FROM THE DATA REVOLUTION TO GENERATIVE AI: WHERE DOES DEMOGRAPHY STAND?

Emilio Zagheni
Max Planck Institute for Demographic Research

IUSSP/PAA Webinar on AI in Demography, Oct. 16, 2026

Everything has to change for everything to stay the same.

adapted from *The Leopard*, by Tomasi di Lampedusa

THIS TALK

Some examples of AI use in Demography

1. BLEND CORE SOCIAL SCIENCE TOOLS WITH THE INFRASTRUCTURE OF THE DIGITAL AGE

1. BLEND CORE SOCIAL SCIENCE TOOLS WITH THE INFRASTRUCTURE OF THE DIGITAL AGE



Was würde Ihre Nachbarschaft über die Ansiedlung von Neuankömmlingen denken?

*

Bitte wählen Sie eine der Antworten von 1 bis 6, wobei 1 bedeutet, dass Ihre Nachbarschaft nicht akzeptieren würde, wenn sich diese Gruppe in Ihrer Nachbarschaft niederlassen würde. Wie würden Sie die Akzeptanz dieser Gruppe in Ihrer Nachbarschaft bewerten?

„Diese jungen Erwachsenen verlassen ihr Heimatland aufgrund von dort herrschenden militärischen Konflikten. Sie sind zwischen 18 und 30 Jahre alt. Einige von ihnen sprechen etwas Englisch, andere nicht. Die meisten hatten keine Möglichkeit, eine höhere Bildung zu erwerben. Ihr Plan für die Zukunft ist, sich niederzulassen und in Deutschland Arbeit zu finden."

Survey experiment on attitudes towards migrants where images of hypothetical groups of migrants with different characteristics (like gender or skin tone) are AI generated.

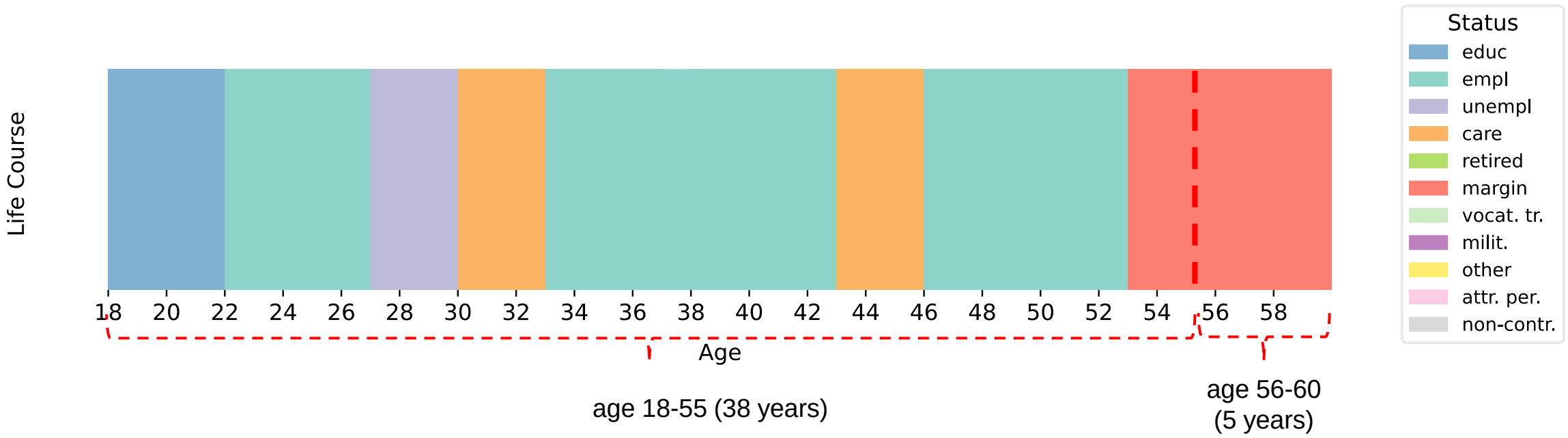
Tsai, Rinderknecht, Palmer and Zagheni (2026) International Migration Review

