



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 870661



A Mixed Blessing: Big Data and AI for Migration

Prof. Tuba Bircan

Interface Demography – Vrije Universiteit Brussel
Kavli Research Centre for Ethics, Science & the Public
Wellcome Connecting Science & Uni of Cambridge

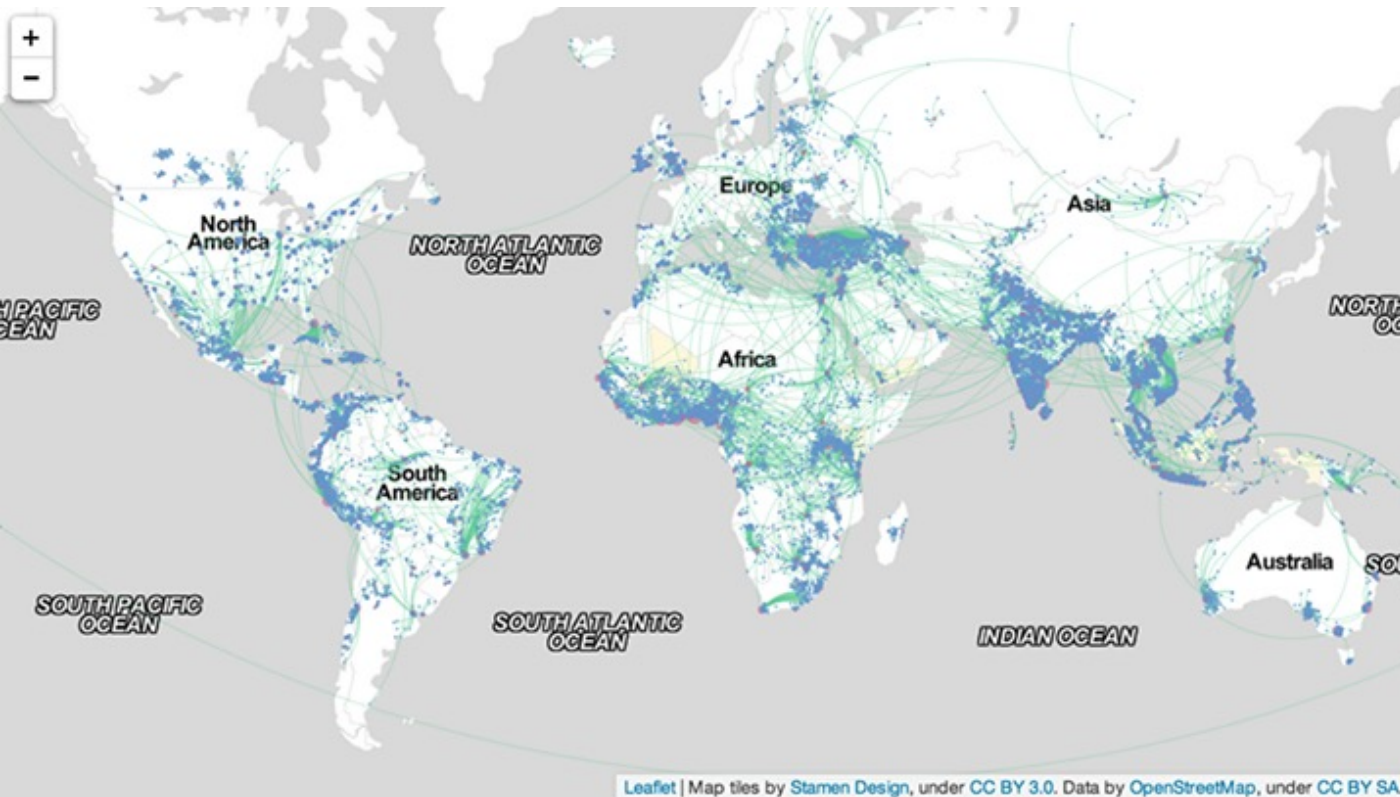
15/11/2022

VAIA - Flanders AI Academy Webinars

To discuss today:

1. Studying migration and mobility (M&M)
2. AI tools for migration monitoring, analysis and forecasting
 - Big Data sources to study M&M
 - AI at the borders
 - Ethical concerns
3. Data collaborations
4. Concluding remarks

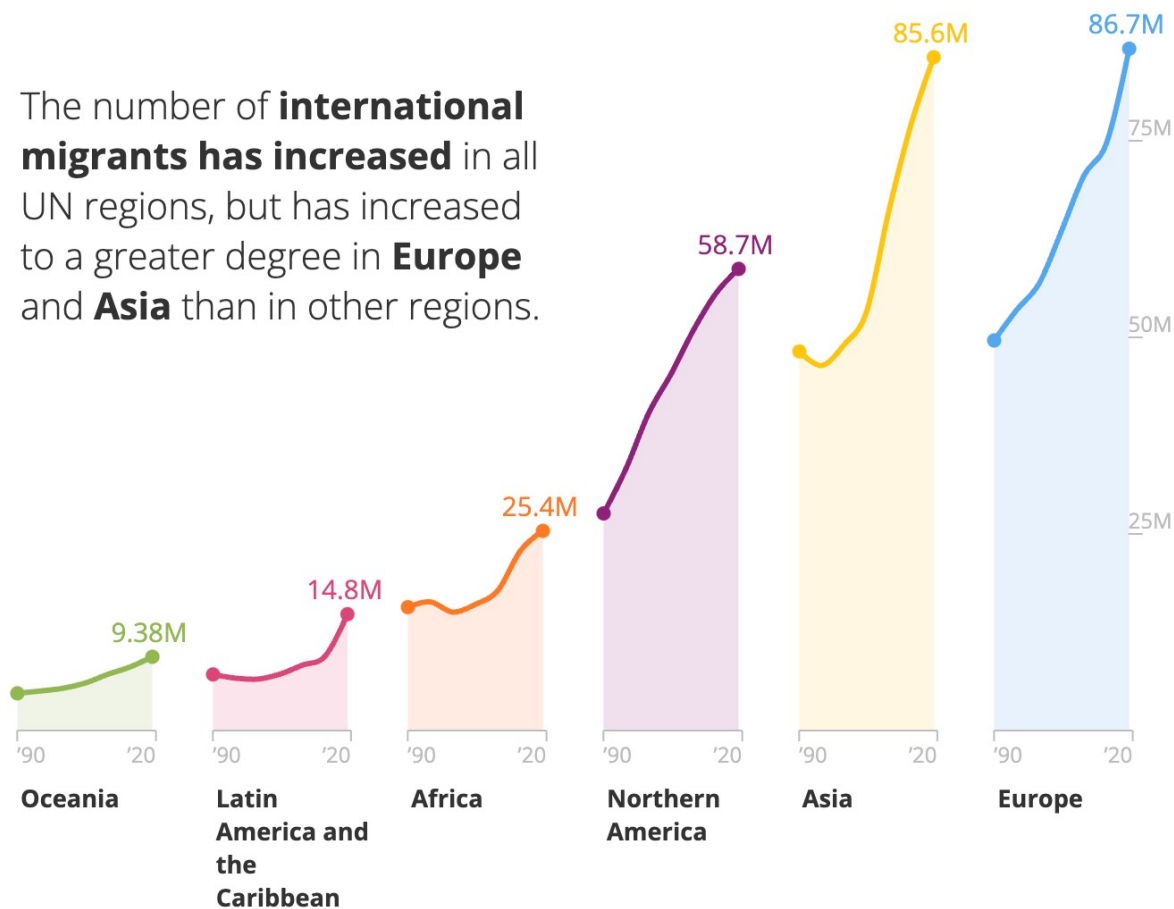
1. Studying Migration and Mobility



- Governments are investing vast sums in developing new technologies, and tech companies are at the forefront of commercializing various tech-based solutions.
- Many critical political, ethical, and social questions remain unanswered

1. Studying Migration and Mobility

The number of **international migrants** has **increased** in all UN regions, but has increased to a greater degree in **Europe** and **Asia** than in other regions.

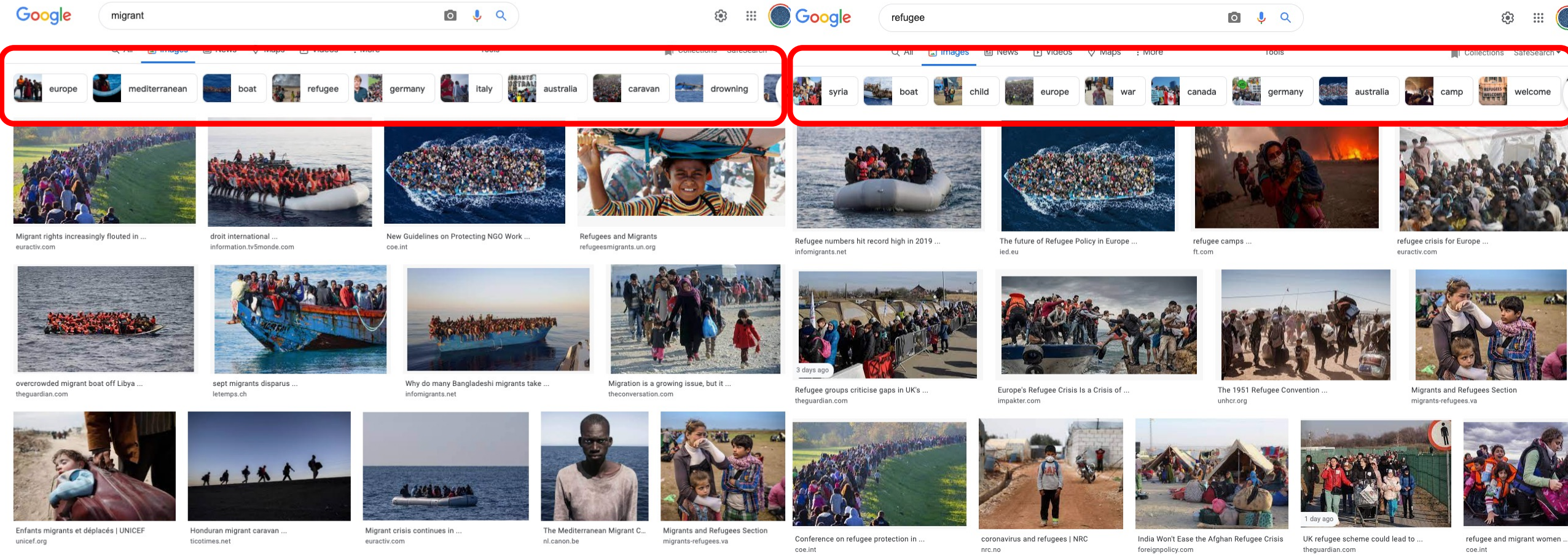


UN DESA 2021.

