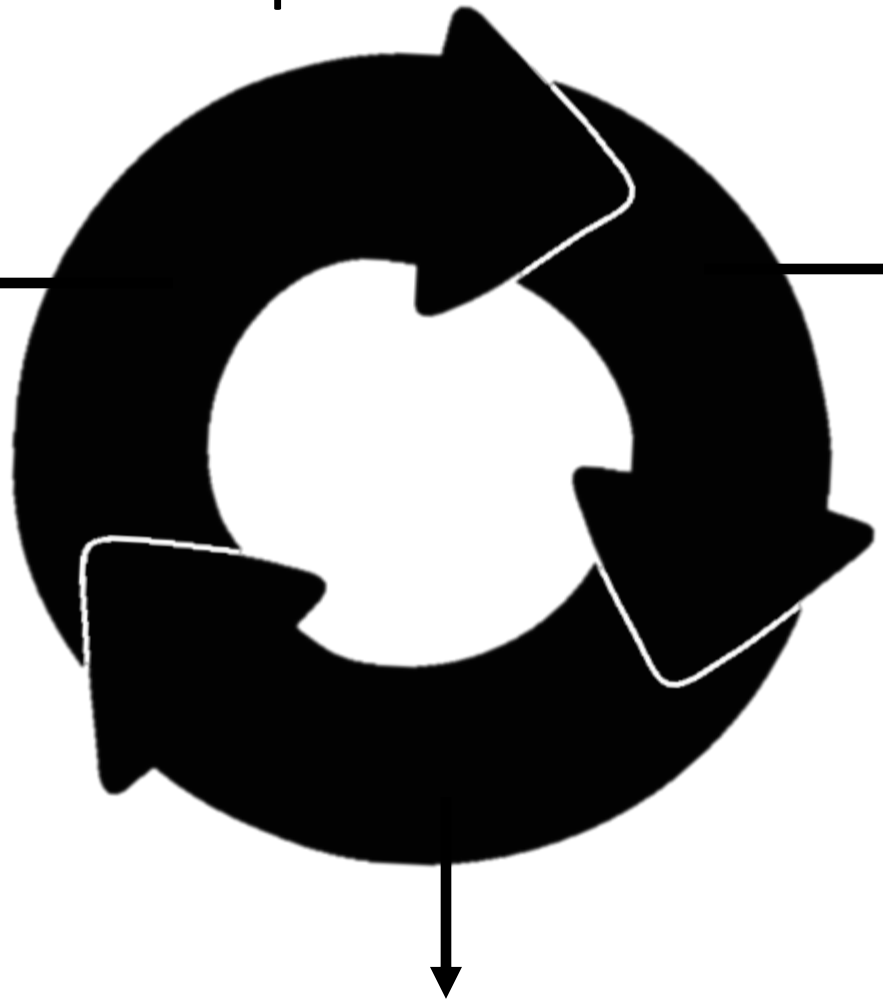


Perpetual relations



The fundamental problem: current society

Plurality in society
conflicting values
and behaviours

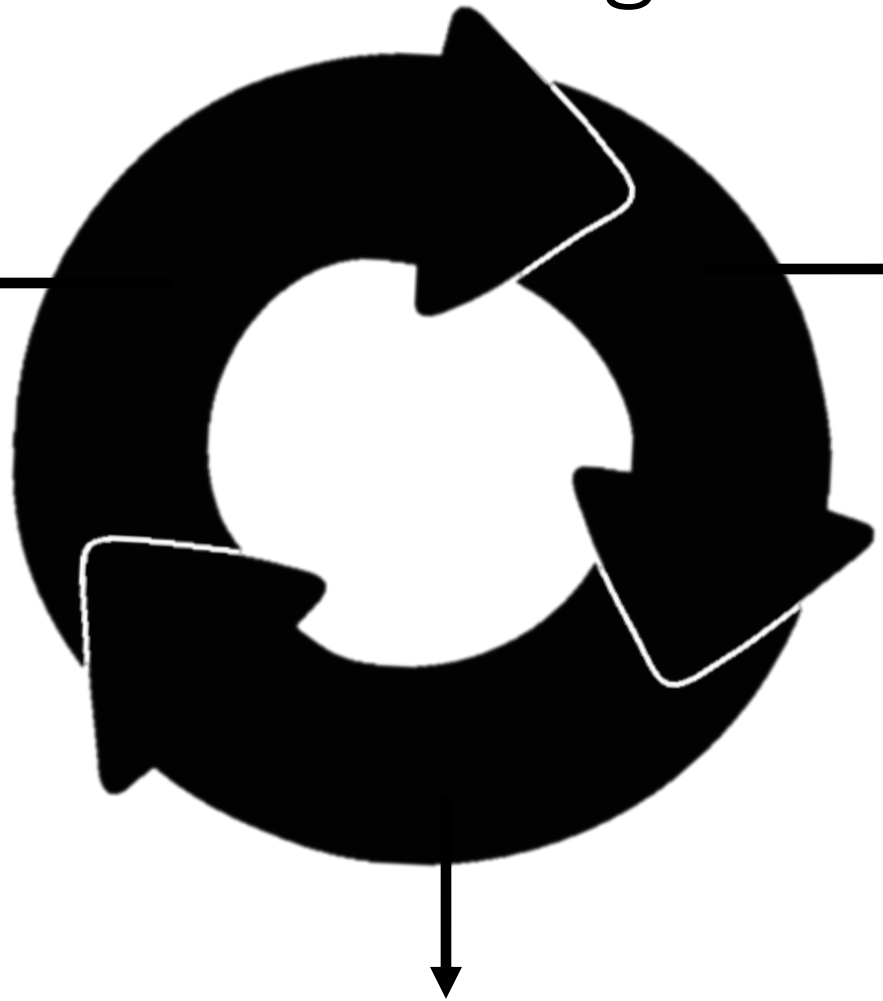


Inequality in society
unequal distribution of
wealth, knowledge,
resources

Hierarchy in society
Oppressive structures and power systems

The fundamental challenge: conscious society

Plurality in society
embracing
differences in values
and behaviours

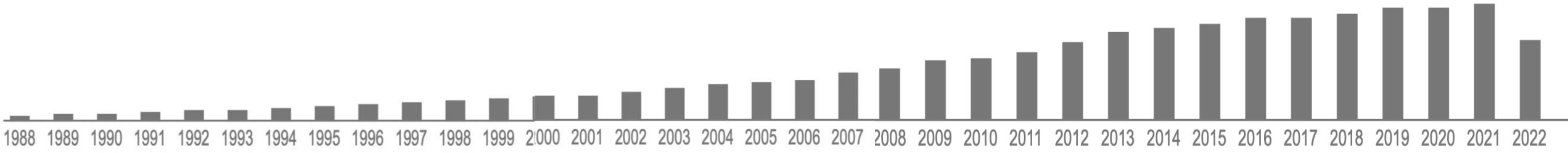


Inequality in society
Sharing wealth,
knowledge resources
in equitable ways

Hierachy in society
breaking with power systems and oppressive structures

Toward a conscious society: Paulo Freire

- Brasilian thinker, educator and changemaker
  - Critical consciousness: from oppression to liberation
- “Pedagogy of the oppressed”, 1970
  - 3rd most cited work in social sciences
- “Theater of the oppressed”
  - most used theatrical methodology in the world today