2. NEW METHODS WITH CLASSIC DATA AND QUESTIONS

Vecgaile, Spata, Vecchietti and Zagheni (2025) SocArxiv

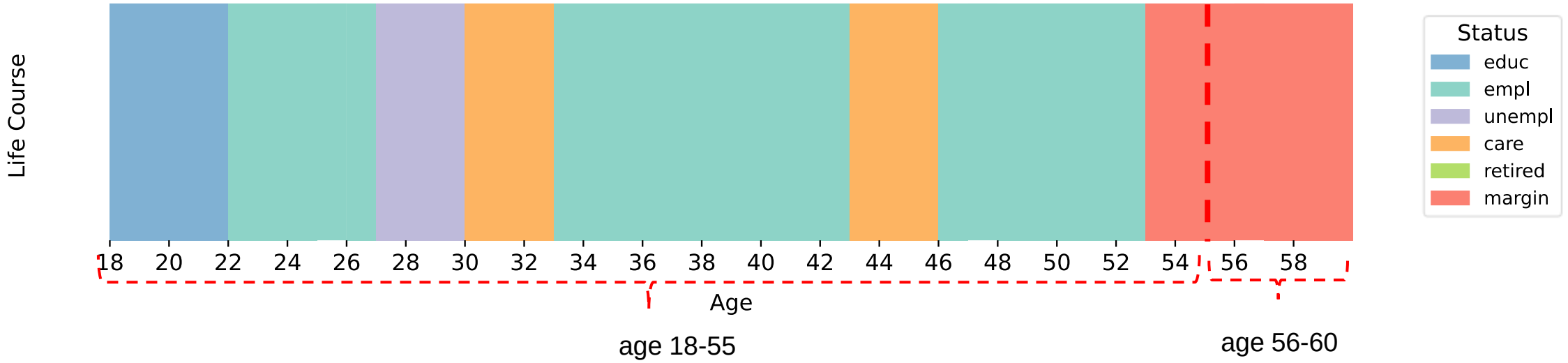
2. NEW METHODS WITH CLASSIC DATA AND QUESTIONS

Modeling life course transitions and the role of early life events on late life outcomes



Data: **Deutsche Rentenversicherung** (German Statutory Pension Insurance)

PREDICTIONS OF EMPLOYMENT STATES TAKING INTO ACCOUNT THE DEPENDENCIES IN LIFE HISTORIES



		Probability distribution (%)				
		Age				
		56	**57**	**58**	**59**	**60**
Social Employment States	**educ**	0.05	0.05	0.1	0.05	0
	empl	0.05	0.2	0.2	0.25	0.2
	unempl	0.25	0.05	0.05	0.1	0.3
	care	0.2	0.25	0.2	0.2	0.25
	retired	0.15	0.1	0.15	0.15	0.2
	margin	0.30	0.35	0.3	0.35	0.05

Vecgaile, Spata, Vecchietti and Zagheni (2025) SocArxiv

3. MAKE UNSTRUCTURED DATA MORE ACCESSIBLE AND USEFUL

3. MAKE UNSTRUCTURED DATA MORE ACCESSIBLE AND USEFUL

YAHOO!

You are where you E-mail: Using E-mail Data to Estimate International Migration Rates*

Emilio Zagheni
Max Planck Inst. for Demographic Research
Konrad-Zuse-Str. 1, Rostock, Germany
zagheni@demogr.mpg.de

Ingmar Weber
Yahoo! Research Barcelona
Av. Diagonal 177, Barcelona, Spain
ingmar@yahoo-inc.com

ABSTRACT
International migration is one of the major determinants of demographic change. Although efforts to produce comparable statistics are underway, estimates of demographic flows are inexistent, outdated, or largely inconsistent, for most countries. We estimate age and gender-specific migration rates using data extracted from a large sample of Yahoo! e-mail messages. Self-reported age and gender of anonymized e-mail users were linked to the geographic