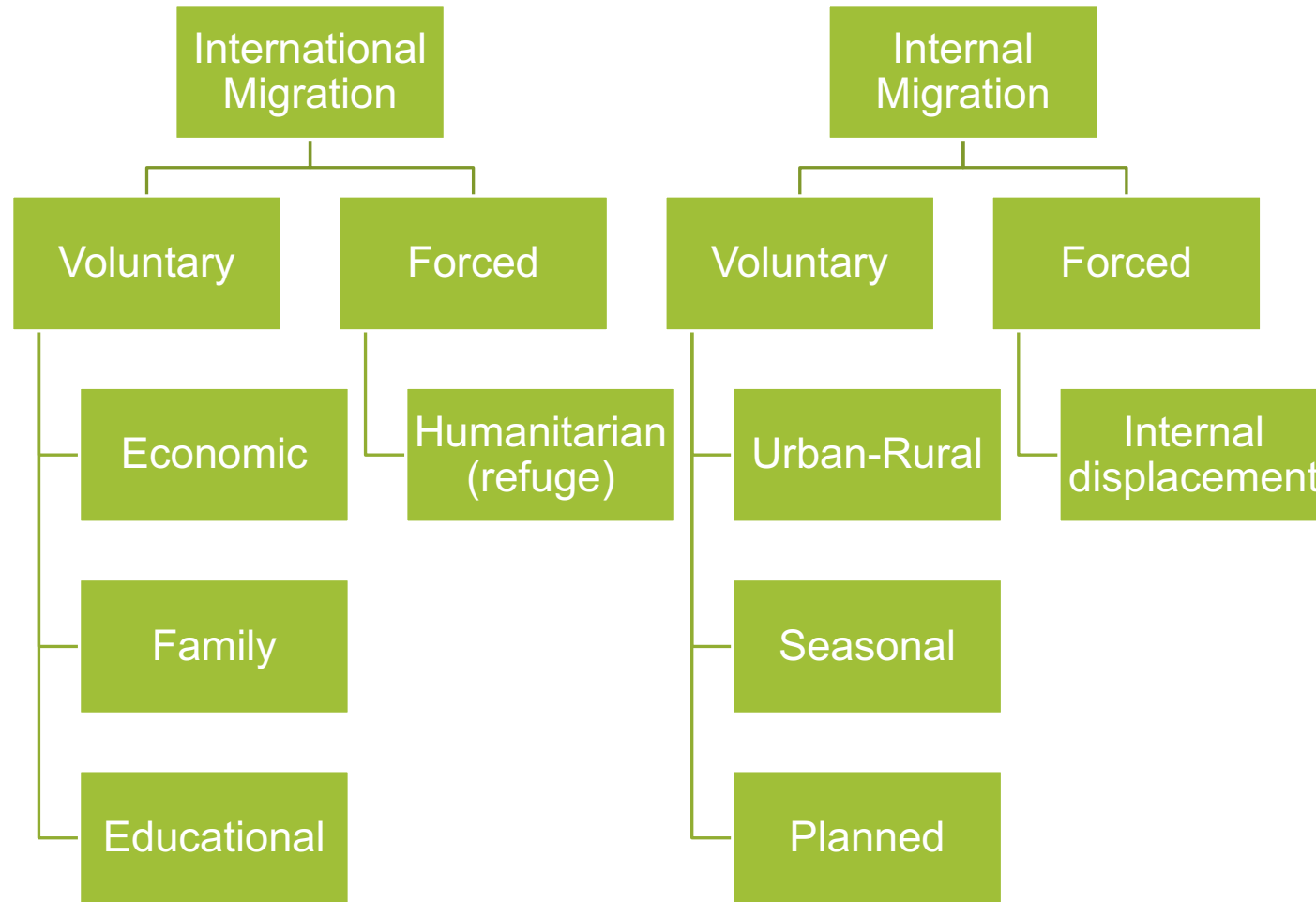
Mobility: Change of location for a limited time (and geography)

Migration: Change of residence for a longer period (usually at least 12 months)

Types of Migration & Migrants



Types of Migration & Migrants



Why Big Data & AI for M&M?

There is a need for:

- Social scientists with Big Data processing skills
- Computer/data scientists with interdisciplinary skills in social sciences
- Data collaboratives and data stewards

Two challenges:

- Technical challenges
- Ethical and legal challenges

2. AI tools for Migration Monitoring, Analysis and Forecasting

- Use of Big Data and AI for understanding **migration indicators** and addressing **data gaps**
 - Migrant stocks
 - Migration flows
- Other uses of Big Data and AI technology related to migration
 - Identity and origin checks
 - Border security and control
 - Analysis of data about visa and asylum applications
 - “AI for social good” applications for migrants and refugees
 - ...

Bircan, T., et al. (2020). Gaps in Migration Research: Review of migration theories and the quality and compatibility of migration data on the national and international level. Deliverable 2.3, Leuven: HumMingBird Project 870661 – H2020.

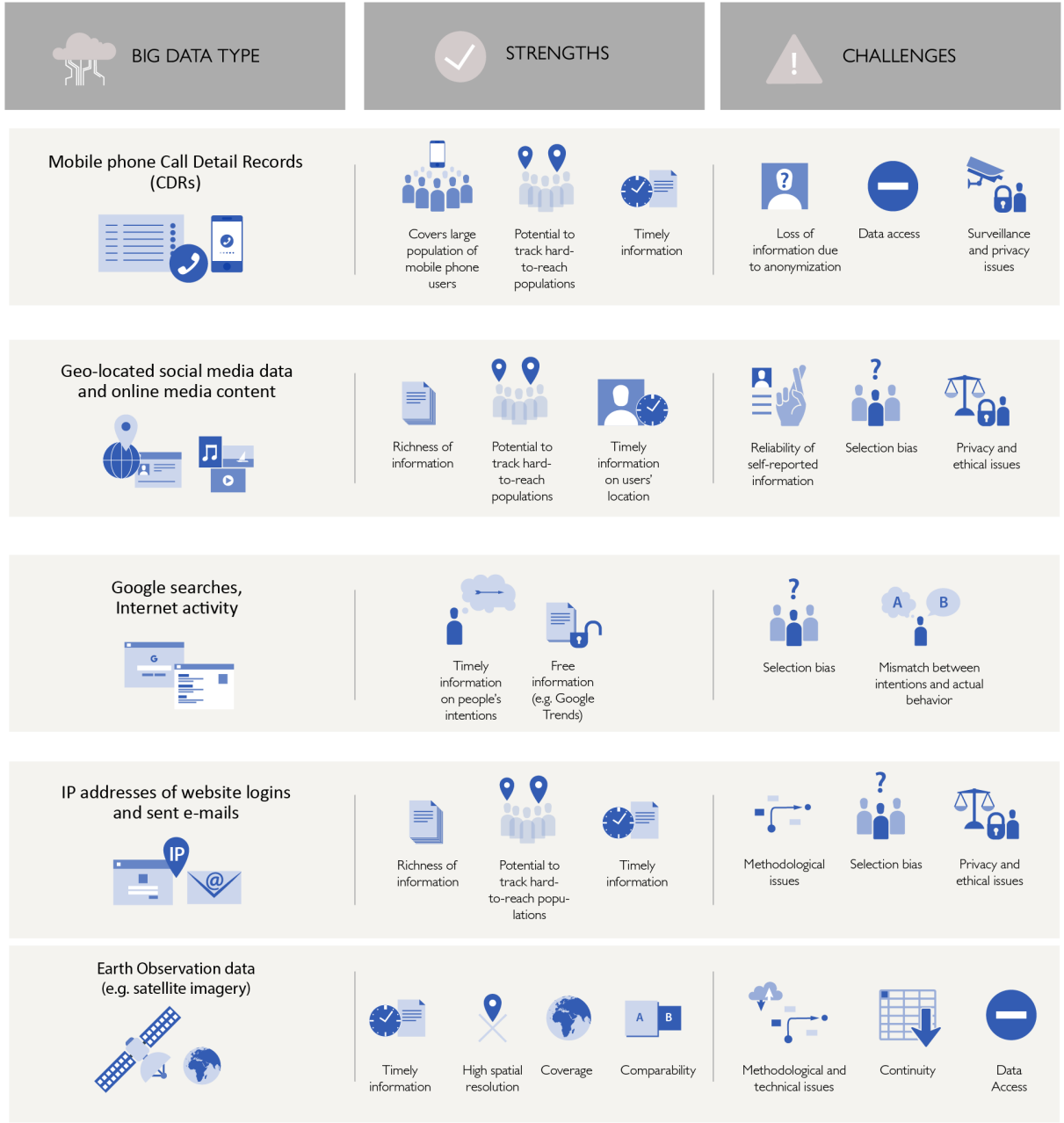
Salah, A.A., T. Bircan, E.E. Korkmaz (2022) New data sources and computational approaches on migration and human mobility, in DSMM. OUP.