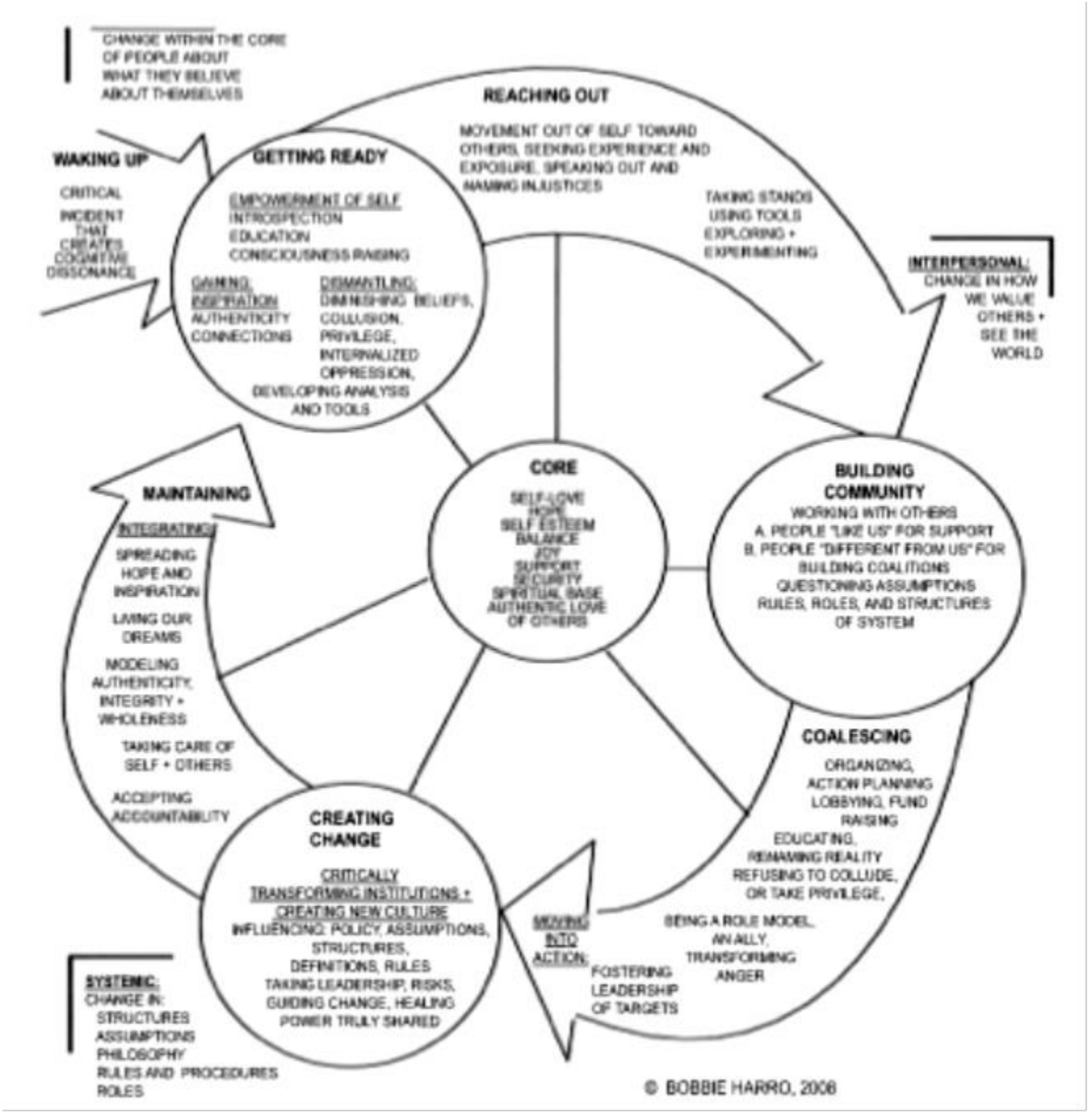
Current society: cycle of oppression

Pedagogy of the Oppressed, Freire, 1970
Cycle of Socialization and Liberation, Harro, 2000

AI's cycle of oppression

- AI has been undergoing the same cycle of socialization as humans
- AI has passed most stages of socialization, entering the second round
- AI is at a stage where it reinforces inequality, bias, power structures etc
- Without action the cycle of oppression will continue and accelarate
- But, there is a way out …

Conscious society: cycle of liberation



A Pedagogy for Liberation, Freire, 1987
Cycle of Socialization and Liberation, Harro, 2008