Author Keywords
Demographics, Migration, Mobility, E-mail data

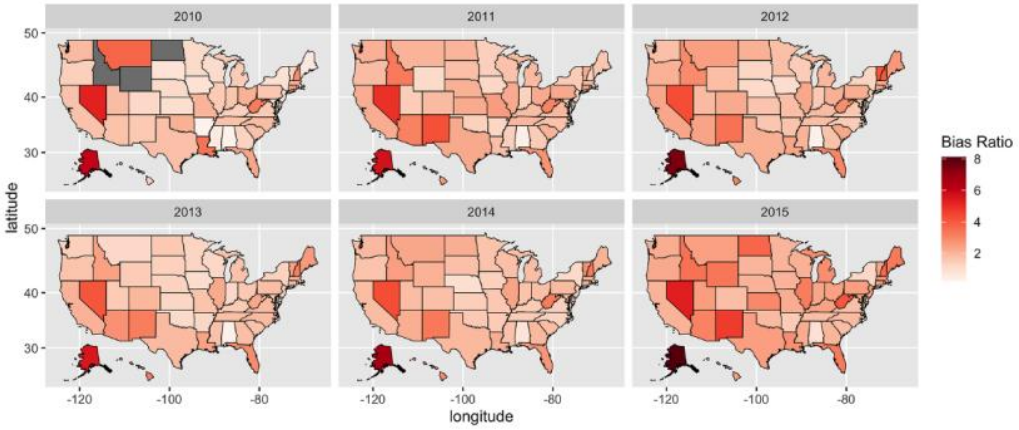
ACM Classification Keywords
J.4 Social and Behavioral Sciences; H.4.3 Communications Applications

General Terms

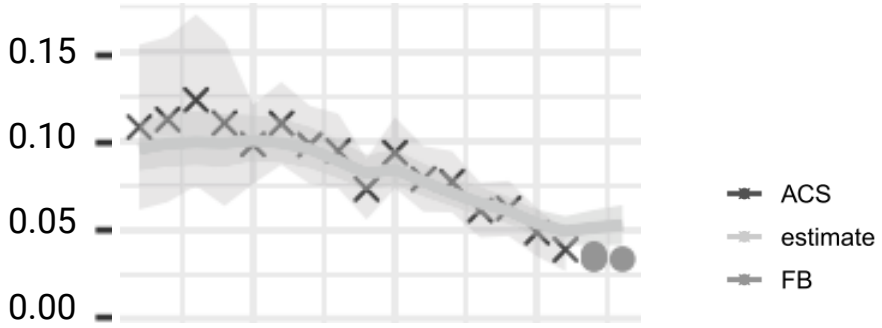
Zagheni and Weber (2012) WebSci

METHODS TO MAKE VALID STATISTICAL INFERENCE FROM SAMPLES THAT ARE NOT NECESSARILY REPRESENTATIVE

Bias ratios by state



Fraction of Mexican men, 25-29, in Georgia US



Original Article

Sociological Methods & Research
1–39
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00491241221140144
journals.sagepub.com/home/smr
SAGE

Modeling the Bias of Digital Data: An Approach to Combining Digital With Official Statistics to Estimate and Predict Migration Trends

Yuan Hsiao[1], Lee Fiorio[2], Jonath
and Emilio Zagheni[4]

Abstract

Obtaining reliable and timely estimates of migrati
cing the migration theory and guiding policy dec
lenge. Digital data provide granular information
not draw from representative samples of the popu

Population Research and Policy Review (2022) 41:1–28
https://doi.org/10.1007/s11113-020-09599-3

ORIGINAL RESEARCH

Check for updates

Combining Social Media and Survey Data to Nowcast Migrant Stocks in the United States

Monica Alexander[1] · Kivan Polimis[2] · Emilio Zagheni[3]

Received: 11 April 2020 / Accepted: 20 July 2020 / Published online: 18 August 2020
© The Author(s) 2020

Demography (2021) 58(6):2193–2218
DOI 10.1215/00703370-9578562 © 2021 The Authors
Published online: 9 November 2021
This is an open access article distributed under the terms of a Creative Commons license (CC BY-NC-ND 4.0).