➤ Big Data and AI for Migration Measures

The potential of big data to complement traditional data sources in migration statistics

- High spatial resolution
- High frequency of update
- Timeliness (virtually real-time)
- Wide coverage (hard-to-reach populations)
- Larger sample size compared to surveys
- Richness of information
- Relatively low costs
- They are collected *already*

Slide by M. Rango & N. Sievers, IOM GMDAC



Big Data Sources for Migration Measures

- Internet derived:
 - Social media, search, online interactions
 - Mobile devices, wearable technology, Internet of Things, etc.
- Tracking data
 - Monitoring the movements
 - GPS, geolocation data, traffic and transport sensor data, CCTV images, etc.
- Satellite and aerial imagery
 - Earth observation and remote sensing
 - Google earth, Landsat, Sentinel 1-2, infrared, radar mapping

Internet Data for M&M

Table 6. Bilateral model including Google Trends Indices and Gallup World Poll migration intentions.

Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Benchmark	GTI	Benchmark	GTI	Benchmark	GTI	Benchmark
	(A)		(B)		(C)		(D)
Log GDP (destination)	1.230*** (0.242)	0.675*** (0.239)	1.625*** (0.258)	0.810*** (0.246)			0.761*** (0.19)
Log Population (destination)	-3.384*** (0.433)	-0.764** (0.425)	-3.816*** (0.456)	-0.742* (0.417)			-1.82*** (0.32)
Log GDP (origin)	-0.915*** (0.125)	-0.503*** (0.121)					
Log Population (origin)	0.280 (0.328)	0.735** (0.323)					
Log GWP1 (intention)	0.0704*** (0.00455)	0.0655*** (0.00397)	0.0911*** (0.00652)	0.0786*** (0.00516)	0.119*** (0.00722)	0.0838*** (0.00566)	-0.001 (0.001)
Log GWP2 (plan)	0.0347*** (0.00902)	0.0375*** (0.00801)	0.0355*** (0.0104)	0.0386*** (0.00869)	0.0484*** (0.0104)	0.0409*** (0.00892)	-0.001 (0.001)
Log GWP3 (preparation)	0.0712*** (0.0126)	0.0484*** (0.0105)	0.0705*** (0.0137)	0.0426*** (0.0109)	0.0808*** (0.0135)	0.0441*** (0.0110)	-0.001 (0.001)
GTI (bilateral)	-	✓	-	✓	-	✓	-
Joint significance GTI (p-value)	-	0.000	-	0.000	-	0.000	-
Fixed effects							
Destination	✓	✓	✓	✓	-	-	-
Origin	✓	✓	-	-	-	-	-
Year	✓	✓	-	-	-	-	-
Destination-year	-	-	-	-	✓	✓	-
Origin-year	-	-	✓	✓	✓	✓	✓
Destination-year	-	-	-	-	-	-	✓

- Searching for a better life: Predicting international migration with online search keywords
 - Google Trends Index (GTI) for migration-related search terms can be exploited to measure migration intentions from a certain country and predict subsequent emigration
 - Online search words
 - a set of close to fifty words specific to the migrant population, such as passport, asylum, embassy, social assistance,
 - GTI indicators for each individual keyword in the official language of the respective country of origin
- The models with GTI included perform better, on average, than the basic one, explaining more of the variation in the migration outcome measure and, at the same time, producing fewer errors.



Social Media Data for M&M

Proxy for Migration Stocks with Facebook Data

Billateral migration flows 2018-2020

NUTS0 level

232 countries

19k+ records x 73 features (World)

40k+ records x 77 features (Europe)

Data sources:

 Humanitarian Data Exchange FACEBOOK Data for Good	Facebook connectivity (SCI), updated October 2021
	World migration flows, population stocks, NET
	European migration flows & population stocks
 United Nations Statistics Division	Integrate missing values
 THE WORLD BANK IBRD • IDA	Integrate missing values



Social Media Data for M&M

Migration flows to/from Spain

International citizens

	provenance	destination	MI_2019	sci
0	Cuba	Spain	386,983.86824	nan
1	Peru	Spain	15,510.57461	13,932.00000
2	Ecuador	Spain	14,882.53624	30,067.00000
3	Philippines	Spain	10,061.36801	1,568.00000
4	Colombia	Spain	2,045.82491	22,668.00000
5	Aruba	Spain	195.59179	20,032.00000
6	Uzbekistan	Spain	191.42113	508.00000
7	Åland Islands	Spain	122.17209	nan
8	Faroe Islands	Spain	100.99846	nan
9	Kenya	Spain	21.53141	858.00000

European citizens

	provenance	destination	MI_2019	sci
0	France	Spain	9.115723e+06	10218.0
1	Italy	Spain	1.607655e+07	17626.0
2	United Kingdom	Spain	1.194262e+07	13252.0
3	Romania	Spain	3.771340e+07	50921.0
4	Belgium	Spain	2.093445e+07	14593.0
5	Portugal	Spain	2.296671e+07	29713.0
6	Germany	Spain	3.078374e+06	9782.0
7	Poland	Spain	2.502049e+06	3559.0
8	Netherlands	Spain	1.563316e+07	11482.0
9	Bulgaria	Spain	2.175125e+07	21141.0

Most connected countries (Equatorial Guinea, Paraguay, Honduras, Dominican Republic, Nicaragua) do not share migration flows, but... they are Spanish colonies! Common history, language, culture, religion.

Connectivity of Spain

Top-30 countries having stronger connections with Spain



11

Social Media Data for M&M

Leveraging Twitter data to analyse profiles, tweets and network structure of migrants and natives

- Twitter, Italy, 2015
 - More than 200K users
 - Assigning nationality and country of residence

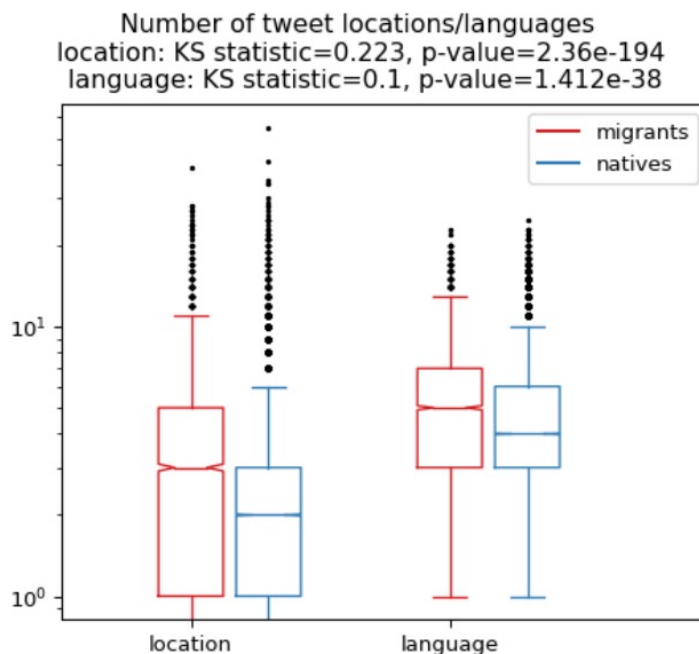


Figure 5: Distribution of tweet locations and languages

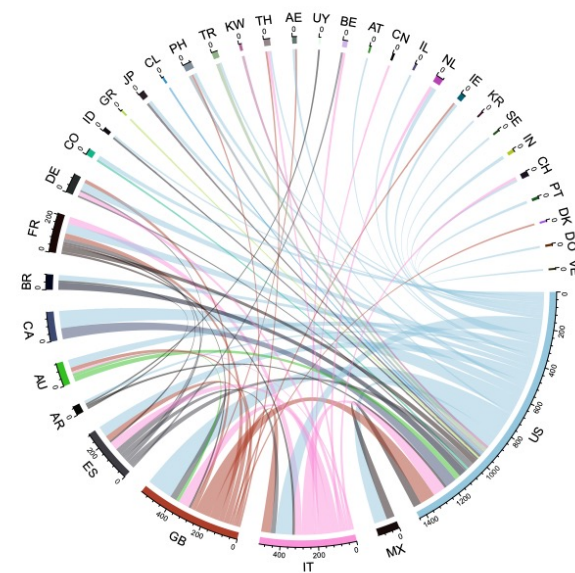


Figure 2: Chord diagram on migration patterns: The number of migrants have moved from a country to another is represented by the links. The colors represent the nationalities of migrants. We show only countries with at least 10 migrants for the visualisation purpose.