AI’s cycle of liberation

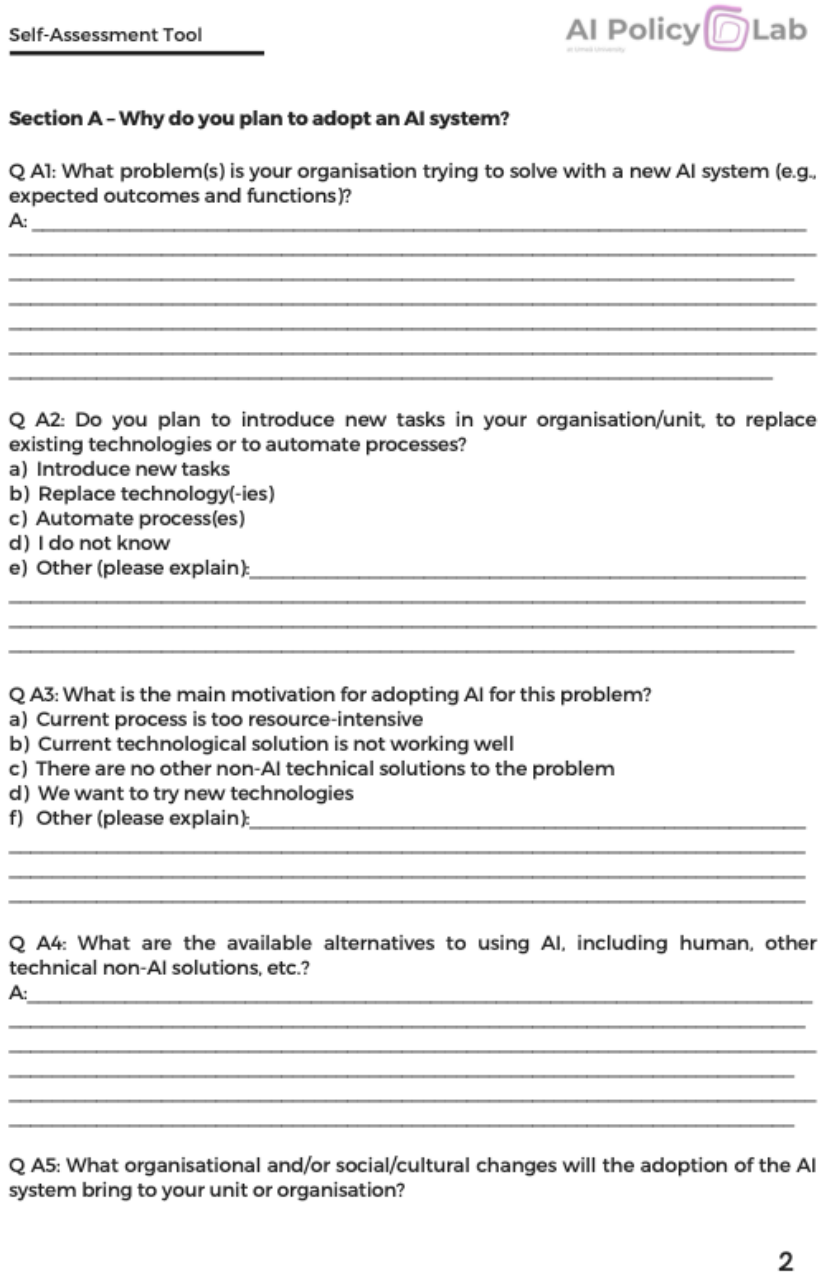
- AI provides an opportunity to get out of cycle of oppression
- AI has the power to empower people, communities
- AI as an instrument to maintain and reinforce cycle of liberation
- How to capitalize on AI’s potential for liberation?
- Socially intelligent design and development

Socially-Intelligent Artificial Systems

- Enable the transition from cycle of oppression to liberation
  - centering “Ethics of life”, not (only) “Ethics of Market”
  - focusing on values such solidarity, dignity, integrity, equity etc
  - allowing for continuous learning, exploration and co-creation
  - appreciating both universality and locality of social values and processes
- Focus on improving human-human and human-planet alignment
- Consider human-machine alignment a by product, not an end goal
  - This is where the AI safety research falls short

Question Zero

- From: ”Should we adopt AI in the first place?”
- To: “How should we use AI to empower people, to empower communities, to help them out of cycle of oppression into cycle of liberation
- This minimizes “Question Zero” risk of
  - Misuse/repurpose, intentionally or unintentionally
  - To Legitimize/reinforce existing unwanted practices
  - To Mask/introduce new unwanted practices



Self-Assessment Tool

AI Policy Lab

Section A – Why do you plan to adopt an AI system?

Q A1: What problem(s) is your organisation trying to solve with a new AI system (e.g., expected outcomes and functions)?
A: _____

Q A2: Do you plan to introduce new tasks in your organisation/unit, to replace existing technologies or to automate processes?
a) Introduce new tasks
b) Replace technology(-ies)
c) Automate process(es)
d) I do not know
e) Other (please explain): _____

Q A3: What is the main motivation for adopting AI for this problem?
a) Current process is too resource-intensive
b) Current technological solution is not working well
c) There are no other non-AI technical solutions to the problem
d) We want to try new technologies
f) Other (please explain): _____

Q A4: What are the available alternatives to using AI, including human, other technical non-AI solutions, etc.?
A: _____

Q A5: What organisational and/or social/cultural changes will the adoption of the AI system bring to your unit or organisation?

2

Question Zero

- What does this means in practice?
- How to translate into assessment?
- It might help to aim at
  - Building socio-technical systems, not AI
  - Centering lived experiences, not dominant theories/narratives
  - Exposing hidden socio-technical patterns, not hiding
  - Giving people/communities co-ownership, not meaningless credits
  - Enabling socio-technical transformations, not disrupting

Question ∞

“The oppressed, instead of striving for liberation, tend themselves to become oppressors”

Paolo Freire