A Framework for Estimating Migrant Stocks Using Digital Traces and Survey Data: An Application in the United Kingdom

Francesco Rampazzo, Jakub Bijak, Agnese Vitali, Ingmar Weber, and Emilio Zagheni

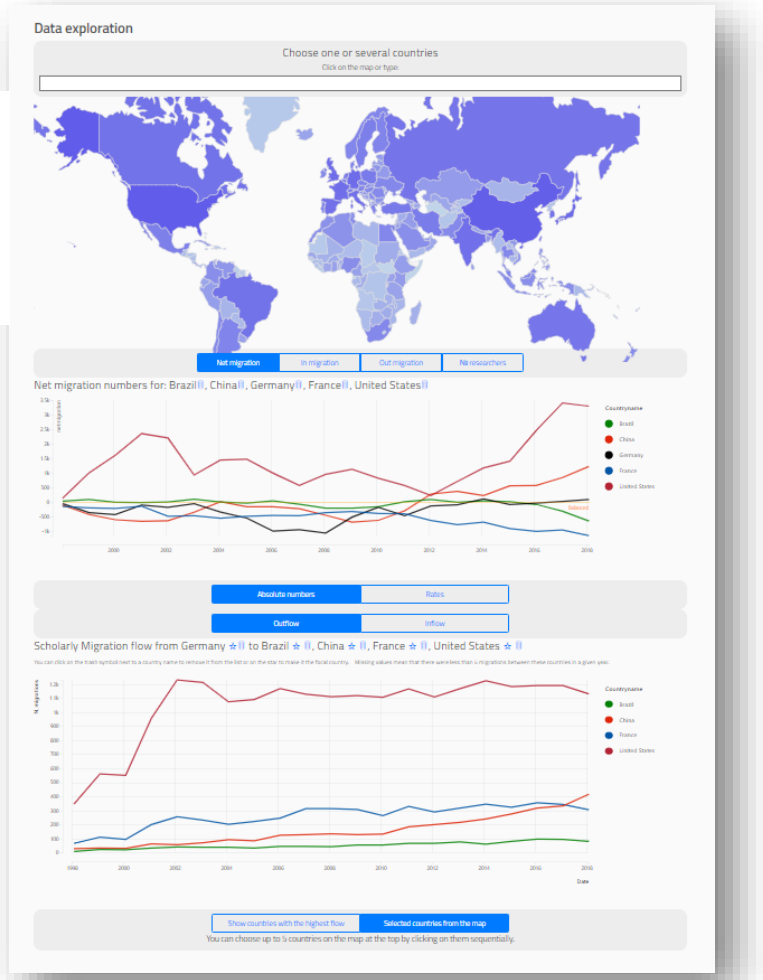
ABSTRACT An accurate estimation of international migration is hampered by a lack of timely and comprehensive data, and by the use of different definitions and measures of migration in different countries. In an effort to address this situation, we complement traditional data sources for the United Kingdom with social media data: our aim is to understand whether information from digital traces can help measure international migration. The Bayesian framework proposed is used to combine data from the Labour Force Survey (LFS) and the Facebook Advertising Platform to study the number of European migrants in the United Kingdom, with the aim of producing more accurate estimates of the numbers of European migrants. The overarching model is divided into a Theory-Based Model of migration and a Measurement Error Model. We review the