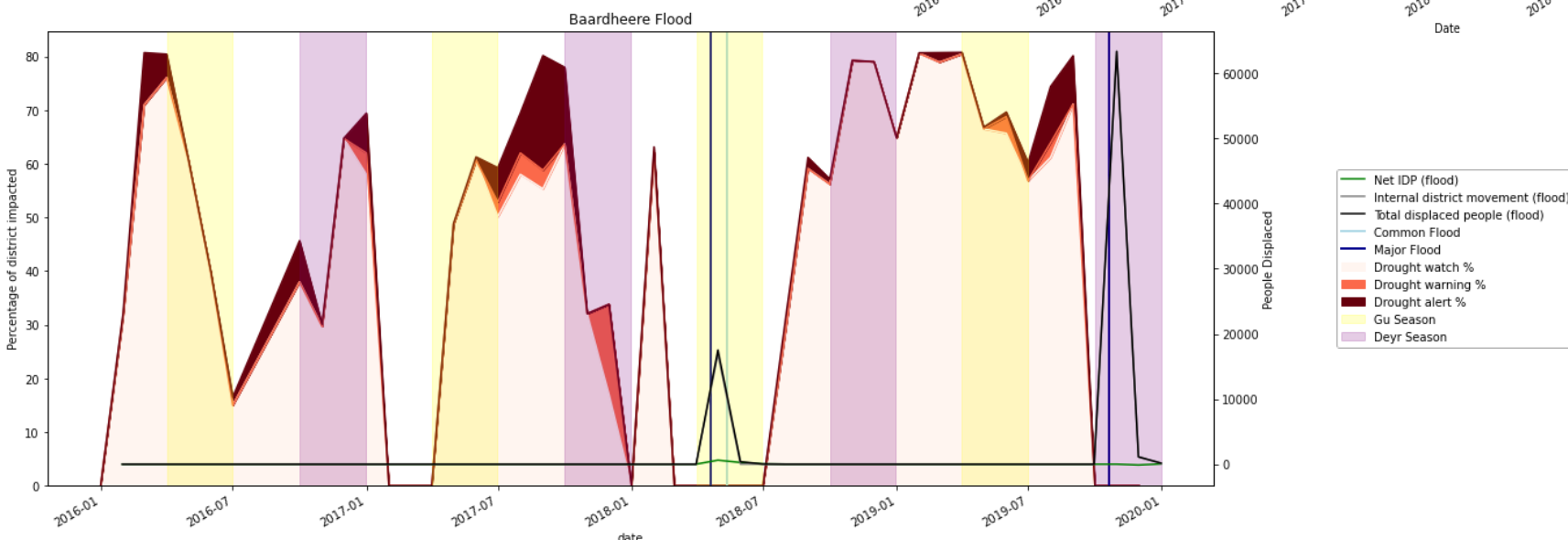
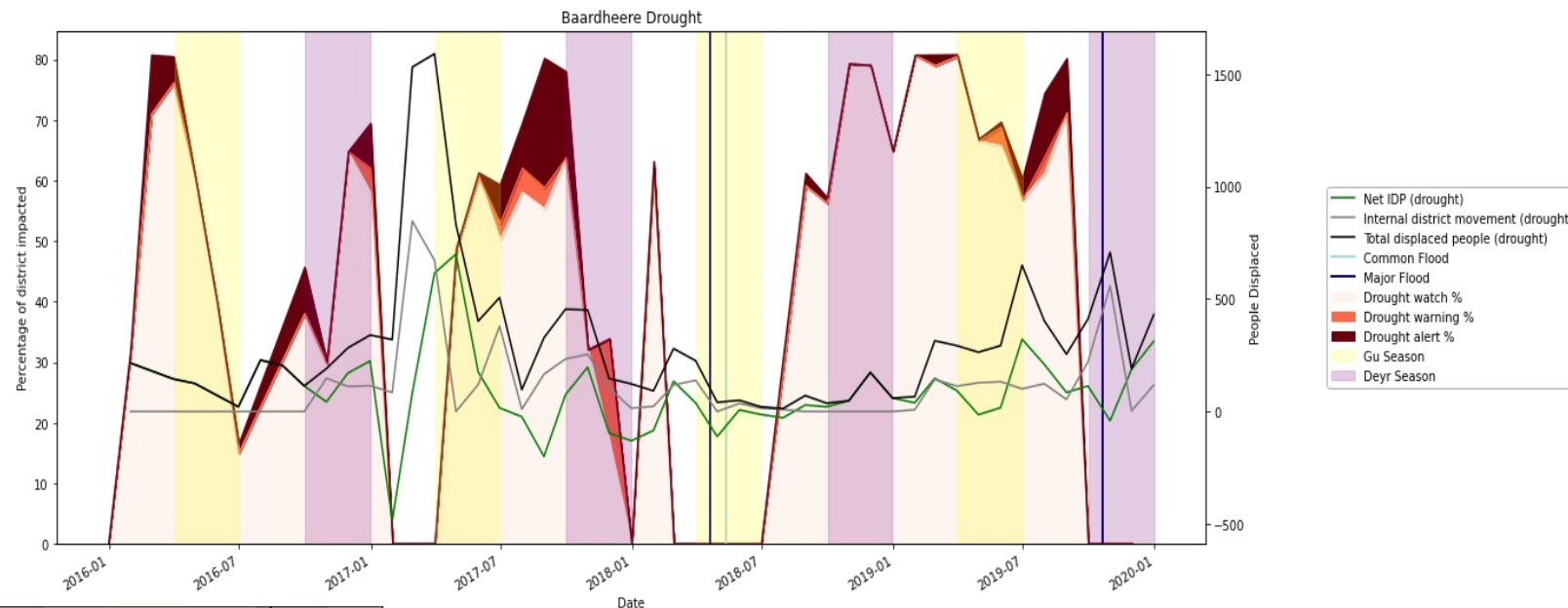
Satellite Data for M&M

Environmentally induced internal displacement

- Satellite data analyses (Somalian case)
 - For the period 2016-2020
 - Sentinel -1
 - A land use and land cover (LULC) map were generated at the beginning of each period
 - the changes in land use/ cover due to drought, floods, or violence break outs, will be analysed.
 - Sentinel - 2
 - bi-monthly time series are converted into different time series of vegetation indices or indicators

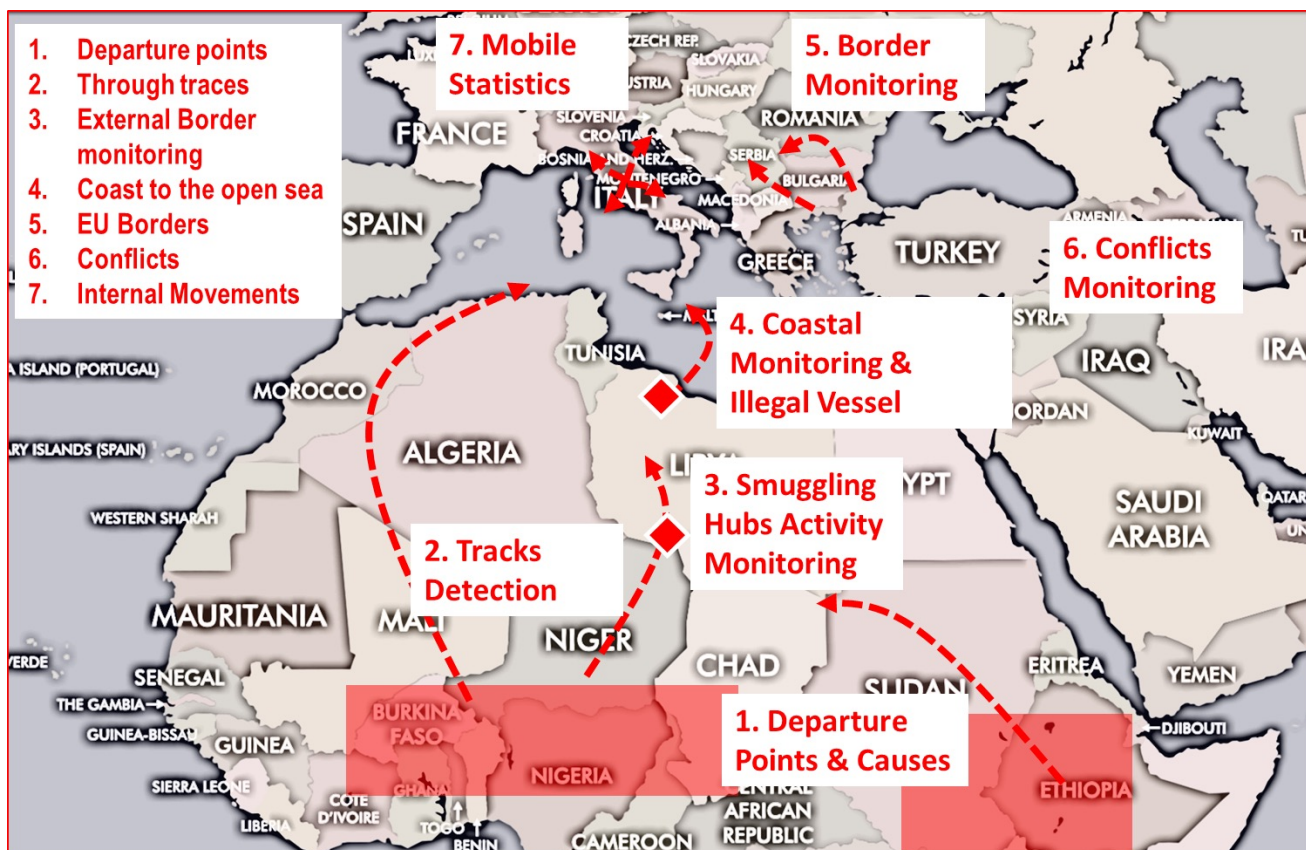


Satellite Data for M&M



Momeni, R., Bircan, T., King, R. (2022). Environment-induced internal displacement. The case of Somalia (Deliverable 4.3). Leuven: HumMingBird project 870661 – H2020.

Satellite Data for M&M



Migration Radar 2.0 by ESA

- Big Data applications to boost mitigation preparedness and response to migration feasibility study

Mobile Phone Data for M&M

Inferring patterns of internal migration from mobile phone call records: evidence from Rwanda

- Rwanda's primary telecom operator an exhaustive log of all phone-based activity that occurred from the beginning of 2005 through the end of 2008 (1.5 million subscribers)
 - time-stamped record of every call that the individual made or received
 - the closest tower to the subscriber at the time of the transaction (Such as a phone call or text-message)

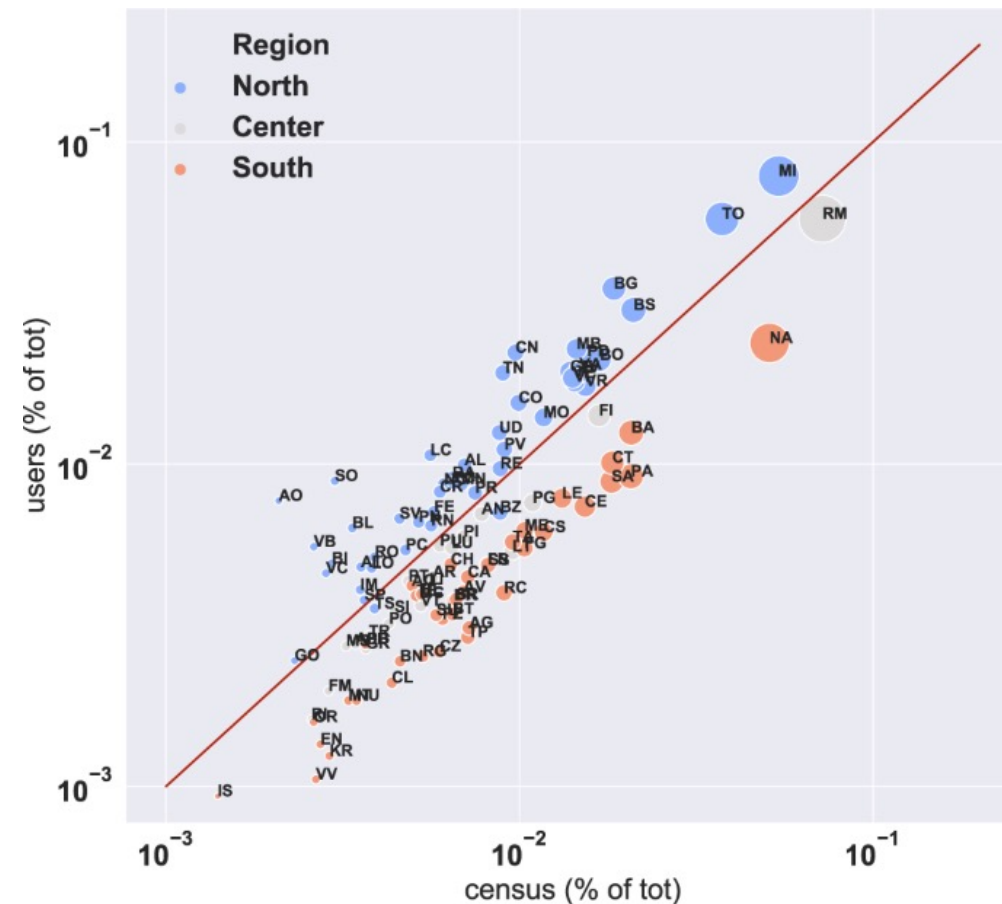
	All	Farmer	Teacher	Student	Unemployed	Transport
ROG (km)	15.32	13.57*	15.81	18.94*	17.15	26.01**
# towers	14.02	8.99***	11.39**	15.97	15.06	40.04***
Max distance (km)	71.62	60.79***	77.03	74.76	76.69	111.84***
3-month migration	0.23	0.14***	0.28	0.22	0.28	0.39
12-month migration	0.016	0.01	0.01	0.00***	0.02	0.09
N	901	269	109	77	47	23
*Indicates mean of occupation is different from mean of group with $p < 0.05$.						
**Indicates mean of occupation is different from mean of group $p < 0.01$.						
***Indicates mean of occupation is different from mean of group $p < 0.001$.						



Mobile Phone Data for M&M

COVID-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown

- Anonymously shared positions of about 170,000 de-identified smartphone users before and during the outbreak, at the sub-national scale
- Daily time-series of three different aggregated mobility metrics
- Data is limited to users who have **opted-in** for anonymously sharing their location with Cuebiq.



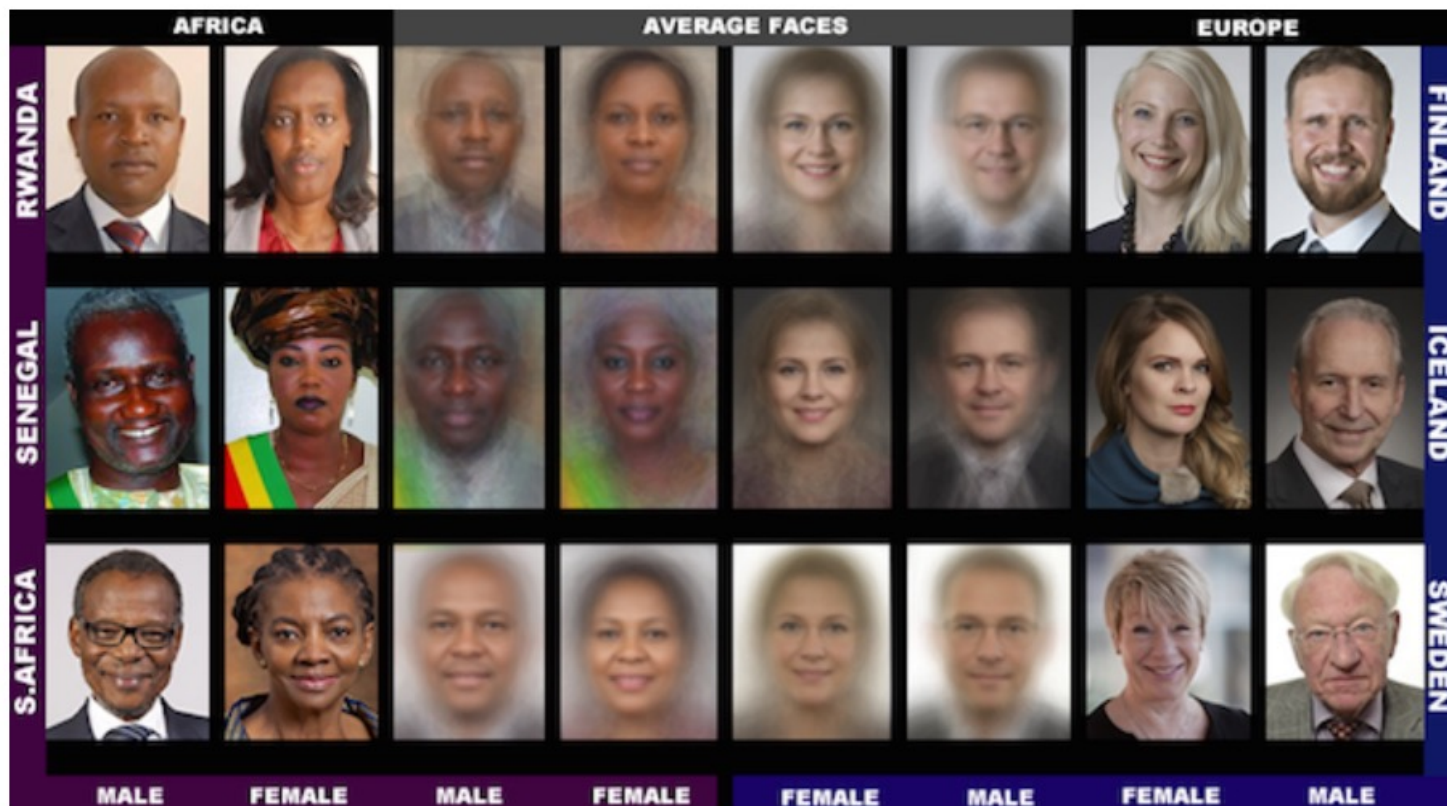
➤ AI at the Borders

- Identification technologies
- EU smart borders
- EU integrated border management
- Automated biometric systems
- Algorithmic profiling

AI at the Borders: Identification Technologies

- Identification of mobile and risky people
- Fingerprints
- Facial recognition
- Automated identification and AI

AI at the Borders: Identification Technologies



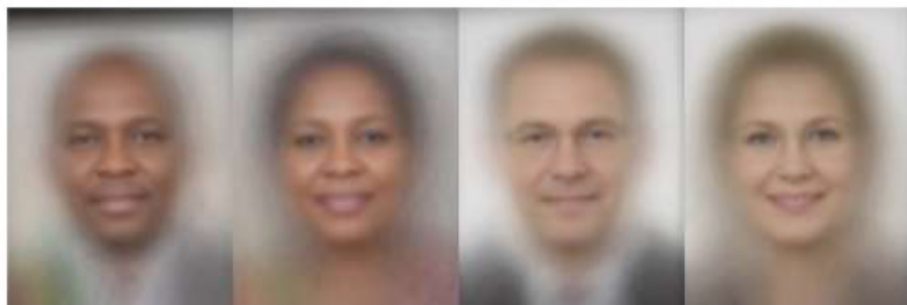
Gender Shades project evaluates the accuracy of AI powered gender classification products

Pilot Parliaments Benchmark



AI at the Borders: Identification Technologies

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%



Leading tech companies' commercial AI systems significantly mis-gender women and darker skinned individuals.