UNSTRUCTURED DATA 2.0

UNSTRUCTURED DATA 2.0

Scholarly Migration Database

Quantifying the Mobility of Scholars

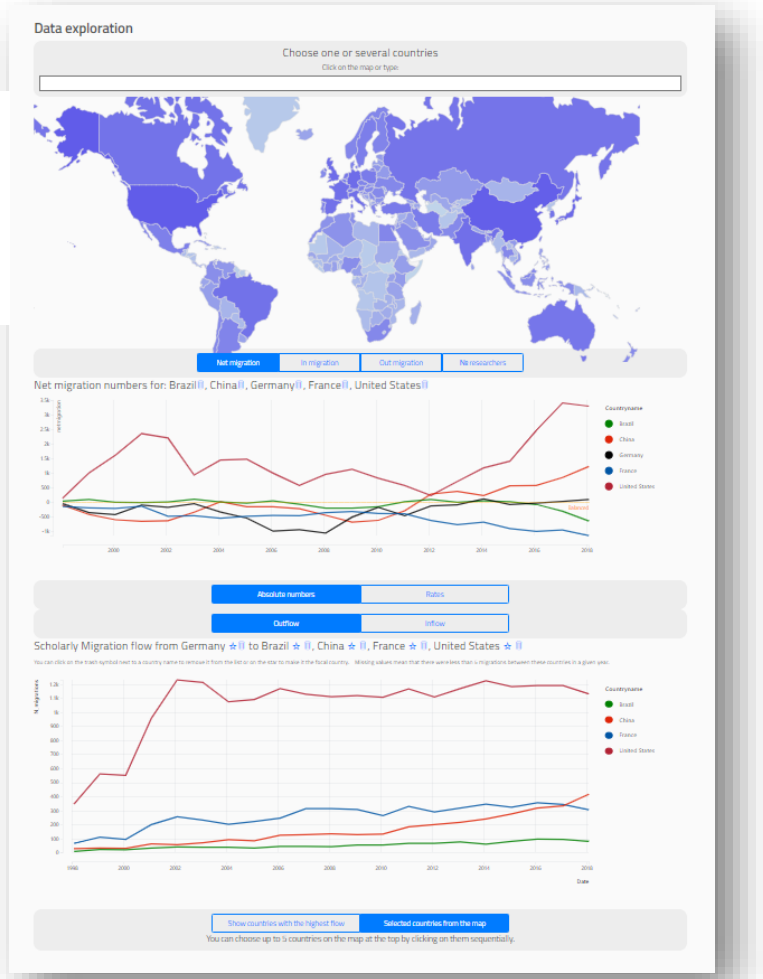


www.scholarlymigration.org

UNSTRUCTURED DATA 2.0

Scholarly Migration Database

Quantifying the Mobility of Scholars



www.scholarlymigration.org

Complement and enhance with CV data that are normally scattered and difficult to organize



4. OPPORTUNITY TO ENHANCE TRANSPARENCY & REPRODUCIBILITY AND TO REDUCE ERRORS

4. OPPORTUNITY TO ENHANCE TRANSPARENCY & REPRODUCIBILITY AND TO REDUCE ERRORS




```php
function register()
{
    if (!empty($_POST)) {
        $msg = '';
        if ($_POST['user_name']) {
            if ($_POST['user_password_new']) {
                if ($_POST['user_password_new'] === $_POST['user_password_repeat']) {
                    if (strlen($_POST['user_password_new']) > 5) {
                        if (strlen($_POST['user_name']) < 65 && strlen($_POST['user_name']) > 1) {
                            if (preg_match('/^[a-z\d]{2,64}$/i', $_POST['user_name'])) {
                                $user = read_user($_POST['user_name']);
                                if (!isset($user['user_name'])) {
                                    if ($_POST['user_email']) {
                                        if (strlen($_POST['user_email']) < 65) {
                                            if (filter_var($_POST['user_email'], FILTER_VALIDATE_EMAIL)) {
                                                create_user();
                                                $_SESSION['msg'] = 'You are now registered so please login';
                                                header('Location: ' . $_SERVER['PHP_SELF']);
                                                exit();
                                            } else $msg = 'You must provide a valid email address';
                                        } else $msg = 'Email must be less than 64 characters';
                                    } else $msg = 'Email cannot be empty';
                                } else $msg = 'Username already exists';
                            } else $msg = 'Username must be only a-z, A-Z, 0-9';
                        } else $msg = 'Username must be between 2 and 64 characters';
                    } else $msg = 'Password must be at least 6 characters';
                } else $msg = 'Passwords do not match';
            } else $msg = 'Empty Password';
        } else $msg = 'Empty Username';
        $_SESSION['msg'] = $msg;
    }
    return register_form();
}
```