High-Level Benchmark Evaluation Results

Government Face Recognition Benchmark: 79.6% Lighter Skin

Research Gender Classification Benchmark: 86.2% Lighter Skin

High-Level Gender Classification Results

All classifiers perform better on male faces than female faces (8.1%-20.6% difference in error rate)

All classifiers perform better on lighter faces than darker faces (11.8%-19.2% difference in error rate)

All classifiers perform worst on darker female faces (20.8%-34.7% error rate)

Microsoft and IBM classifiers perform best on lighter male faces (error rates of 0.0% and 0.3% respectively)

Face++ classifiers perform best on darker male faces (0.7% error rate)

The maximum difference in error rate between the best and worst classified groups is 34.4%

AI at the Borders: Smart Borders

Migration management vs Migration governance?

Face recognition technologies have come under increased scrutiny due to concerns about:

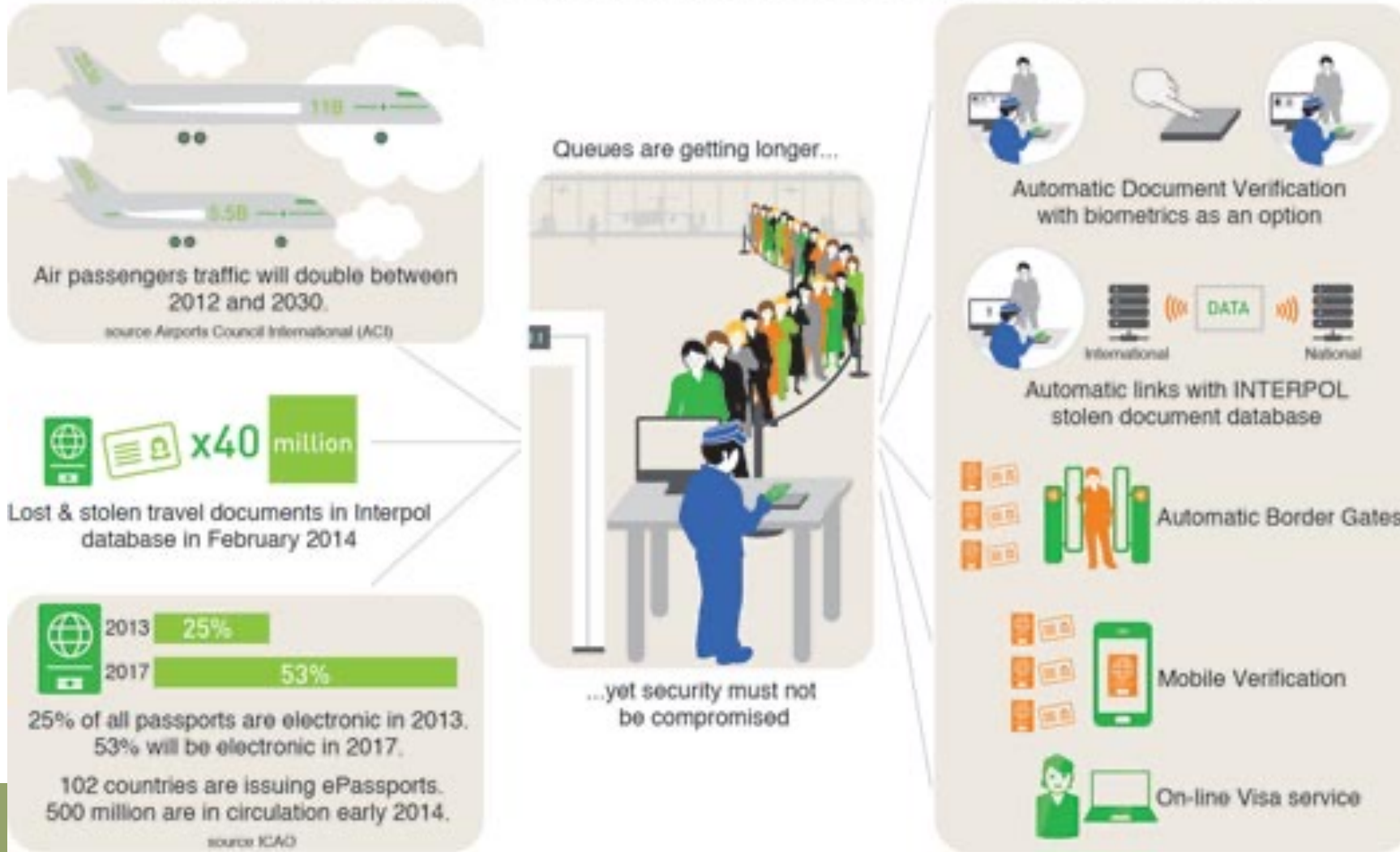
- fundamental rights
- risks related to bias and discrimination, data protection and mass surveillance.



AI at the Borders: Integrated Border Management

Innovation at the borders

Best-practice immigration processes and state-of-the-art technologies can cut the bottlenecks.



AI at the Borders: Automated Biometric Systems

Refugees are buying groceries with iris scans. What could go wrong?

More than 80% of the refugees in Jordanian camps now use iris scans to pay for their groceries. Refugee advocates say this is a huge future privacy problem.



WFP Introduces Iris Scan Technology To Provide Food Assistance To Syrian Refugees In Zaatari



TOPICS

Jordan

Syrian Arab Republic

AI at the Borders: Algorithmic Profiling

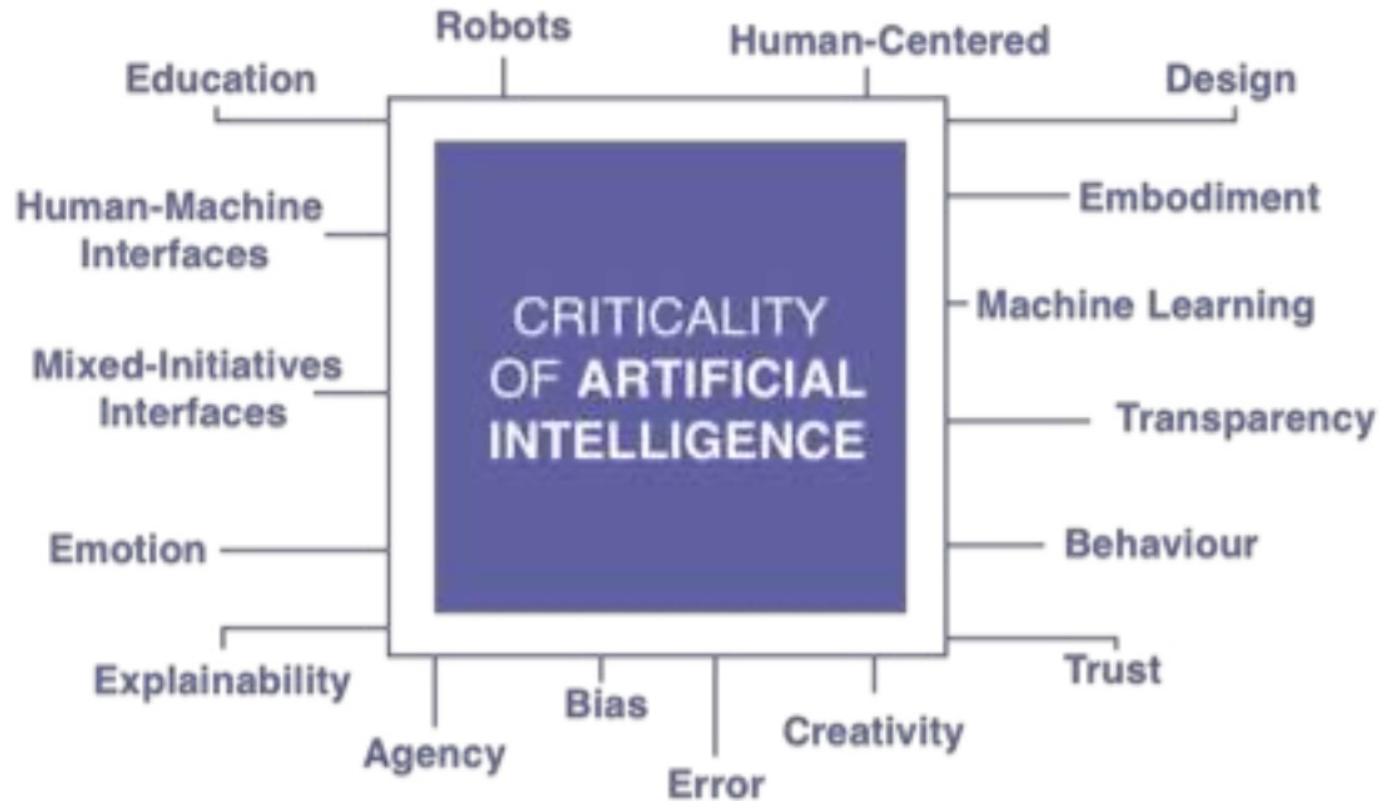


- In 2016, ProPublica analysed a commercially developed system (COMPAS) that predicts the likelihood that criminals will re-offend, created to help judges make better sentencing decisions, and found that it was biased against blacks.

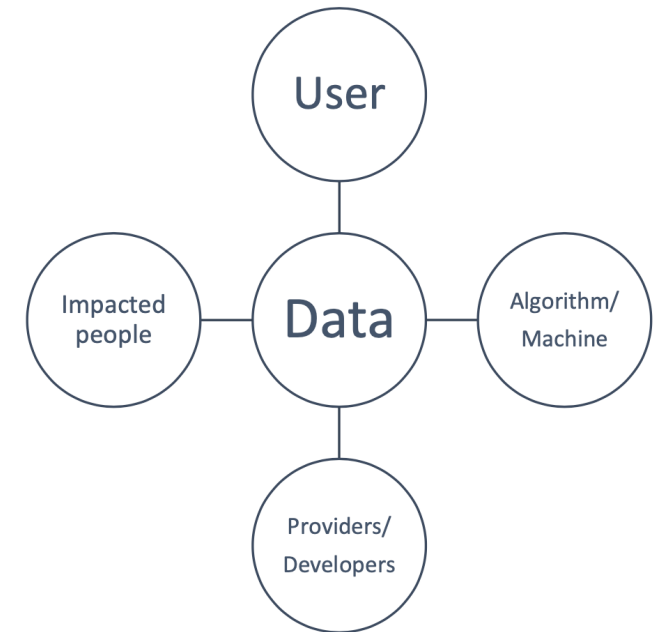
AI at the Borders: Algorithmic Profiling

- Automated decision making processes:
 - Can be used to ignore/prevent fundamental rights of people to seek asylum
 - Can exacerbate pre-existing vulnerabilities by adding on risks such as bias, error, system failure and theft of data

➤ Ethical Concerns



Biased AI ⇔ Biased Society



Consequences of Biased AI for M&M?

Bias in AI => Real-life implications on real people:

- Unequal treatment
- Violation of human rights
- Preventing access to fundamental rights

3. Data collaboratives

‘Co-operation between the private and public sectors in the field of data sharing and statistics’



3. Data collaboratives: Mobile Phone Data

Dataset	Data	Year	#pp	Observations
D4D Ivory Coast	CDRs	2011-2012 (1 year)	5 million	Antenna-to-antenna traffic (hourly), 50k individual trajectories at antenna level; 500k individual trajectories at prefecture level; 5k comm graphs
D4D Senegal	CDRs	2013 (1 year)	9+million	Antenna-to-antenna traffic, 300k users mobility
Telecom Italia	CDRs, electricity, weather, rain, news, geolocated tweets	2014	300+k	Milan and Trento
Telefonica	Smart steps, hospital admission, transportation, Twitter, crime	2013 (3 weeks)	500+k	London metropolitan area
D4R Turk Telekom	CDRs, labeled as refugee & non-refugee	2018	1 million	Syrian refugee movements in Turkey
HummingBird Turkcell	CDRs	Est. 2021	-	To study migration from Turkey into Europe



Data for Refugees (D4R) Challenge

Collected from 200K refugees and 800K non-refugee residents, over the entire country, over one year.

- 1- Cell tower locations, lists of cell towers for each prefecture
- 2- Site-to-site antenna traffic on an hourly basis.
- Total number and duration of calls given, separated into “originating from refugees” and “not originating from refugees”.
- 3- Fine grained mobility of a small subset of (anonymous) users. Only 15 days for each user, at cell tower level.
- 4- Coarse grained mobility of a small subset of (anonymous) users. For the entire data collection period, but provided at prefecture level.



Data for Refugees (D4R) Challenge: Priority Themes

Safety & Security

Violence, theft, illegal trafficking

Health

Access to resources, spread of diseases, vaccination

Education

Access to education, language learning, schools

Unemployment

Movement due to unemployment, skill and resource management

Social Integration

Events and institutions for social integration, segregated/mixing patterns of behavior

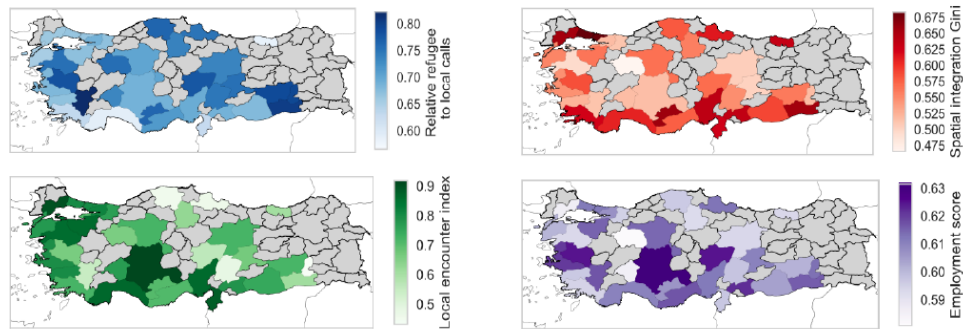


Aim of D4R

- Contribute to the **welfare of the refugee** populations,
- Gain **insights** on key issues, including safety and security, health, education, unemployment, social integration and segregation, mobility, and distribution of resources and infrastructure,
- Help governments and international bodies model the dynamics of the refugee populations and to discover vulnerabilities (socioeconomic vulnerabilities, gaps in education and services, etc.), **policy** insights,
- Seed **further projects**, co-created with refugees, resulting in new applications, services, and innovative solutions for refugees in Turkey and elsewhere.

Example applications from D4R

Social integration indicators



Access to healthcare

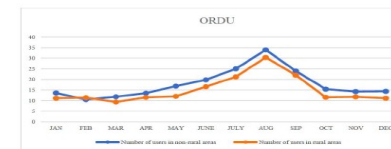
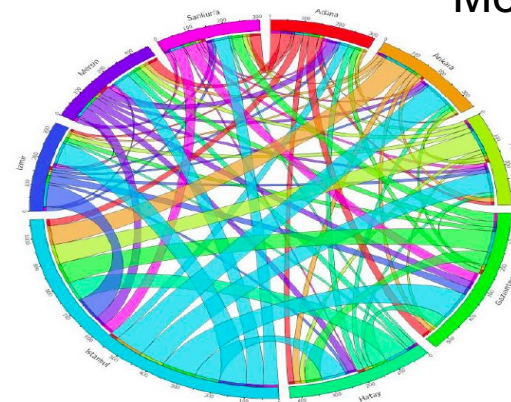


Fig. 2. Map with red pins showing the locations of current MHCs over the choropleth which illustrates the density of refugee residents at different districts in Istanbul.

Altuncu T, Sevensan N, Kaptaner AS (2019) Optimizing the access to healthcare services in dense refugee hosting urban areas: a case for Istanbul. In: Data for Refugees Challenge Workshop

Bakker M, Piracha D, Lu P, Bejgo K, Bahrami M, Leng Y, Balsa-Barreiro J, Ricard J, Morales A, Singh V, Bozkaya B, Balcişoy S, Pentland A (2019) Measuring fine-grained multidimensional integration using mobile phone metadata: the case of Syrian refugees in Turkey. In: Data for Refugees Challenge Workshop

Mobility of seasonal workers



Alışık ST, Aksel DB, Yantaç AE, Baruh L, Salman S, Kayı I, İçduygu A, Bensason I (2019). UDMIT: an urban deep map for integration in Turkey. In Data for Refugees Challenge Workshop

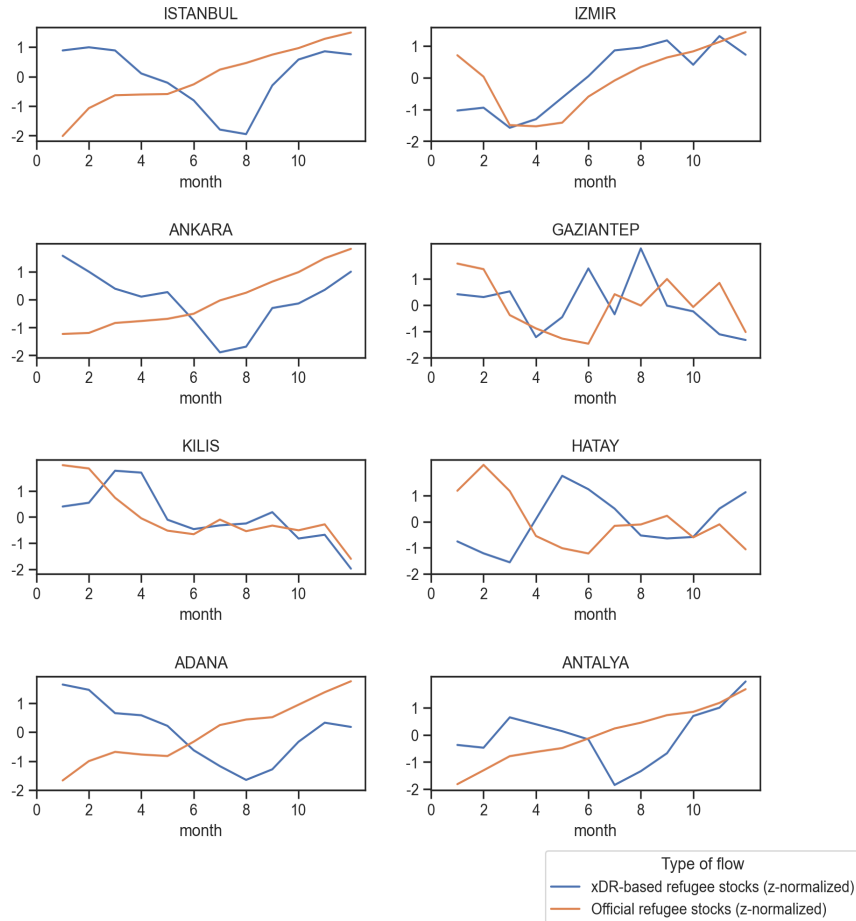
HumMingBird Turkcell Collaborative

- Focusing on refugees and undocumented migrants
 - CDR data analyses (Istanbul case)
 - A new dataset in collaboration with Turkey's primary telecommunications operator, Turkcell
 - containing an exhaustive log of all phone-based activity that occurred from the beginning of 2019 through the end of 2019 for a carefully sampled subsample of the population, including refugee and migrant groups
 - Ethics and OPAL
 - The CDR's will be anonymized and aggregated to remove all personal 39 information, along the guidelines established by ethics committees in past endeavours
 - aggregation and anonymization scheme closely follows the Data for Development Challenge setup and the Data for Refugees Challenge setup, where it was shown to provide mobility indicators without jeopardizing privacy or security.