4. OPPORTUNITY TO ENHANCE TRANSPARENCY & REPRODUCIBILITY AND TO REDUCE ERRORS




```php
function register()
{
    if (!empty($_POST)) {
        $msg = '';
        if ($_POST['user_name']) {
            if ($_POST['user_password_new']) {
                if ($_POST['user_password_new'] === $_POST['user_password_repeat']) {
                    if (strlen($_POST['user_password_new']) > 5) {
                        if (strlen($_POST['user_name']) < 65 && strlen($_POST['user_name']) > 1) {
                            if (preg_match('/^[a-z\d]{2,64}$/i', $_POST['user_name'])) {
                                $user = read_user($_POST['user_name']);
                                if (!isset($user['user_name'])) {
                                    if ($_POST['user_email']) {
                                        if (strlen($_POST['user_email']) < 65) {
                                            if (filter_var($_POST['user_email'], FILTER_VALIDATE_EMAIL)) {
                                                create_user();
                                                $_SESSION['msg'] = 'You are now registered so please login';
                                                header('Location: ' . $_SERVER['PHP_SELF']);
                                                exit();
                                            } else $msg = 'You must provide a valid email address';
                                        } else $msg = 'Email must be less than 64 characters';
                                    } else $msg = 'Email cannot be empty';
                                } else $msg = 'Username already exists';
                            } else $msg = 'Username must be only a-z, A-Z, 0-9';
                        } else $msg = 'Username must be between 2 and 64 characters';
                    } else $msg = 'Password must be at least 6 characters';
                } else $msg = 'Passwords do not match';
            } else $msg = 'Empty Password';
        } else $msg = 'Empty Username';
        $_SESSION['msg'] = $msg;
    }
    return register_form();
}
```



4. OPPORTUNITY TO ENHANCE TRANSPARENCY & REPRODUCIBILITY AND TO REDUCE ERRORS

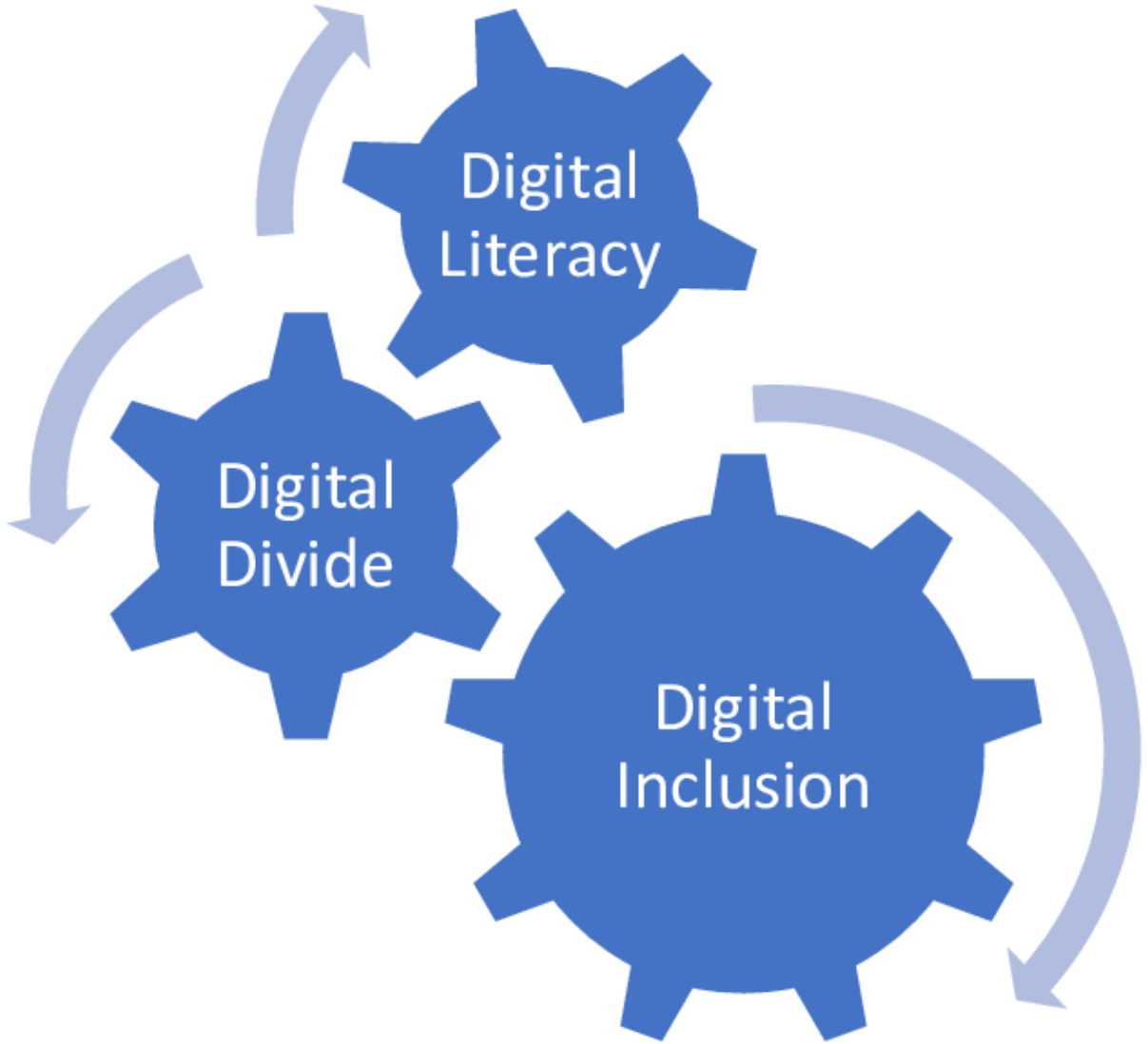



```php
function register()
{
    if (!empty($_POST)) {
        $msg = '';
        if ($_POST['user_name']) {
            if ($_POST['user_password_new']) {
                if ($_POST['user_password_new'] === $_POST['user_password_repeat']) {
                    if (strlen($_POST['user_password_new']) > 5) {
                        if (strlen($_POST['user_name']) < 65 && strlen($_POST['user_name']) > 1) {
                            if (preg_match('/^[a-z\d]{2,64}$/i', $_POST['user_name'])) {
                                $user = read_user($_POST['user_name']);
                                if (!isset($user['user_name'])) {
                                    if ($_POST['user_email']) {
                                        if (strlen($_POST['user_email']) < 65) {
                                            if (filter_var($_POST['user_email'], FILTER_VALIDATE_EMAIL)) {
                                                create_user();
                                                $_SESSION['msg'] = 'You are now registered so please login';
                                                header('Location: ' . $_SERVER['PHP_SELF']);
                                                exit();
                                            } else $msg = 'You must provide a valid email address';
                                        } else $msg = 'Email must be less than 64 characters';
                                    } else $msg = 'Email cannot be empty';
                                } else $msg = 'Username already exists';
                            } else $msg = 'Username must be only a-z, A-Z, 0-9';
                        } else $msg = 'Username must be between 2 and 64 characters';
                    } else $msg = 'Password must be at least 6 characters';
                } else $msg = 'Passwords do not match';
            } else $msg = 'Empty Password';
        } else $msg = 'Empty Username';
        $_SESSION['msg'] = $msg;
    }
    return register_form();
}
```



…But also risks and changing incentives for open science and open data

5. AI AS AN OBJECT OF STUDY IN ALGORITHMICALLY-MEDIATED SOCIETIES



Source: Nguyen et al (2020)

Everything has to change for everything to stay the same.

adapted from *The Leopard*, by Tomasi di Lampedusa

Everything has to change for everything to stay the same.

adapted from *The Leopard*, by Tomasi di Lampedusa

We tend to overestimate the effect of a technology in the short run, and underestimate the effect in the long run.

-Roy Amara

THANK YOU!