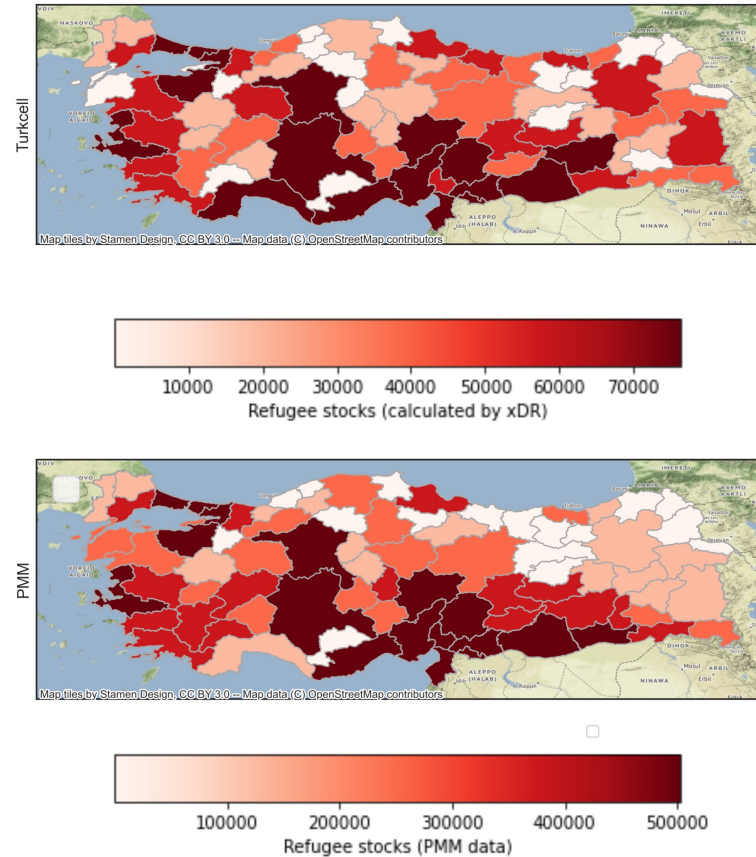


Examples from HumMingBird Turkcell Collaborative

Migrant flow indicators (internal migration)



Migrant stock indicators (national)



4. Concluding Remarks

- AI technology => innovative, cost- and time-effective for international migration governance
- AI literacy is of extreme importance for policymakers and decisionmakers
- Citizen engagement & trust are of great importance
- Essential questions arise
 - legal requirements, confidentiality, and rules of engagement
 - ethical concerns of (mis)use of new technologies for more conservative and preventive migration policies and hence violated human rights.

4. Concluding Remarks

- State-industry relations should be clear to identify
 - accountability, responsibility and GDPR compliance
- Research can be improved through
 - More investment
 - More multidisciplinary multi-stakeholder collaborations
 - AI researchers/developers/engineers should be able to see “human” behind the algorithms
 - Transparency, explainability, interpretability are crucial
 - Data collaboratives

Migration is a complex phenomenon with multi-dimensional factors. With ever-expanding data storage and processing capabilities, new data sources (such as social media, mobile call data records, and satellite imaging) are becoming available to study migration from both qualitative and quantitative perspectives. *Data Science for Migration and Mobility* addresses the needs of both migration scholars who stand to benefit from the analysis of these new sources but lack the computational tools, as well as data scientists who have practical and theoretical knowledge in dealing with these data sources but have no familiarity with the relevant questions of migration research. It describes the main conceptual frameworks, explains techniques of data collection and processing, provides case studies, discusses the strengths and limitations of each data source, and critically discusses the ethical, legal, and privacy-related issues specific to each data source.

"This interdisciplinary book presents the state of the art of data science for migration and mobility, offering many scientific and practical insights for data-savvy social scientists. The authors understand that with big data comes big responsibility: they critically discuss the risks of data science throughout the book."

Tijs van den Broek, Assistant Professor of Organization Sciences, VU University of Amsterdam

"This timely volume establishes new common ground for the emerging interdisciplinary research area of digital migration studies. It is highly recommended to data scientists and migration researchers who want to learn about both the pros and cons as well as the nuts and bolts of accessing, processing and combining big data sources to study human movement."

Koen Leurs, Associate Professor of Gender, Media and Migration Studies, Utrecht University

Cover: Viktor Sira Salah



Published for The British Academy
by Oxford University Press
Part of the Proceedings of the British Academy series



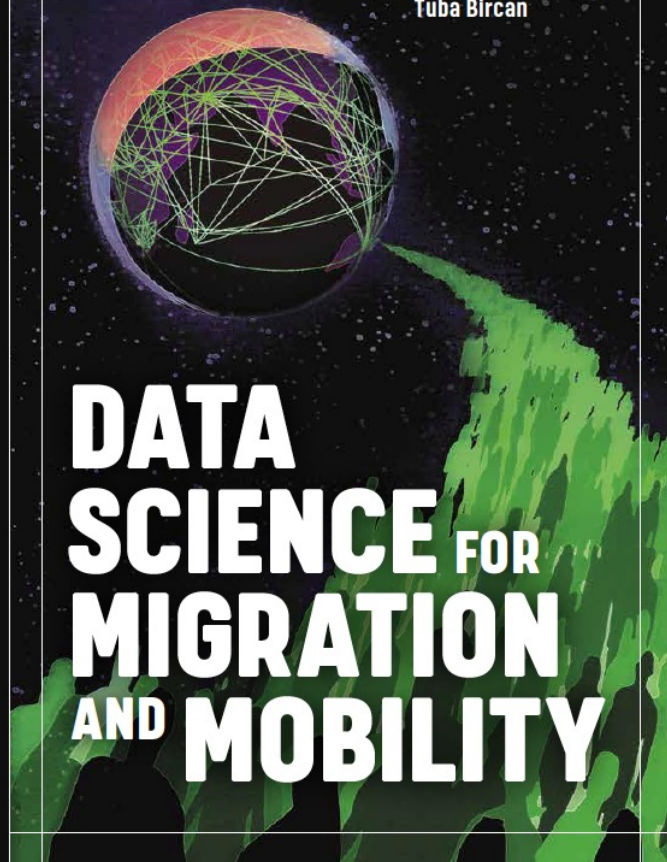
SALAH
KORKMAZ
BIRCAN

DATA SCIENCE FOR MIGRATION AND MOBILITY

PBA
251



Editors
Albert Ali Salah
Emre Eren Korkmaz
Tuba Bircan



19 chapters, 48 contributors
Includes chapters on mobile phone data, satellite data, social media, financial datasets, and case studies.

Book Launch

Data Science for Migration and Mobility

Wednesday 23rd November, 4.30 - 6.30 PM

Seminar Room 1 (ODID, Queen Elizabeth House, 3
Mansfield Road, Oxford, OX1 3TB

Thanks!

For questions and collaboration:

Tuba.Bircan@vub.be

tb25@sanger.ac.uk



@bircantub



COORDINATOR

HIVA - Research Institute for Work and Society, KU Leuven (BE)

PARTNERS

University of Utrecht, Department of Information and Computing Sciences (NL) • University of Salamanca, Faculty of Social Sciences (ES) • Otto-Friedrich-University of Bamberg, Geographic Migration and Transition Studies (DE) • CESSDA ERIC (NO) • EKKE - National Centre for Social Research (GR) • IEN - Institute of Economic Sciences, Data Center Serbia for Social Sciences (RS) • Malmö University, Malmö Institute for Studies of Migration, Diversity and Welfare (SE) • Turkcell Technology (TR) • GMV Innovating Solutions Ltd, Remote Sensing Services and Exploitation Platforms (UK) • MPG - Migration Policy Group (BE) • ENoMW - European Network of Migrant Women (BE) • White Research (BE) • University of Pisa, Department of Computer Science (IT) • CNR - National Research Council (IT) • VUB - Vrije Universiteit Brussel, Interface Demography, Department of Sociology (BE) • University of Zurich, Department of Social Anthropology and Cultural Studies (CH) • CILD - Italian Coalition for Civil Liberties and Rights (IT) • Caritas International (BE)

VISIT: <http://www.hummingbird-H2020.eu>

CONTACT US: HumMingBird@kuleuven.be

FOLLOW US:

